An engineering Research Data Management (RDM) literature review

Version 1.0. March, 2012
Paul Stainthorp
pstainthorp@lincoln.ac.uk
Orbital project, University of Lincoln

Abstract

"The management of research data is recognised as one of the most pressing challenges facing the higher education and research sectors. Research data generated by publicly-funded research is seen as a public good and should be available for verification and re-use" (Hodson and JISC, 2011**a**). To gain a clearer understanding of the more complex and unfamiliar concepts in the emerging discipline of Research Data Management, the Orbital project conducted a review of published literature on the subject (mainly web sites, project reports and guidance documents), with particular reference to RDM in the discipline of engineering. The project team identified nine themes in the literature – for each theme, a recommendation is made which will support the development of RDM infrastructure at the University of Lincoln.

Keywords
bibliography, engineering, glossary, Higher Education, literature review, MRD, Orbital, RDM, research data

Table of contents

This is version 1.0 of the document. An updated list of the sources referenced here is available in an online RefShare database at: http://lncn.eu/bcf6

1. Background to the literature review

This document contains a summary and review of major themes in the published material relating to all aspects of Research Data Management (RDM) in Higher Education Institutions – in particular, to RDM as it relates to the discipline of engineering. Where appropriate, particular recommendations are made for the future of RDM at the University of Lincoln. It contains a bibliography of selected published works relating to all aspects of RDM. It has been produced to support the work of the Orbital project[1] at the University of Lincoln, which has been funded through the JISC Managing Research Data Programme 2011-13[2].

The concepts and arguments within RDM are complex and—in the author's experience—the domain has surprisingly little terminology in common with other HE professional disciplines (library; ICT; educational technology). HE staff have access to a sometimes overwhelming amount of good advice and guidance on RDM. This literature review is an attempt to summarise that information.

> The management of research data is recognised as one of the most pressing challenges facing the higher education and research sectors. Research data generated by publicly-funded research is seen as a public good and should be available for verification and re-use. In recognition of this principle, all UK Research Councils require their grant holders to manage and retain their research data for re-use, unless there are specific and valid reasons not to do so.
> (Hodson and JISC, 2011**a**)
>
>

Most of the material was retrieved from open web (using Google and Google Scholar[3]). In addition, a number of websites and documents were discovered through attendance at MRD conferences and workshops. The bulk of the literature discussed here consists of web pages, project documentation, and briefing documents aimed at staff in HEIs. It is worth highlighting that the following authors (individual and corporate) have several entries in the bibliography each:

- Dr Mansur Darlington and colleagues at the University of Bath, Innovative Design & Manufacturing Research Centre (I*d*MRC)[4], and at UKOLN[5]. Their discipline-specific research into engineering RDM through two JISC-funded projects, along with Dr Darlington's rôle as consultant on the Orbital project, have provided Orbital with a number of immediately useful [re-usable] findings and models. See: Darlington *et al.* (2010; 2012**a**;**b**;**c**); Ball *et al.* (2010**a**;**b**); Howard *et al.* (2010)
- The Digital Curation Centre (DCC)[6] is a UK centre of research data management expertise, and exists to "promote and publicise curation concepts" (Allan, 2011) to anyone involved in RDM.
- JISC[7] is a body which supports and develops the use of ICT in Higher and Further Education in the UK. They have co-ordinated two major programmes of funded RDM projects (2009-11; 2011-13).

1 Orbital project: http://orbital.blogs.lincoln.ac.uk/
2 JISC Managing Research Data Programme 2011-13: http://lncn.eu/c47
3 Google Scholar: http://scholar.google.com/
4 University of Bath, I*d*MRC: http://www.bath.ac.uk/idmrc/
5 UKOLN: http://www.ukoln.ac.uk/
6 Digital Curation Centre (DCC): http://www.dcc.ac.uk/
7 JISC: http://www.jisc.ac.uk/

2. Themes in the RDM literature

The following 9 themes were identified through notes made at RDM conferences and workshops, through advice and guidance documents provided (mainly) by the Digital Curation Centre—many of which can be found in the bibliography section of this document—and by copying-and-pasting the text of some of the works into the word cloud generator, Wordle[8], which made it easy to recognise the key terms in each work.

Word cloud of the RSS feed of the JISC Managing Research Data Programme blog[9], March 2012

2.1. Fundamentals of research data and RDM

Darlington (2012**b**) gives a formal definition of research data: “re-interpretable representations of information in a formalized manner suitable for [...] processing”, and goes on to define terminology for data *records* (the real, digital representation of the data), *cases* (the set of data records associated with a discrete research activity), and other entities encountered in research (Darlington, 2011)[10]. In the terminology promoted by the DCC, a single ‘data object’ consists of “all homogenous data, documentation and metadata relating to a specific [research] project.” (Alexogiannopoulos, McKenney and Pickton, 2010).

Researchers may keep data on their own computer or laptop, on personal storage media (*e.g.* USB ‘thumb’ drives), on an institutional networked file store or dedicated data server, and in cloud storage services *e.g.* Dropbox[11] – with different storage choices preferred “at the data collection, analysis and project completion stages” (Alexogiannopoulos, McKenney and Pickton, 2010). Researchers in UK HEIs can also contribute data to one of several national research data stores (Allan, 2011) which support “thousands of researchers and millions of downloads each year” (JISC, Research Information Network, and

8 Wordle: http://www.wordle.net/

9 JISC Managing Research Data Programme: http://researchdata.jiscinvolve.org/wp/

10 It should be noted that some of these terms are not in common use, even in the RDM community, but that detailed discussion of RDM practice between researchers and others can quickly run into ambiguity if clear terminology is not agreed upon in advance.

11 Dropbox: https://www.dropbox.com/

Technopolis Ltd, 2011**c**): all of these methods of deposit, storage and preservation are within the scope of RDM. **It is recommended** that the Orbital project continue its work to assess the storage and other requirements of University of Lincoln researchers using surveys and interviews, and make its findings available to the University to assist RDM planning.

Models for characterising and describing the different stages and processes involved in RDM include: the *DCC Curation Lifecycle Model* (Digital Curation Centre, 2010**c**; described in Ball, 2010**a**); and the *I2S2 Research Activity Life Cycle Model* (Beagrie, 2011**b**). Howard *et al.* (2010) evaluates both models with respect to engineering research data, noting that while the DCC model is aimed at institutional 'curators' of someone else's data, the I2S2 model "recognizes the – often quite different – needs of researchers in supporting data re-use" (*ibid.*), and that the I2S2 Life Cycle provides for a more detailed description of "typical" scientific research.

