DATA MANAGEMENT PLANNING IN ENGINEERING DESIGN AND MANUFACTURING RESEARCH

Mansur Darlington[1], Tom Howard[2], Alex Ball[3], Steve Culley[1] and Chris McMahon[1]
(1) I*d*MRC, University of Bath, UK (2) Dept. of Manufacturing Engineering, DtU, Denmark (3) UKOLN, University of Bath, UK

ABSTRACT
There is a growing interest in maximizing the value of research data through their sharing and re-use. This desire is hampered by the prevailing culture in data management during the research activity which largely ignores the potential for re-use, and by a lack of understanding of the character of research data and of the barriers to and opportunities for their re-use. The work reported here characterizes engineering design and manufacture research data and explores the context of their development and current management. Insights from the work have resulted in the authors proposing a number of new approaches and tools which provide the basis for better management practice.

Keywords: data management; information management, engineering design, manufacturing

1. INTRODUCTION
Motivated by the general drive toward efficiency and accountability, there is a developing interest in the ways that the data gathered and generated in research might be made more available for use, re-use and re-purposing [1], [11], [3], [4]. This has become of particular interest as the generation and use of digital data has become dominant.
The work reported here attempts to identify and characterize in particular the spectrum of research data that is characteristically collected and generated during research activities associated with engineering design and manufacture. The purpose is to provide a sounder basis for the management of such data so that they can be made more amenable to use, re-use and re-purposing (see definitions in Section 4) by which means the value of the data is increased. In particular the work aims to provide means for better data management planning (DMP).
From greater sharing of data would be derived a number of benefits including the reduction in duplicated work, greater transparency of research, an improved basis for validation and obviating the need for collection and generation [5], [9].
The research work and findings reported here concern engineering research carried out within the academic domain; yet it would be difficult to argue that engineering research carried out within industry, and the research data so generated, would be characteristically different in any particular. Furthermore, it is not merely from research that data is generated. Much of the complex and interactive activities in the practice of engineering will produce data in a similar manner and of a sort similar to that found in research. Thus it is the case that improved methodologies for the management of diverse research data derived from and collected from complex and interactive research activities – irrespective of the environment in which the research occurs – may be directly applicable to the management of similar non-research data sets which are the products of the engineering process.
The authors' research consists of three main parts. The first part consists of the identification and characterization of engineering research data through the inspection of data generated and collected in a representative sample of engineering design and manufacturing research. The second part of the research is theoretical, and is the basis for the analysis of the subject research data in respect of their management. Understanding how to manage research data requires not only that the nature of the data is understood, but so too the context in which the data are collected, gathered, manipulated and used. The third part of the research consists of the early manifestations of a number of new approaches, derived from the two other research elements, which support not just data management but also data management planning.
The focus of the research is research data associated with engineering design and manufacture. Because the diversity of data found is very great it can be said that the findings of the research may be

usefully applied more generally, that is to any data generated through complex activities such as those found widely in commercial enterprises and in design and engineering practice.

2. BACKGROUND

The quantity of digital data in existence is growing at an explosive rate. In some areas of science – such as particle physics, astronomy and bioinformatics – the amount of data being generated has increased by a factor between a hundred and a thousand over the past decade [12]. This increase poses numerous challenges for the researcher, not only in terms of storing, backing up and preserving one's own output, but also when it comes to discovering relevant, high-quality data amongst that which is available, and in the case of large datasets, gaining access to them over crowded networks.

Having a sizeable pool of shared data on which to draw provides the researcher with many opportunities, not least that it can allow established forms of research to be conducted with greater efficiency and speed: for example, researchers working with the Sage Federation's pilot data commons managed to publish three papers in *Nature* after only four months [13]. By connecting up with existing data, researchers can extend the baseline of their studies through time, space and human context; hence, for instance, the efforts to plug gaps in the climate data record with meteorological information from ships' logs [14], [15]. Most exciting is the opportunity that data afford for new forms of research, whether this is mining vast collections of homogeneous data for emerging patterns, as is the case in some bioinformatics research [16], or combining heterogeneous data to answer new research questions, such as the effect of environmental conditions on species traits and populations [17].

Such instances of re-use and repurposing are only possible if the data have been properly and carefully managed throughout the course of the research and beyond. They need to be generated and collected using common, well-respected methods. They need to be recorded in a well-supported format, using a data model congruent with those in widespread use within the discipline. They need to be fully documented so that the methodologies, data models, structures and semantics associated with the data are clear to potential re-users. They need to be properly packaged, so that none of them are misplaced or otherwise missed during batch actions (transfers, fixity checks, etc.). They need to have an unambiguous status with regards intellectual property rights and licences, confidentiality and privacy, so that whatever can be shared, can be shared with confidence. They need to be appropriately described, and the descriptions available to harvest by data search services, so that potential re-users can find the data easily. Finally, the data and their associated metadata need to be stored securely, preserved at the bit level, and adapted over time in response to changes in available software, data models and semantics..

These activities cannot be left to chance, not least because some of them require effort substantial enough to require dedicated funding. To ensure that the data are managed carefully and properly, their management must be planned and budgeted for right from the start, in accordance with institutional/departmental policies and best practice. The formal expression of such planning is termed a Data Management Plan (DMP), which should be created for each project. The importance of doing this is becoming widely recognized: at least five major research funding bodies in the UK (the AHRC, BBSRC, ESRC, MRC, and Wellcome Trust) require a DMP to be submitted as part of every grant application [18], as do two major research funders in the US (the NIH [19] and NSF [20]). This is linked to an increased appreciation of the benefits of data sharing. In addition to the opportunities outlined above, data sharing also permits greater scrutiny of the research process (driving up standards of quality), increases efficiency in terms of lower data collection costs and reduced duplication, and arguably increases the academic, economic and social impact of research [21], [22], [23]. For these reasons, some journals (e.g. *American Economic Review*, *Journal of Evolutionary Biology*, *Clinical Infectious Diseases*) are now mandating that the data underlying submitted papers are deposited either with the journal itself or with a recognized data depository.

While a consensus is forming around the need for DMPs, there is a much weaker consensus on what such plans should contain, at least at the point of bid submission. A comparison of funder requirements conducted by the UK's Digital Curation Centre revealed significant differences in expectations for both scope and level of detail [18]. Within UK engineering research, the uncertainty is particularly acute as the principle funding body, the Engineering and Physical Sciences Research Council, has not yet mandated DMPs and provides little guidance on what to include in one. It is in this context that the research presented here was conducted.

3. METHODOLOGY

Little is really known about the content and diversity of research data gathered and developed in the course of engineering design and manufacturing research. To improve the understanding of such data and their management needs, in order to improve management methods and tools, a representative set of research data was subjected to close scrutiny and characterization. The data were drawn from two repositories. One of these is the KIM Grand Challenge Project [7] 'virtual' repository and the other is the repository of research data held by the I*d*MRC at the University of Bath (http://www.bath.ac.uk/idmrc/).

The work was carried out in a number of stages, which included:

1. Clarifying terminology.
2. Identifying exemplar projects.
3. Identifying the main sorts of data represented in the exemplar projects, and their relationships and data development paths.
4. Interviewing key participants concerning opportunities and requirements from reuse and the issues and challenges that arise.
5. Preparation of a set of complementary support approaches and tools.

This paper reports in brief on these items and the outputs which emerged. A fuller account may be found in [10] and other documentation associated with the Engineering Research Information Management Project (ERIM Project).

4. TERMINOLOGY

The terms *use*, *re-use* and *re-purposing* have been introduced above. In order to be able to communicate about the data life cycle and the sub-activities and processes that go on during data generation, collection and manipulation, these terms, together with many others were found to require identification and definition. This motivated the adoption of a set of prescriptive terms, some, but not all, borrowed from existing work; the resulting terminology can be found in its current form in [10]. The key terms introduced above, together with a number of others of particular importance to this paper, are defined thus:

- **Data Use** Using research data for the current research purpose/activity to infer new knowledge about the research subject.
- **Data Re-use** Using research data for a research purpose/activity other than that for which it was intended.
- **Data Purposing** Making research data available and fit for the *current research activity*.
- **Data Re-purposing** Making existing research data available and fit for a *future* ***known*** *research activity*.
- **Supporting Data Re-use** Managing existing research data such that it will be available for a *future* ***unknown*** *research activity*. Unfortunately, it has not been possible to coin an apposite verb for this activity; however it seems likely that 'supporting data re-use' consists of some or all of the activities associated with archiving, preservation and curation.
- **Data** Reinterpretable representations of information in a formalized manner suitable for communication, interpretation or processing.
- **Data Object** Either a physical object or a digital object containing data.
- **Data Record** A data object created, received and maintained as evidence of an activity.
- **Data Case** The set of data records associated with some discrete research activity (project, task, experiment, etc.).
- **Information.** Any type of knowledge that can be exchanged. In an exchange, it is represented by data.