The ERIM project described and made distinctions between different types of high-level research data preparation activities and use activities (Ball *et al.*, 2010**b**; Darlington, 2012**b**) – their categories include Data *Purposing* (making data available and fit for current research activity); *Re-purposing* (making data fit for a known future activity); *Supporting re-use* (preparation for future unknown activity); *Use* and *Re-use* (for the purpose for which it was gathered, and purposes not originally intended, respectively).

Factors affecting the requirements of a RDM system include:

1. The growth in the amount of digital data being produced by researchers. Some 'big' science disciplines are now capable of generating "truly challenging quantities" Gray, Carozzi and Woan, 2011) of data for an institution "near the edge of what it is technically feasible to store and transport": research data measurable in petabytes—*i.e.* millions of gigabytes—per year.[12]
2. The differences of 'big science' are not just ones of quantity (Allan, 2011); big science data is "shared and exploited in ways which differ from other disciplines" (Gray, Carozzi and Woan, 2011). The phrase "data-driven science" (Allan, 2011) is used to describe research that uses existing accessible research data to generate new research outcomes, while perhaps not generating any new primary data of its own. Reproducibility in research (*Science*, 2011) is awarded a higher importance and increases the requirements on access to data.
3. In general, a move toward multi-disciplinary research is increasing the demand to share data and make it more retrievable across space, time, and different working methods (Allan, 2011).
4. Changes in the desktop and lab technology available at HEIs are leading to more complex research workflows (Alexogiannopoulos, McKenney and Pickton, 2010; Allan, 2011). Partly offset by a tendency toward standardisation in approaches to generating, collecting, and manipulating data (Darlington *et al.* 2010).
5. Pressure from Government for Open Access to data, Freedom of Information legislation, and a culture of accountability (Allan, 2011; Ball, 2010**a**). Mandates from research funding bodies (see section 2.4 below).
6. Demand generated within institutions by the success of publication repositories (Allan, 2011).

12 It is worth noting that the data being produced at the University of Lincoln nothing like on this scale. Initial survey responses from researchers suggest data of between a few gigabytes and 1 terabyte is being held by individuals at Lincoln – but considering the 'edge case' will help the University to plan for the tendency for all disciplines to produce more and more data.

2.2. Particular requirements of the discipline of engineering

With reference to the ERIM (Engineering Research Information Management)*[13] *and REDm-MED (Research Data Management for Mechanical Engineering Departments)*[14] *projects at the University of Bath, Innovative Design and Manufacturing Research Centre (Id***MRC).***

Considered RDM systems can benefit engineering research by helping to reduce duplication of effort and the unnecessary re-collection of data, by providing inspiration for new research, by improving the transparency and thus the reproducibility of research, by providing a consistent process for data description and citation – leading to new research opportunities and collaborations, new sources of funding, and increased scholarly output (Darlington, 2012**b**).

Darlington *et al.* (2010) describe a set of "consistent, high-level" ideal principles for managing information (and, by extension, research data) in the discipline of engineering (reproduced below); those principles are given further weight by a set of statements about researcher behaviours, planning, and documentation designed to ensure reproducibility, and ease of data re-usability.

The Principles of Engineering Information Management
(Darlington *et al.*, 2010)

1. The Principle of *Parsimony*: Create, record and retain information only if necessary.
2. The Principle of *Granularity*: Record information in a storable information object at a granularity appropriate for use and re-use.
3. The Principle of *Identity*: Give an information object a unique and persistent identifier.
4. The Principle of *Uniqueness*: Create an information entity once only and explicitly reference it everywhere else.
5. The Principle of *Usability*: Design an information object explicitly to achieve its intended goals.
6. The Principle of *Reusability*: Design an information object explicitly to maximise its potential for reuse wherever appropriate.
7. The Principle of *Evaluation*: Assess and assign the value of an information object throughout its life from creation to disposal.
8. The Principle of *Portability*: Create an information entity and its annotations systematically using representations supporting perpetual reuse.
9. The Principle of *Robustness*: Use robust methods to capture, create and manipulate information entities.
10. The Principle of *Discovery*: Actively employ the information repository as a resource for learning and discovery.
11. The Principle of *Design*: Design all aspects of information management to satisfy the organisation's current and future needs.

13 ERIM project: http://www.bath.ac.uk/idmrc/erim/
14 REDm-MED project: http://www.ukoln.ac.uk/projects/redm-med/

Engineering research data itself is “diverse in character, spanning everything from material properties to questionnaire responses and interview transcripts” (Howard *et al.*, 2010); engineers’ research and RDM *activities* are both equally wide, and the discipline require a flexible and accurate way of describing and categorising research data, research activities, and research data management activities.

Key ERIM Research Findings
(Darlington, 2012**b**)

- Great diversity of data type and quality
- Complex and chaotic nature of data development
- Outputs not linked to data
- Supporting documents not situated with the data files
- Little use of metadata to support future use
- Immature understanding of benefits of sharing and thus need for management
- Limited understanding of the barriers to opportunities for information sharing and re-use
- Poor framework for:
    - pre-project considerations of data management
    - data management during the research
    - during-project data management for post-project re-use
- Poor knowledge of context in which data were generated:
    - engineering research data are very diverse
    - large number of diverse research data records
    - relations between data records complex

Knowing the context is vital to understanding data.

After reviewing the DCC and I2S2 RDM lifecycle models (described above), the ERIM project (Howard *et al.* 2010) proposed and tested experimentally a new, detailed model aimed specifically at engineering research. Building on the notions of *Data Purposing, Re-purposing, Re-use*, *etc.*, ERIM’s model consists of:

1. A taxonomy of data objects: different types of record which form the building blocks of research (*Research Data Record*; *Contextual Data Record*, *Experimental Apparatus Data Record*, *etc.*)
2. An enumeration of the processes to which research data is subjected during the research activity (*Addition*; *Aggregation*, *Deletion*, *Refinement*, *etc.*)
3. RAID (Research Activity Information Development) diagrams, a formal system for visually modelling the relationship between data objects and the development processes described above.

This ERIM project model a.k.a. RAID is comprehensive—albeit inevitably more complex than the more well-used RDM lifecycle modelling tools—and its development of an “emerging terminology” for engineering is absolutely suited to the requirements of that School’s research at the University of Lincoln – though its relevance is by no means limited to that one discipline.