In addition to these, detailed scrutiny of the way in which data is worked on during the management activities of data purposing, supporting data re-use and data re-purposing resulted in the identification of a non-exhaustive set of data 'development' activities to which data and data records are subjected. The set (defined in the terminology) includes: Addition, Association, Aggregation, Annotation, Augmentation, Collection, Collation, Deletion, Derivation, Duplication, Extraction, Generation, Migration, Population and Refinement. An understanding of the development activities and their relationships and their association with the management activities of data purposing, supporting data-

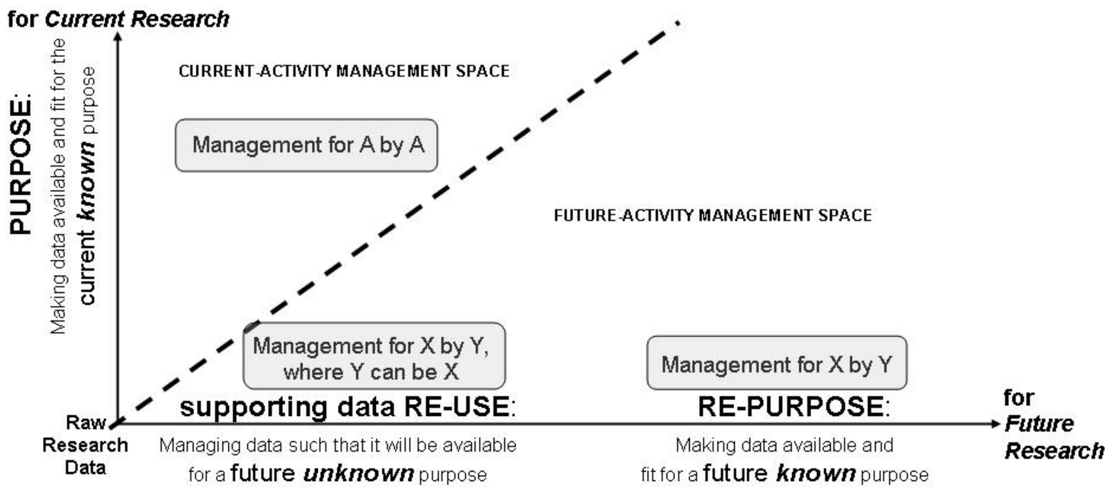

Figure 1. The three data management activities and their relationship providing a management-centric view of data

reuse and data re-purposing provides the basis for a management-centric view of data development as shown in Figure 1. It is important to see that, whilst the above considerations were derived from engineering design and manufacturing research data *per se*, they can equally be applied in relation to engineering design and management data and data records themselves in the commercial context.

5. RESEARCH PROJECT STUDY

As mentioned above, two repositories of research project data were selected for inspection. The first of these repositories is constituted of the data assets held by researchers in the Innovative Design & Manufacturing Research Centre (I*d*MRC) located in the University of Bath (http://www.bath.ac.uk/idmrc/). As the name of the Centre suggests, the focus of interest is in engineering design and manufacturing research. The second repository consists of the data assets of the KIM Project (http://www.kimproject.org) which concerned research associated with understanding the information and knowledge needs of product-service provision associated with a large-scale, long-term capital projects such as the design and manufacture of airports, aircraft, naval vessels and so on.
From the wealth of data assets available, twelve cases (7 from the IdMRC and 5 from the KIM Project) were selected from the engineering design and manufacturing domain which the authors believed displayed a good mix of research topics, methods, data size and data format; the projects are identified here:

1. Airframe Stress Data Reuse
2. Snow Mobile Design Activity Observation
3. Aerospace Cost Forecasting
4. Large-Scale Metrology Shared Resources
5. Form-fill-feed Packaging Modelling
6. CNC Machine Measurement
7. Cryogenic Machining
8. Information Management Tool
9. Knowledge Enhanced Notes (for the recording and management of design review meetings)
10. Service Design Research
11. Design Activity & Knowledge Capture Research
12. Understanding the Learning Organization

A scoping survey was carried out on the twelve asset sets or Data Cases in order to understand how to characterize research data and to understand the context in which they were generated. For each of the above selected projects a semi-structured interview was carried out with the principal researcher associated with the project. This interview had two stages, first to identify individual data records associated with the project then to assign attributes to each of the data records. From this it was possible to refine the data and with the aim of selecting a further, reduced set for a, later, full data audit (see below).