The University of Bath’s new REDm-MED project is working specifically with researchers in mechanical engineering, developing RDM planning tools “specialized for each new research project” and including integration with Bath ICT services (Darlington, Ball and Thangarajah, 2012**c**) to provide practical RDM support (Darlington, 2012**a**). REDm-MED will develop and implement a prototype of the RAIDmap tool to model the information environment of researchers in engineering.

It is recommended that the Orbital project continue to work with Dr. Darlington and the REDm-MED project team on implementing the findings of ERIM into Orbital.

2.3. The behaviour of researchers

Any system of institutional RDM infrastructure must be designed around the researcher: either to so seamlessly work with their existing ways of conducting and managing research activities that there is no additional burden placed upon them which might dissuade them from taking care of their data; or—if changing their behaviour is unavoidable or very desirable—to offer such clear and obvious benefits that the researcher is happy to abandon old ways of working and move to new RDM-aware methods.

However, cohorts of academic researchers tend to be made up of a diverse, heterogeneous group of individuals (Alexogiannopoulos, McKenney and Pickton, 2010) which can make it difficult to generalise about their behaviour, or to use any one researcher to make predictions about another. Studies by the Research Information Network (RIN)[15] of groups of researchers in the life sciences found divergence between the RDM behaviour of individuals “in the same team, between teams in the same discipline, and between disciplines” (Ball, 2010**a**).

However there are *some* identifiable patterns. Alexogiannopoulos, McKenney and Pickton (2010) identified clear differences between researchers of different ages and/or lengths-of-service in academia. More established researchers tended to show greater awareness of RDM issues, and to be more willing to adapt to new régimes of data management where it could be demonstrated to them that a change in their behaviour would have a positive effect on their time or quality of their work. In contrast, researchers from “Generation Y”[16] were more blasé about the need to look after their data, even if this was masked by a greater surface familiarity with the use of ICT to manipulate and store digital data (*ibid.*).

There will also be particular behaviours more common in particular subject disciplines. In engineering, Ball (2010**a**) recorded a tendency toward the “gathering and organization of data for immediate use in the course of the present research”: a functional, pragmatic approach to RDM in contrast to other disciplines where data management practices may involve more long-term storage of data for purposes unknown or understood only vaguely at the point of collection.It is worth noting that the Orbital project team has already noticed this approach to RDM n the Lincoln School of Engineering.

Engineers tend to adopt data management activities which lead to immediately visible benefits such as new industrial applications and collaborations (Darlington, 2012**b**); by extension, they may be less responsive to more ‘aspirational’ practices in RDM such as Open Access to data ‘for the public good’. In an unusual 3D visualisation of perceived benefits[17] influencing the behaviour of researchers in physics and astronomy disciplines, Gray, Carozzi and Woan (2011) suggested that researchers have no particular reason to be motivated by “‘open data’, and only incompletely interested in ‘good metadata’”.

In addition to cultures of behaviour within the ‘lab’, we must try and understand the behaviour of the institution as a corporate whole, and of researchers’ relationship with central support departments and services. Researchers value support and “centralised links to services” (Beagrie, 2011**a**), and although there is often divergence between what a centralised information or policy service might recommend and

15 Research Information Network: http://www.rin.ac.uk/

16 Generation Y: applied to people born from the mid 1970s onwards, approximately, so researchers who were in their late 20s and 30s in the year 2010.

17 Arising out of the Keeping Research Data Safe benefits framework (Beagrie [and others], 2011**b**).

what researchers actually *do* in practice (Ball, 2010**a**), where ICT and library services can develop the right sort of competencies to support RDM they can impact very positively on RDM across the institution: some institutions have created data libraries, data *librarians*, and data stewards/managers in recognition of the importance of RDM for 21st-century academia.

Clearly it is important to gain an understanding the ‘culture’ of any given set of researchers, and of their parent institution, as is understanding what might motivate researchers to consider changing their RDM behaviour. ‘Carrots’ within RDM infrastructure can be designed to reward constructive behaviour without placing unnecessary burdens on existing busy researchers – for instance, by unlocking more beneficial features in return for documenting their data, or for making data available for re-use (Beagrie, 2011**a**).

In presentations to engineers at the University of Lincoln, Darlington (2012**a**) listed both ‘sticks’ and ‘carrots’ including funder and statutory requirements (see section 2.4 below), but also a list of discipline-resonant benefits to engaging in active RDM, particularly in processes which lead to data re-use and sharing: *e.g.* “knowledge extraction from aggregated data” (*ibid.*), in language which speaks to researchers in the subject.

It is recommended that Orbital develop on its initial use of surveys, interviews, and close working alongside researchers to understand their requirements and existing RDM ‘culture’ at Lincoln. It should work to develop advocacy/training materials that speak to what motivates researchers, as well as highlighting obligations to their institution, funders, and to the ‘public good’. The University of Lincoln may also want to consider future job roles (“data librarians” or similar) that would have a positive effect on a culture of RDM.

2.4. RDM policies and legal aspects

Higher Education Institutions in England and Wales are subject to legislation including the Freedom of Information Act 2000[18]. While the act was not drawn up with research data in mind, public controversies involving research data such as ‘Climategate’ at the University of East Anglia[19] have concentrated institutional focus on HEIs’ statutory obligations to research data (Allan, 2011; Ball, 2010**a**). More ‘ethical’ considerations such as the confidentiality of research subjects are better understood by researchers and their institutions and will usually be covered by institutional research ethics procedures (even where the issues overlap with Data Protection legislation); RDM infrastructure must be capable of recording information about the legal and ethical sensitivities of particular data objects.

For research datasets produced by university employees in the course of their employment, the copyright of the data will by default belong to the institution. By long convention, most universities have policies where the copyright in ‘scholarly’ outputs is assigned to its original author(s), although for research data with potential commercial applications this may be qualified by IP and enterprise policies where the institution retains control over how the data is exploited. These arrangements are not always understood by researchers. who, when asked the question, “‘Who owns your data?’ … were often unable to give a clear answer.” (Alexogiannopoulos, McKenney and Pickton, 2010).

All of the UK Research Councils, when awarding funding to institutions to conduct research projects, now have specific policies on how any research data produced must be managed (Digital Curation Centre, 2010**a**). All require researchers to produce a data management plan (DMP); all specifiy some measure of data sharing or publication (see section 2.5 below). Of particular interest to the Orbital project are the

18 *Freedom of Information Act 2000*: http://www.legislation.gov.uk/ukpga/2000/36/contents
19 ‘Climategate’ controversy: http://en.wikipedia.org/wiki/Climatic_Research_Unit_email_controversy

EPSRC's "expectations concerning the management and provision of access to EPSRC-funded research data" (Engineering and Physical Sciences Research Council, 2011**a**[20]). The expectations contain a number of statements with which the University of Lincoln must comply if it wishes to continue to receive EPSRC funding, including promotion of the principles to staff, statements and policies on access and re-use, and metadata and data itself being made freely available on the Internet within a specified period of the data being generated, with access being maintained for a rolling period of 10 years after the most recent access (Engineering and Physical Sciences Research Council, 2011**a**;**b**).

The EPSRC have asked all institutions it funds to have developed a "roadmap to align their policies and processes with EPSRC's expectations by 1st May 2012, and to be fully compliant with these expectations by 1st May 2015" (Engineering and Physical Sciences Research Council, 2011**b**). The EPSRC will respond with sanctions—likely to include the withholding of future research funding—if "proper sharing of research data is being obstructed" by non-compliant institutions.

EPSRC first principle on management of research data

1. "EPSRC-funded research data is a public good produced in the public interest and should be made freely and openly available with as few restrictions as possible in a timely and responsible manner."

(Engineering and Physical Sciences Research Council, 2011**a**)

The interests of researchers and institutions are protected, in part, by a period of 'privileged access'—elsewhere described as an embargo or 'proprietary period' (Gray, Carozzi and Woan, 2011)—before the end of which they will not be expected to share the data, so that the originators of the data can be the first to publish. Researchers can also expect appropriate acknowledgement (in the form of citations—see section 2.7 below—or attribution statements) for their data.

Institutions also have a duty to protect the interests of their researchers and to consider appropriate safeguards. Universities will need to provide funds and staff support for their RDM infrastructure, though the EPSRC (2011**a**) along with other Councils has stated that it is "appropriate to use public funds" to support RDM in order to meet their data policy requirements.

The Digital Curation Centre have identified four universities in the UK—Edinburgh, Hertfordshire, Northampton and Oxford—that have their own, public RDM policies or statements of commitment (Hodson and JISC, 2012), with more in the pipeline.The EPSRC funding statement (above) is the greatest driver encouraging universities to develop policies, alongside JISC funding, with all 17 infrastructure projects in the JISC MRD funding programme (which funds Orbital) developing policies for their institutions (Hodson and JISC, 2011**a**), often supported by "guidance materials in the form of researcher targeted scenarios or workthroughs" (Hodson and JISC, 2012)

It is recommended that in the short term, Orbital must support the University's response to the EPSRC data policy roadmap (due 1st May, 2012). Longer term, Orbital (both the project and the RDM application it will produce) can support the University of Lincoln in developing and implementing institutional policies that complement its obligations to funders and its statutory requirements.

20 EPSRC: http://www.epsrc.ac.uk/

2.5. Data sharing

"Data are most useful when they are interoperable with other data" (Ball, 2010**a**). The principle that the fruits of publicly-funded research should be publicly accessible has been given a boost first by Open Access mandates for published outputs, and more recently by the funding councils' data policies mention in section 2.4 above (Allan, 2011; Digital Curation Centre, 2010**a**).

However, the distribution of Research Council money is uneven: at some universities a relatively small proportion of researchers are in receipt of funding which carried with it a requirement to provide Open Access to research data (Alexogiannopoulos, McKenney and Pickton, 2010); other universities rely more heavily on NHS-funded or commercially-sponsored research, where there are no OA mandates, and/or where patient confidentiality or commercial IP contracts may preclude open sharing of data. However, even small amounts of funding—even one project—from a body mandating data sharing would require a university to adopt the same kinds of RDM procedures as an institution where 100% of research income was from mandating funders: the requirement for OA-publishing *capability* exists irrespective of the amount of funding received.

Some institutions (and some disciplines) prefer to publish data to national data centres where there is evidence that data pooling has had a beneficial effect on collaboration within a field (JISC, Research Information Network, and Technopolis Ltd, 2011**c**); many institutions are investing in data repositories as a part of their RDM infrastructure. However, and despite the benefits to researchers in sharing data (identified in sections and 2.3 and 2.6), RDM and data publishing tends to be regarded as a low priority in institutions, even when compared to the arguably *already* low priority of meeting funders' OA publication mandates. There are "contextual, technical, legal and social barriers to the re-use and re-purposing" (Darlington, 2012**b**) of data which dissuade researchers from tackling the subject; data are neglected once a project is complete (Alexogiannopoulos, McKenney and Pickton, 2010); and institutions are often instinctively wary of Open models of business and worried about maintaining institutional IP.

It is recommended that the Orbital team should continue the initial explorations into Open Access, Open Data, and Open Source models that are appropriate for the University, and work with Research & Enterprise to formulate clear policies on data sharing and licensing alongside a business case for Open Access to data, and clear guidance to researchers about their rights, obligations, and the benefits of data sharing.

2.6. Costs and benefits

Beagrie (2011**b**) highlights that the "costs of archival storage and preservation activities are ... a very small proportion of the overall" costs involved in building and maintaining an RDM infrastructure, instead highlighting the costs incurred at the "acquisition and ingest" stages Costs of curation and maintenance tend to find their own institutional level irrespective of the size of the data collection (*ibid.*), although data storage costs are certainly non-trivial for large datasets, and some institutions do not have effective costing mechanisms for providing multi-terabyte-scale storage. Some institutions have adopted a transparent costing model (Digital Curation Centre, 2011), while others use the 'carrot/stick' approach where data storage is awarded to researchers on the basis of their compliance with RDM procedures.

Effective RDM can produce a number of measurable *academic* benefits for institutions. Data can be retrieved more quickly, typically from one or more days "down to an average of around five minutes" (Beagrie, 2011**a**), "shortening time gaps and leading to a more rapid progression towards final publication". Data sharing and data pooling facilities such as the UK data centres were felt by academic users to improve the impact of their data through re-use (JISC, Research Information Network, and

Technopolis Ltd, 2011**c**): despite not all re-use being recorded because of the limited use of data citations.