6. MODELLING

It became apparent as a result of the scoping survey that the context in which data were generated, collected and manipulated was of prime importance to their later interpretation and understanding and that recording of the context was a necessary precursor to the later re-use and re-purposing of data. It follows that recording of context must become part of the management of data during their generation and first use. The scoping study and development of the theoretical aspects of the research (including the terminology) suggested that the relationships between data could be mapped in principle not only at the data level, but at the data record and data case level.

In Figure 2 what the authors refer to as *horizontal* and *vertical* relationships can be observed. In the figure a distinction is made between data objects which are Research Data Records (RDRs) and those which contain contextualizing information which supports interpretation and understanding of the research data, that is Contextualizing Data Records (CDRs). In fact a number of distinct data record types were identified of which these are chiefly representative.

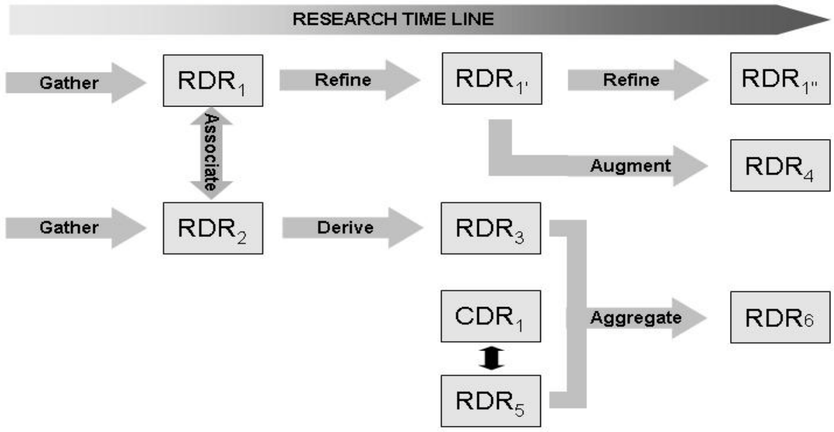

Figure 2. An illustration of some horizontal and vertical relationships between data records occurring as a result of data development activities being carried out at the data level

The horizontal sort of relationship occurs as a result of specific and appropriate processes being carried out on data at the data level, resulting in data change or generation. Characteristically these processes include such things as data *refinement*¸ data *derivation* and data *aggregation.* Sometimes these activities result in new RDRs, sometimes in changes which occur within an existing RDR (for example, the addition of a chart or a new page to an existing spreadsheet).

The vertical relationship is the sort of relationship which is either inherent or developed as a result of specific and appropriate management sub-activities being carried out at the data record level. Characteristically these activities include such things as *generation, collection, association.* These relationships can exist between not only RDRs but also between CDRs and between CDRs and RDRs. The contextualizing data record shown here serves the purpose of enhancing in some way the understanding of the content of the associated RDR. Where these vertical relations are implicit only, good management practice would require explicit *association* by some mechanism. For example, in a research project it is often the case that a number of similar experiments will be made, each one of which will have a 'case' of records in which research data is contained. Such collections of cases will, too, be the subject of management, and will be of greatest value only whilst their association at the 'super'-case level is recorded and perpetuated.

Analysis of the data records and the development of a means of classifying such records into types and assigning other useful attributes allowed development of a formal means of modelling of what the authors refer to as Research Activity Information Development (RAID), from which the visualization know as the RAID diagram has emerged. The RAID diagram, which uses UML as a representational language, provides a visual mapping of the data records associated with a research activity (at any chosen level of activity granularity), showing the chief attributes of each data record and the relationship between it and other records in the data case.

As can be seen in Figure 5 in one view can be shown a useful contextualizing mapping consisting of attributional, relational and temporal information. Thus, the RAID modelling approach becomes a means of managing data during their development so as to support their later interpretation and understanding for re-use and re-purposing.

7. DATA AUDITS

The scoping survey provided the basis for understanding how research data might be usefully characterized for the purposes of their better management. In addition a terminology was developed together with the modelling formalism introduced above. The set of data assets used in the scoping survey were refined into a sub-set of five data cases for closer inspection using a binary selection method acting on important attributes by which the data had been characterized. The binary decision path, which aimed to maximize diversity in the research data to be audited, is shown in Figure 3.