Cost savings are more difficult to identify, he KRDS ("Keeping Research Data Safe") model (Beagrie, 2011**b**) uses elements of Full Economic Costing "in order to support more accurate cost–benefit analysis" of RDM (Ball, 2010**a**), and persuasive estimates of staff and process efficiency savings of up to 37% were made in one project (Beagrie, 2011**b**). Darlington (2012**a**) lists pragmatic ways in which engineering researchers can benefit from RDM – reproduced below.

Wanting to Manage Research Data
[...a list of benefits]

(Darlington, 2012**a**)

- Reduction in duplicated work
- Inspiration for new/continuation research & funding
- Basis for knowledge extraction from aggregated data
- Greater transparency of research
- Improved basis for validation
- Obviating the need for re-collection and generation
- Providing basis for reliable data citation
- Increasing scholarly output

It is recommended that guidance and advocacy materials drawn up as part of the Orbital project concentrate on articulating the *academic* benefits of RDM to researchers, and that institutional policies, along with support materials, develop the practice of reflecting appropriate RDM costs into future research funding bids, in consultation with Research & Enterprise, commercial partners, and funding bodies. It is also recommended that the University develop appropriate plans for meeting the costs of long-term storage, preservation and curation of research data.

2.7. Curation standards, metadata and citation

"The need to cite data is starting to be recognised as one of the key practices underpinning the recognition of data as a primary research output rather than as a by-product of research."
(Australian National Data Service, 2011)

If there are such demonstrable benefits to the management, sharing, and re-use of data, then we need a robust system to identify and track individual data elements and their provenance (Ball, 2010**a**). Un-citable data, abandoned and disowned after the end of a project or PhD thesis, is all but unusable. Beagrie (2011**a**) argues there can be little progress in effective curation and sharing of research data "unless the practice of citing data used is better encouraged", alongside agreed standards for citing data.

DataCite[21] is an organisation (convened by a group of leading research libraries and technical information providers) and an application that makes it possible to 'mint' identifiers, in the form of DOIs (Digital Object Identifiers)[22], for datasets and data objects, allowing them to be "handled as independent, citable, unique scientific objects" (Australian National Data Service, 2011).

21 DataCite: http://datacite.org/
22 The Digital Object Identifier system: http://www.doi.org/

Moving outwards from an identifier for each data object, data stored by a researcher can benefit from consideration given to “file and directory naming schemes” (Allan, 2011), and data have their own metadata requirements. Without a system for associating and preserving metadata for datasets *and for individual data elements within datasets*, the “relevance of data and information extracted from it can be lost” (*ibid.*) and scientists unable to query and select relevant pieces of data, even if the file ‘container’ is itself relatively well described (Ball, 2010**a**). The Digital Curation Centre (2010**b**) provide advice on metadata standards and their application to research data.

As identified in section 2.6, above, data curation is a “non-trivial investment” (Ball, 2010**a**) which when performed with discrimination, adds value to the data through appropriate description and facilitation of discovery, as well as maintaining its basic existence. Analysis of the different rôles of the various people involved in RDM, and associated terminology (Ball *et al.*, 2010**b**), might provide the “data librarian” with a ‘to do’ list of questions to ask his or her colleagues: how is data being managed, what is it being used for, what documentation exists, *etc.* – a special data case of the reference interview. Curation of data might be seen as a “natural progression for a preservation-focused repository” (Hitchcock, 2011).

There are a number of different models for data curation which are all necessarily simplifications of complex and varied practice, and all not necessarily observed to the letter by even the most RDM-committed researchers and institutions (Ball, 2010**a**). The models include the *DCC Curation Lifecycle Model*, *ANDS data sharing verbs*, *DDI life cycle model*, *UK Data Archive data model* (described in Ball, 2010**a**) – it is important to note that the models are not exclusive and elements of several can be incorporated into good RDM practice.

The Open Archival Information System (OAIS) Reference Model is a conceptual framework for describing the functions of archival repositories. Ball (2006) evaluates the OAIS Reference Model with respect to the discipline of engineering. OAIS is a useful “reference model” for [data] repositories (Allinson, 2006): it provides guidance on standards for ingest, administration, provenance, access, and preservation of objects in repositories – but it does not tackle the question of metadata standards (above), and it can be a daunting and sometimes confusing document. Ball (2006) identified a low level of compliance with the model in repositories in practice.

Finally, RDM systems and associated data repositories should be cognisant of the CERIF (Common European Research Information Format)[23] data model for interoperability between RDM and other research-information-management systems (such as funding databases, audit/submission systems, *etc.*)

It is recommended that Orbital incorporate the functionality of DataCite via its API, into the Orbital application, to allow Lincoln researchers to secure a DOI (Digital Object Identifier) for their data objects. This more than almost any other feature will secure the ability of the University of Lincoln and its staff to publish and share datasets with confidence. It is also recommended that the Library be engaged specifically to consider the curation and description of assets held in data repositories with reference to appropriate curation standards, as a natural extension of its (the Library’s) support for the Lincoln (publications) Repository, and to offer training and support for other University staff.

23 CERIF: http://www.eurocris.org/Index.php?page=CERIFintroduction&t=1

2.8. Technical considerations

The range of file formats involved in engineering research is a significant area of complexity, reflecting the diversity of research which goes on under the banner of engineering. “There are often many different file formats available for storing the same data” (Australian National Data Service, 2011), and the choice of file format is a “subtle” (*ibid.*) and important choice when it comes to sharing and preservation of said data.

In engineering and other physical sciences there is a distinction between truly raw data (sometimes instrument-specific and proprietary, certainly difficult to share), and treated data ‘products’ (prepared from the raw data for analysis, scientifically meaningful, using a standard [possibly Open] format). It is this latter stage of data which is “naturally archived, most carefully documented, and which will eventually be made public” (Gray, Carozzi and Woan, 2011). *N.B.* this distinction is already a subject of necessary discussion in the Orbital project, where proprietary data from Siemens instrumentation must be made available for use by engineers “who do not have an intimate connection with, and knowledge of, the instrument” (*ibid.*). On the subject of proprietary data, there have been “several attempts to use XML to describe binary data” (Allan, 2011), but with little lasting success: any RDM system and data repository must expect to receive data which it is not immediately possible to interpret or query, and this is a problem for re-use and scientific transparency.

It is recommended that Orbital continue to work with Siemens, the School of Engineering, the University of Bath and the DCC to develop institutional expertise in handling engineering data formats.