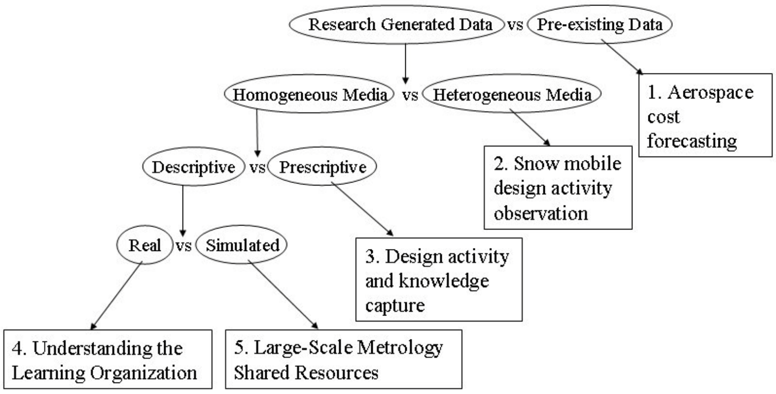

Figure 3. The selection of cases for full audit

Following selection the researcher involved in each research project from which the data case was selected was interviewed again, this time principally to capture the research activity information development (RAID) for the data case, but also to gain a better understanding of the data following the preliminary enquiry in the scoping study and to understand the researcher's views on the barrier to and opportunities for re-use.

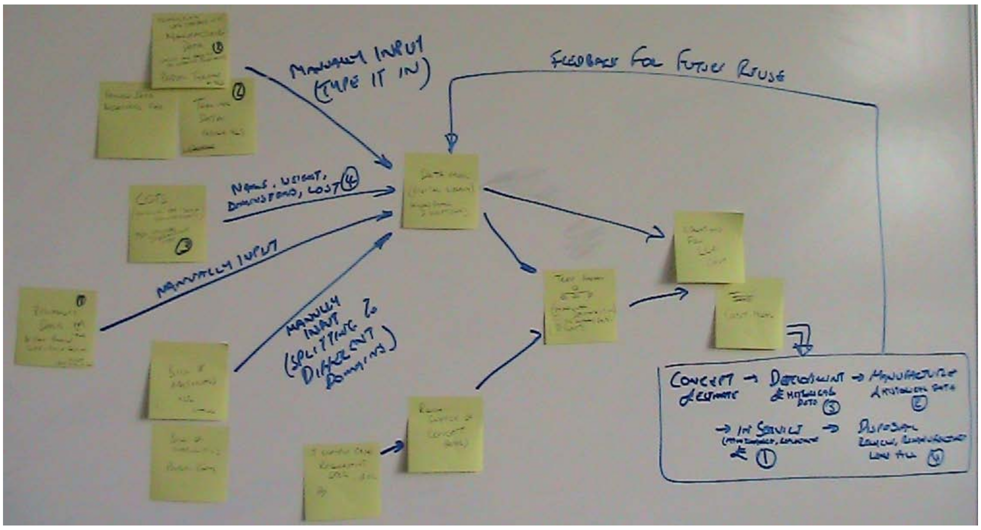

Figure 4. An example of RAID capture at the point of construction

Figure 4 shows the method of capture and Figure 5 a resulting RAID diagram for an example case. The data related to the research associated with the design, implementation and development of a means of cryogenic machining of soft polymers.

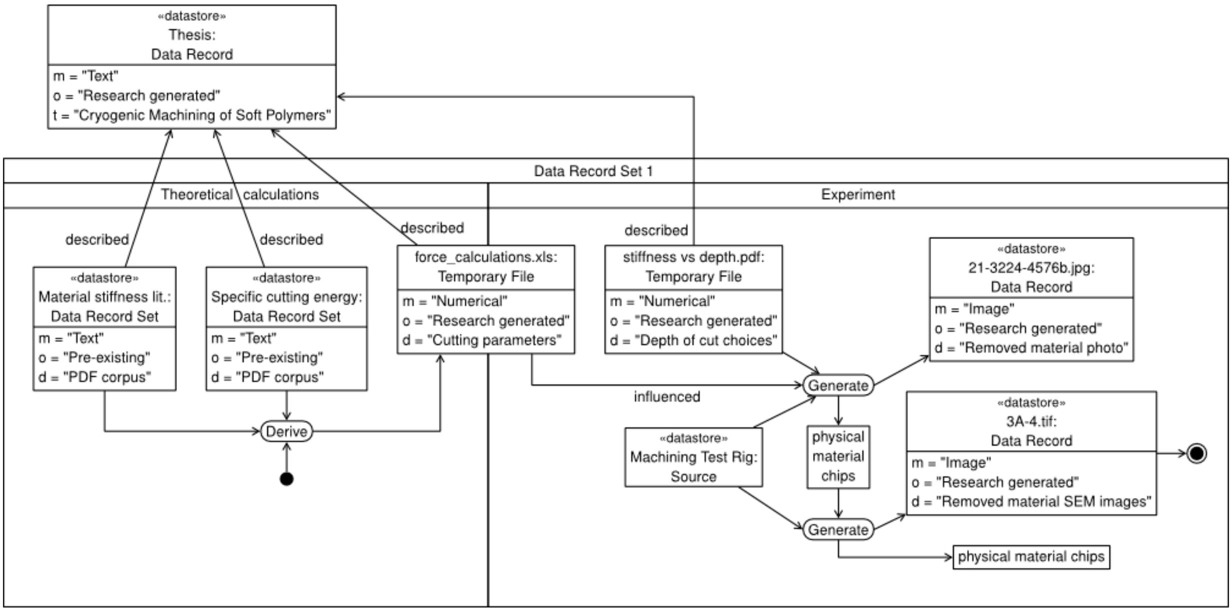

Figure 5. A RAID diagram of CRYMAN Project Data Records

8. ISSUES AND CHALLENGES IN RESEARCH DATA MANAGEMENT

The research which has been reported in brief above extended the understanding of the breadth and width of engineering research data, provided an improved means by which the subject could be communicated and discussed, and provided insights into barriers to and opportunities for data sharing and re-use. The following constitute the main findings and insights from the work:

1. **Great diversity of data type and quality.** The data found in engineering design and manufacturing research are very diverse in their nature both in terms of content and in representation: one end of the spectrum being represented by, say, numerical point cloud data and the other by text and sketch field notes, or more diversely still audio and video data. Indeed, it is difficult to imagine any form of medium or representation which might not reasonably be expected to be encountered. Similarly, there is a great diversity of quality, this being dictated often by what is available or the (non-uniform) needs of the research. In spite of this diversity, management techniques must be found that can be applied in general.
2. **Complex and chaotic nature of data development.** As can be seen by inspection of even a fairly simple RAID diagram, the development of data is often very chaotic. In research, as in many other knowledge-based activities, the recording and development of information, and the forms it takes can be both diverse and myriad, often dictated by the character and direction of the research. At the same time, the explicit organization of the data within a data case when it exists at all, is often partial or idiosyncratic.
3. **Outputs not linked to data.** Hitherto, largely dictated by what is possible, it has not been a requirement or expected that data supporting reports of research findings should be made available with the report nor, indeed, that the data should necessarily be made available or accessible for scrutiny at all. The dominance of electronic data and new means for linking information objects is encouraging a change from the current situation to one where the provision of supporting data will become more customary.
4. **Supporting documents not situated with the data files.** Often good data are made meaningful only by information that is external to the data themselves. Currently there is neither the culture nor the necessary support to ensure that such information is associated with the data for which it provides context and by which those data are understood and interpreted.
5. **Little use of metadata to support future use.** A powerful way of providing additional information about data which helps future discovery, access and interpretation is through the use of metadata, either attached locally within the file which contains the data it explains, or remotely in ‘stand-off’ documents.

6. **Immature understanding of benefits of sharing and thus need for management.** The benefits of sharing data or making them available for future use through the preparation activities of re-purposing and 'supporting re-use' are poorly articulated, not least to the users and re-users of data. At the same time a lack of understanding promotes a situation where researchers themselves express the desire to share in only a limited way.
7. **Limited understanding of the barriers to or opportunities for information sharing and re-use.** Little work has been done to understand why data owners are reluctant to share data and what real and imaginary obstacles exist which impedes their doing so. Similarly, a widespread clear understanding of the opportunities of sharing and re-use is missing.
8. **Low investment in management support or activity.** A widespread understanding of the benefits of sharing referred to above is required not only by those who generate data (as in 6 above) but also those who are responsible for implementing strategies for data management. Until such understanding is more generally shared and a culture developed where good data management practice becomes a matter of course, proper investment in data management for future use will be neither requested nor expected.