2.9. Tools, support and training

A number of bodies offer training, support, and tools to help researchers and institutions with RDM. Firstly, the *research funding councils* (as described in section 2.4 above) are introducing data management policies, and some have produced comprehensive guides for grant applicants on meeting their requirements. Other bodies such as the UK Research Integrity Office ([n.d.]) and data centres such as the UK Data Archive (*et al.*, 2011) have also produced guides aimed at researchers.

Centrally (and aimed more at institutional information managers), *JISC* have supported RDM through at least two programmes of funding, including the Managing Research Data Programme 2011-13 (Hodson and JISC, 2011**b**) which funds Orbital. The outputs of the first JISC MRD programme (2009-11) are numerous and include:

- ‘How-to’ guidance on RDM for researchers and institutions;
- Tools for conducting data management planning and institutional benchmarking;
- A number of case studies (JISC, 2011**a**);
- Training guides from JISC (2011**b**) both generic and discipline-specific.

The *Digital Curation Centre* provides the introductory ‘DCC 101’ (2010**d**) which are designed to encourage “data custodians, producers and users” to share best practice in order to create an institutional culture of RDM, along with guidance on selection, citation, planning, and licensing (Digital Curation Centre, 2010**b**).

The DCC also hosts and maintains/develops a number of important tools which support RDM, including:

1. DMP Online, aimed at the researcher, a web-based tool designed to help grant applicants write data management plans “according to the requirements stipulated by the major UK funders” (Digital Curation Centre, 2011**b**). Three iterations of the tool exist, to support minimal (application stage),

core, and full planning. The ERIM project gives examples of indicators of when various data re-use/re-purposing scenarios would be appropriate and desirablem and how to prepare data and DMPs for each case (Darlington, 2012**b**). The Australian National Data Service (2011) and Ball *et al.* (2010**b**) provide detailed explanations about producing DMPs.

2. The Data Asset Framework (DAF), designed to help institutions assess the data management practices and the data held by individual researchers and research groups. It consists of a set of survey methodologies, and an online recording tool Designed to “help with planning a strategy to ensure research data produced in UK Higher Education Institutions are preserved and remains accessible in the long term.” (Digital Curation Centre and University of Glasgow, 2011**c**).
3. Wider organisational planning using the CARDIO tool (Digital Curation Centre, 2011**a**) which is used to assess institutional requirements and capacity of the organisation to manage data; build consensus between parties and different departments; identify goals, inefficiencies, opportunities, and to help make case for investment in RDM infrastructure.

Finally, in the HE library sector there are two groups supporting librarians in engineering RDM. The UK University Science & Technology Librarians Group (USTLG, 2011), and the Engineering Libraries Division of ASEE (Cook, 2012): both have a remit to help science/engineering librarians to develop RDM skills.

It is recommended that Orbital review the available DCC and JISC training materials, and use them to design a comprehensive RDM training programme for researchers and support staff at the University of Lincoln, consisting of workshops as well as printed/online guides – also that Data Management Planning (DMP) Online tools are incorporated within the Orbital application (using the DMP Online APIs as soon as they are available) to support researchers in meeting their funders’ requirements for plans at the grant application stage.

3. Conclusion

In light of the discussions above, the initial objectives of the Orbital project were appropriate and on the mark as far as the scope of the researcher and the project itself is concerned, but that those same objectives indicate a broad area of institutional responsibility for RDM that goes beyond scholarly communication to affect strategic areas such as recruitment and training, business intelligence and continuity, IP and income generation, as well as future curriculum design and corresponding investment in infrastructure and estate.

This is no small task. The nine recommendations made in section 4, below, will be taken forward by Orbital within the scope of the JISC-funded project. However, they are long-term commitments, and if it is to benefit from research data re-use and publication, the University of Lincoln will have to adapt certain functions to sustain an RDM infrastructure beyond the life of the Orbital project. A business case for that sustained maintenance will be made later in the project.

As described in the introduction to this document, there is a large body of published material relating to Research Data Management; and this despite the relative youth of RDM as a discrete discipline in HE information management.

There is such a wealth of good-quality material, training guides, tools and questionnaires that HE staff new to RDM can easily be overwhelmed by the amount of advice and support on offer – despite that, some of the terminology is not used outside RDM and the concepts discussed can feel obscure. Finally, this is a fast-changing area, and new material is being generated literally daily by projects, institutions, and funders. For those reasons, this literature review should be maintained as a 'live' document, reviewed and updated regularly.

4. List of recommendations for the University of Lincoln

It is recommended that:

1. The Orbital project continue its work to assess the storage and other requirements of University of Lincoln researchers using surveys and interviews, and make its findings available to the University to assist RDM planning.
2. The Orbital project continue to work with Dr. Darlington and the REDm-MED project team on implementing the findings of ERIM into Orbital.
3. Orbital develop on its initial use of surveys, interviews, and close working alongside researchers to understand their requirements and existing RDM ‘culture’ at Lincoln. It should work to develop advocacy/training materials that speak to what motivates researchers, as well as highlighting obligations to their institution, funders, and to the ‘public good’. The University of Lincoln may also want to consider future job roles (“data librarians” or similar) that would have a positive effect on a culture of RDM.
4. In the short term, Orbital must support the University's response to the EPSRC data policy roadmap (due 1st May, 2012). Longer term, Orbital (both the project and the RDM application it will produce) can support the University of Lincoln in developing and implementing institutional policies that complement its obligations to funders and its statutory requirements.
5. The Orbital team should continue the initial explorations into Open Access, Open Data, and Open Source models that are appropriate for the University, and work with Research & Enterprise to formulate clear policies on data sharing and licensing alongside a business case for Open Access to data, and clear guidance to researchers about their rights, obligations, and the benefits of data sharing.
6. Guidance and advocacy materials drawn up as part of the Orbital project concentrate on articulating the *academic* benefits of RDM to researchers, and that institutional policies, along with support materials, develop the practice of reflecting appropriate RDM costs into future research funding bids, in consultation with Research & Enterprise, commercial partners, and funding bodies. It is also recommended that the University develop appropriate plans for meeting the costs of long-term storage, preservation and curation of research data.
7. Orbital incorporate the functionality of DataCite via its API, into the Orbital application, to allow Lincoln researchers to secure a DOI (Digital Object Identifier) for their data objects. This more than almost any other feature will secure the ability of the University of Lincoln and its staff to publish and share datasets with confidence. It is also recommended that the Library be engaged specifically to consider the curation and description of assets held in data repositories with reference to appropriate curation standards, as a natural extension of its (the Library’s) support for the Lincoln (publications) Repository, and to offer training and support for other University staff.
8. Orbital continue to work with Siemens, the School of Engineering, the University of Bath and the DCC to develop institutional expertise in handling engineering data formats.
9. Orbital review the available DCC and JISC training materials, and use them to design a comprehensive RDM training programme for researchers and support staff at the University of Lincoln, consisting of workshops as well as printed/online guides – also that Data Management Planning (DMP) Online tools are incorporated within the Orbital application (using the DMP Online APIs as soon as they are available) to support researchers in meeting their funders’ requirements for plans at the grant application stage.

5. Bibliography

N.B. an updated list of sources is available in an online RefShare database at: http://lncn.eu/bcf6

Alexogiannopoulos, E., McKenney, S., and Pickton, M. (2010) *Research Data Management Project: a DAF investigation of research data management practices at The University of Northampton* [Online]. Northampton: University of Northampton. Available from: http://nectar.northampton.ac.uk/2736/ [Accessed 4 October 2011].

Allan, R. (2011) *Management and Analysis of Large Research Data Sets* [Online]. Warrington: Science and Technology Facilities Council; Daresbury Laboratory. Available from: http://www.grids.ac.uk/NWGrid/LargeData/ [Accessed 11 February 2012].

Allinson, J. (2006) *OAIS as a reference model for repositories: an evaluation*. Bath: UKOLN, University of Bath (Report number: Drs-OAIS-evaluation-0.5).

Australian National Data Service (2011) *Data Management for Researchers* [Online]. Clayton, Victoria: ANDS. Available from: http://www.ands.org.au/researchers/manage-data.html [Accessed 11 October 2011].

Ball, A. (2006) *Briefing paper: the OAIS Reference Model*. Bath: UKOLN, University of Bath.

Ball, A. (2010**a**) *Review of the state of the art of the digital curation of research data* [Online]. Bath: University of Bath (Report number: erim1rep091103ab12). Available from: http://opus.bath.ac.uk/19022/ [Accessed 11 October 2011].

Ball, A., Darlington, M., Howard, T., McMahon, C., and Culley, S. (2010**b**) *Engineering Research Data Management Plan Requirement Specification*. Bath: University of Bath (Report number: erim6rep100901ab11).

Beagrie, N. (2011**a**) *Benefits from the Infrastructure Projects in the JISC Managing Research Data Programme* [Online]. Bristol: JISC. Available from: http://www.jisc.ac.uk/whatwedo/programmes/mrd/outputs/benefitsreport.aspx [Accessed 11 October 2011].

Beagrie, N. [and others] (2011**b**) *Keeping Research Data Safe (KRDS): cost/benefit studies, tools, and methodologies focusing on long-lived data* [Online]. Salisbury: Charles Beagrie. Available from: http://beagrie.com/krds.php [Accessed 20 December 2011].

Consultative Committee for Space Data Systems (2009) *Reference Model for an Open Archival Information System (OAIS)* [Online]. Washington, DC: CCSDS Secretariat (Report number: CCSDS 650.0-P-1.1). Available from: http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/CCSDSAgency.aspx [Accessed 4 October 2011].

Cook, J. (2012) *Engineering Libraries Division of ASEE* [Online]. Washington, DC: American Society for Engineering Education. Available from: http://depts.washington.edu/englib/eld/ [Accessed 12 February 2012].

Darlington, M., Ball, A., Howard, T., Culley, S., and McMahon, C. (2010) *Principles for Engineering Research Data Management*. Bath: University of Bath (Report number: erim6rep101028mjd10).

Darlington, M. (2012**a**) Managing Engineering Research Data at the University of Lincoln: Part 1 of 2 [Presentation]. Available from: http://www.slideshare.net/ulcerd/erim-redmmed-and-orbital [Accessed 30 January 2012].

Darlington, M, (2011) *ERIM Terminology V4* [Online]. Bath: University of Bath. http://wiki.bath.ac.uk/display/ERIMterminology/ERIM+Terminology+V4 [Accessed 30 January 2012]

Darlington, M. (2012**b**) *Managing Engineering Research Data at the University of Lincoln: Part 2 of 2* [Presentation]. Available from: http://www.slideshare.net/ulcerd/erim-redmmed-and-orbital-ii [Accessed 30 January 2012].

Darlington, M., Ball, A., and Thangarajah, U. (2012**c**) *Research Data Management for Mechanical Engineering Departments (REDm-MED) Project* [Online]. Bath: University of Bath. Available from: http://www.ukoln.ac.uk/projects/redm-med/ [Accessed 20 January 2012].

Science (2011). Dealing with data. *Science*, (11 February) [Online]. Available from: http://www.sciencemag.org/site/special/data/ [Accessed 20 December 2011].

Digital Curation Centre (2010**a**) *Funders' data plan requirements* [Online]. Edinburgh: DCC. Available from: http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements [Accessed 11 October 2011].

Digital Curation Centre (2010**b**) *How-to Guides* [Online]. Edinburgh: DCC. Available from: http://www.dcc.ac.uk/resources/how-guides [Accessed 11 October 2011].

Digital Curation Centre (2010**c**) *DCC Curation Lifecycle Model* [Online]. Edinburgh: DCC. Available from: http://www.dcc.ac.uk/resources/curation-lifecycle-model [Accessed 11 October 2011].

Digital Curation Centre (2010**d**) Digital Curation Training for All [Online]. Edinburgh: DCC. Available from: http://www.dcc.ac.uk/training [Accessed 11 October 2011].

Digital Curation Centre (2011**a**) *CARDIO: Collaborative Assessment of Research Data Infrastructure and Objectives* [Online]. Edinburgh: DCC. Available from: http://cardio.dcc.ac.uk/ [Accessed 20 December 2011].

Digital Curation Centre (2011**b**) *DMP Online: the DCC Data Management Planning Tool* [Online]. Edinburgh: DCC. Available from: https://dmponline.dcc.ac.uk/ [Accessed 20 December 2011].

Digital Curation Centre and University of Glasgow (2011**c**) *Data Asset Framework: four steps to effective data management* [Online]. Edinburgh: DCC. Available from: http://www.data-audit.eu/ [Accessed 20 December 2011].

Engineering and Physical Sciences Research Council (2011**a**) *EPSRC Policy Framework on Research Data* [Online]. Swindon: EPSRC. Available from: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx [Accessed 20 January 2012].