In the above, items 1, 4 and 5 are supported by findings from the data audit, items 2, 3, 5 and 8 from the analysis of the RAID diagrams, and items 4, 5, 6, 7, and 8 from the interviews with the researchers responsible for the data.

9. RESEARCH DATA MANAGEMENT PLANS FOR ENGINEERING DESIGN

It is clear from the research work done and the insights listed above that there is much to be done to improve the management of research data to enhance the opportunities for sharing and re-use. The work has resulted in proposals for a number of remedies to shortcomings that the authors found in current management in respect of the wide variety of data encountered in engineering design and manufacture research. In particular it has become clear that good data management is dependent on good data management planning practices and on methods and tools which support such management. Currently there are few.

The RAID method of modelling the development of data has been introduced above. The authors believe that such a method, implemented as a data development mapping tool, would provide a useful way to provide an integrating contextual record of the data assets in a data case. Equally, because this approach is a general one – not reliant on any special properties of research data, but of data in general – its lends itself to a wide application.

As reported in Howard, et al. [10] one of the greatest impediments to re-using data is the lack of support for recording and maintaining context amongst the diverse digital and paper records which constitute the research data case(s) developed during the course of research activity. It is argued that without providing some sort of integrating contextualization the understanding necessary for data re-use cannot be guaranteed. In order to provide a good basis for interpreting data, it is necessary to understand the rôles and data content of records and the relationship between data and between records. Such context in which data can be understood is particularly necessary after the event, when the observer of the data has little or no personal knowledge of the circumstances surrounding their original collection or generation. A mapping of the sort provide by the RAID approach would provide a better basis than exists at present for sharing and re-use, through improved understanding and interpretation of the data. A high-level specification for such a tool, based on the RAID method, is part of the output of this work.

Good management requires also support in the form of best practice guidance. To this end, the authors have proposed and developed a number of guidance tools at different levels of activity. The most high-level of these is a proposed set of Principles for Engineering Research Data Management [6] which identifies the fundamental foundations to such management and provides motivation for doing so. At a more detail level, informed by the Principles, an Engineering Research Data Management Plan Requirements Specification has been prepared [2]. From this, can be implemented data management plans for individual research activities.

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 repurposed, where applicable, and how data re-use has

been supported. 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 Librarians or Managers to construct a suitable DMP for the long-term care of the data.
The DMPs under this specification

- 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 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.

10. CONCLUSIONS

The research described in this report has contributed considerably to the greater understanding of the diversity and character of engineering design and manufacturing research data and provided the basis for improved methods of their management for re-purposing and re-use. Part of the work has been development of an emerging terminology to help describe the different sorts of such information and the different forms of development and management activity to which it is subjected to during research. In particular the notions of ‘data purposing’, ‘data re-purposing’ and ‘supporting data re-use’ have been identified as data development activities which motivate data management..

The concepts defined in the terminology provide the elements in a new means of modelling the research activity and the development of associated research and contextual data. These data are recorded in a number of types of data record identified by the authors. Scrutiny of the data, and investigation of the research by which they were gained, provides the basis for characterizing a broad spectrum of data, the development of which can be captured for each data case using the new modelling method, in a Research Activity Information Development (RAID) diagram.

In considering the different types of development process to which research data are commonly subjected, the authors believe that ‘association’ is perhaps foundational in supporting good data management for its easier re-use and re-purposing. To support the researcher in the better management of data as they are developed, and to provide a ‘map’ to aid their later understanding, the authors propose the use of an automated ‘association’ tool, based on the information that can be captured and represented in a RAID diagram. By capturing and recording the development of research data in individual activities or projects it is proposed that data assets can be made more findable, interpretable, verifiable, repeatable, replicable and useful. This approach is applicable not only to research data *per se*, but to data of any sort irrespective of the process by which they have been generated.

During the research work a number of barriers to re-use were identified. It is envisaged that these barriers could be tackled at a number of levels. Many of the barriers could have been removed by giving consideration to re-use during use, or prior to the gathering of the data. The authors envisage a progression of tools, each one more specializing than the last, which will assist in supporting better data management during the research activity. In particular, the provision of the means to achieve a well-documented context for the records which make up a data case might usher in a new era in which consideration of re-use of data to maximize their value becomes an expectation during their collection and generation for first use. In an attempt to encourage this trend the authors submit their own data for scrutiny, which can be found on the ERIM Project wiki.