Engineering and Physical Sciences Research Council (2011**b**) *Impact, Timescales and Support* [Online]. Swindon: EPSRC. Available from: http://www.epsrc.ac.uk/about/standards/researchdata/Pages/impact.aspx [Accessed 12 February 2012].

Gray, N., Carozzi, T., and Woan, G. (2011) *Managing Research Data – Gravitational Waves Final Report*. Glasgow: University of Glasgow (Report number: LIGO-P1000188-v10).

Hitchcock, S. (2011) *Setting institutional repositories on the path to digital preservation: Final report from the JISC KeepIt Project*. Unpublished.

Hodson, S. and JISC (2011**a**) *Managing Research Data Programme 2011-13* [Online]. Bristol: HEFCE, on behalf of JISC. Available from: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx [Accessed 1 February 2012].

Hodson, S. and JISC (ed.) (2011**b**) Managing Research Data Launch Meeting, National College for School Leadership, Nottingham, 1-2 December 2011 [Online]. London: HEFCE, on behalf of JISC. Available from: http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata/mrdlaunch.aspx [Accessed 20 December 2011]

Hodson, S. and JISC (2012) *JISC Managing Research Data Programme: blog for the JISCMRD Programme* [Online]. London: JISC. Available from: http://researchdata.jiscinvolve.org/ [Accessed 15 February 2012].

Howard, T., Darlington, M., Ball, A., Culley, S., and McMahon, C. (2010) *Understanding and Characterizing Engineering Research Data for its Better Management*. Bath: University of Bath (Report number: erim2rep100420mjd10).

JISC (2011**a**) *JISC Managing Research Data Programme 2009-11 Outputs* [Online]. Bristol: JISC. Available from: http://www.jisc.ac.uk/whatwedo/programmes/mrd/outputs.aspx [Accessed 11 October 2011].

JISC (2011**b**) *Research data management training materials (RDMTrain)* [Online]. Bath: HEFCE, on behalf of JISC. Available from: http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmtrain.aspx [Accessed 20 December 2011].

JISC, Research Information Network, and Technopolis Ltd (2011**c**) *Data centres: their use, value and impact* [Online]. Bristol: JISC. Available from: http://www.jisc.ac.uk/publications/generalpublications/2011/09/datacentres.aspx [Accessed 4 October 2011].

Lavoie, B.F. (2004) *The Open Archival Information System Reference Model: Introductory Guide*. Dublin, OH: OCLC Online Computer Library Center, Inc. and Digital Preservation Coalition (Report number: 04-01 04-01).

Lyon, L. (2007) *Dealing with Data: Roles, Rights, Responsibilities and Relationships*. Bath: UKOLN, University of Bath.

O'Brien, R. (2011) Opening Up. *JISC Inform*, (32) [Online]. Available from: http://www.jisc.ac.uk/inform/inform32/ [Accessed 12 February 2012].

Patrick, M. ([n.d.]) *Data management bibliography* [Online]. Oxford: University of Oxford. Available from: http://sudamih.oucs.ox.ac.uk/docs/Data%20management%20bibliography.pdf [Accessed 11 October 2011].

Research Councils UK ([n.d.]) *RCUK Common Principles on Data Policy* [Online]. Swindon: RCUK. Available from: http://www.rcuk.ac.uk/research/Pages/home.aspx [Accessed 4 October 2011].

Tsinoremas, N.F., Zysman, J., Mader, C., and Blaire, J. (2011) Data in Motion: a new paradigm in Research Data Lifecycle Management. **In**: *Princeton Workshop, 18-20 July 2011* [Online]. Available from: http://www.columbia.edu/~rb2568/rdlm/Tsinoremas_UMiami_RDLM2011.pdf [Accessed 20 December 2011].

UK Data Archive Data Management ([n.d.]) *Planning for ESRC Research Data-rich Investments project (DMP-ESRC)* [Online]. Colchester: UK Data Archive, University of Essex. Available from: http://www.data-archive.ac.uk/create-manage/projects/JISC-DMP [Accessed 4 October 2011].

UK Data Archive, van den Eynden, V., Corti, L., Woollard, M., Bishop, L., and Horton, L. (2011) *Managing and sharing data: best practice for researchers* [Online]. 3rd edition. Colchester: UK Data Archive, University of Essex. Available from: http://www.data-archive.ac.uk/media/2894/managingsharing.pdf [Accessed 4 October 2011].

UK Research Integrity Office (2009) *Code of Practice for Research: promoting good practice and preventing misconduct* [Online]. Falmer: UKRIO. Available from: http://www.ukrio.org/what-we-do/code-of-practice-for-research/ [Accessed 19 November 2011].

University of Bath, Innovative Design and Manufacturing Research Centre (I*d*MRC) ([n.d.]) *The Engineering Research Information Management (ERIM) Project* [Online]. Bath: UKOLN, University of Bath. Available from: http://www.bath.ac.uk/idmrc/erim/ [Accessed 4 October 2011].

University of Cambridge ([n.d.]) *Support for Managing Research Data* [Online]. Cambridge: University of Cambridge. Available from: http://www.lib.cam.ac.uk/dataman/ [Accessed 20 December 2011].

University of Cambridge (2010) *Incremental project* [Online]. Cambridge: University of Cambridge. Available from: http://www.lib.cam.ac.uk/preservation/incremental/ [Accessed 20 December 2011].

University of Edinburgh (2011) *Research data management guidance* [Online]. Edinburgh: University of Edinburgh. Available from: http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt [Accessed 4 October 2011].

University of Glasgow ([n.d.]**a** *Data management support for researchers* [Online]. Glasgow: University of Glasgow. Available from: http://www.gla.ac.uk/services/datamanagement/ [Accessed 20 December 2011].

University of Oxford ([n.d.]) *Research Data Management* [Online]. Oxford: University of Oxford. Available from: http://www.admin.ox.ac.uk/rdm/ [Accessed 4 October 2011].

University Science and Technology Librarians Group (2011) *USTLG* [Online]. Available from:

http://ustlg.org/ [Accessed 20 December 2011].

Valentine, M., Macdonald, S., Donnelly, A., Rice, R., and Ekmekcioglu, C. (2011) *Research Data Mantra project wiki* [Online]. Edinburgh: University of Edinburgh. Available from: https://www.wiki.ed.ac.uk/pages/viewpage.action?pageId=119345088 [Accessed 11 October 2011].

Wellcome Trust (2010) *Policy on data management and sharing* [Online]. London: Wellcome Trust. Available from: http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm [Accessed 9 December 2011].