ACKNOWLEDGEMENTS

The paper reports work carried out within the Engineering Research Information Management (ERIM) project funded by JISC through the Managing Research Data (JISCMRD) Programme.

REFERENCES

[1] Ball, A. Review of the State of the Art of the Digital Curation of Research Data. ERIM Project Document, 2010, http://opus.bath.ac.uk/18713/.

[2] Ball, A., Darlington, M., Howard, T., McMahon, C. & Culley, S. Engineering Research Data Management Plan Requirement Specification (version 0.2). ERIM Project Document erim6rep100901ab02, 2010, Bath, UK: University of Bath, http://opus.bath.ac.uk/21280/

[3] Beagrie, N., Beagrie and Rowlands, I. Research Data Preservation and Access: the views of researchers, Ariadne, 2009, Issue 60.

[4] Bechhofer, S., Ainsworth, J., Bhagat, J., Buchan, I., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Goble, C., Michaelides, D., Missier, P., Owen, S., Newman, D., De Roure, D. and Sufi, S. (2010) Why Linked Data is Not Enough for Scientists. In: Sixth IEEE e–Science conference (e-Science 2010), December 2010, Brisbane, Australia. (In Press) downloaded 20.12.2010 http://eprints.ecs.soton.ac.uk/id/eprint/21587

[5] Birnholtz, J., & Bietz, M. J. Data at work: Supporting sharing in science and engineering, Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work, 2003, (pp. 339-348). New York, NY: ACM Press. DOI: 10.1145/958160.958215.

[6] Darlington, M., Ball, A., Howard, T., McMahon, C. & Culley, S. *Principles for Engineering Research Data Management* (version 1.0)., 2010, Bath, UK: U. of Bath, http://opus.bath.ac.uk/22201/

[7] Darlington, M., Heisig, P., Leiringer, R. & Burt, G. *KIM Resources Inventory*, 2009. (KIM Project document: kim50too007gb10).

[8] [ERIM Project.] JISC Project web site: http://www.jisc.ac.uk/whatwedo/programmes/mrd/rdmp/erim, downloaded 6 January 2011.

[9] Fry, J., Lockyer, S., Oppenheim, C., Houghton, J. and Rasmussen, B. *Identifying Benefits Arising from the Curation and Open Sharing of Research Data* Produced by UK Higher Education and Research Institutes, 2008. Final report. JISC. URL: http://ie-repository.jisc.ac.uk/279/ (10/13/2009).

[10] Howard, T., Darlington, M., Ball, A., McMahon, C. and Culley, S. *Understanding and Characterizing Engineering Research Data for its Better Management* (ERIM Project Document erim2rep100420mjd10) 2010, Bath, UK: University of Bath. http://opus.bath.ac.uk/20896

[11] Jones, S., Ball, A. and Ekmekcioglu. C. The Data Audit Framework: A First Step in the Data Management Challenge, *International Journal of Digital Curation* [ISSN 1746-8256], 2008, 3:2, 112-120.

[12] Bell G., Hey T. and Szalay A. Beyond the data deluge. *Science*, 2009, 323, 1297-1298.

[13] Friend S. Open data for open science: physicists and software engineers do it, why can't biologists do it too? *6th International Digital Curation Conference*, Chicago, December 2010.

[14] Herrera, R.G., et al. *CLIWOC Final Report*, 2004 (Universidad Complutense de Madrid, Spain). URL: http://www.webcitation.org/5vYyXWQ0Q

[15] WWI ships to chart past climate. *BBC News*, 13 October 2010. URL: http://www.webcitation.org/5vYydYTR0

[16] Mons B., et al. Calling on a million minds for community annotation in WikiProteins. *Genome Biology*, 2008, 9(5), R89. DOI: 10.1186/gb-2008-9-5-r89

[17] Bechhofer S. The elevator pitch. *FISH.Link Project Blog*, 21 October 2010. URL: http://www.webcitation.org/5vYyid8eo

[18] Jones S. Summary of UK research funders' expectations for the content of data management and sharing plans, 2010 (Digital Curation Centre, Edinburgh). URL: http://www.webcitation.org/5vYyq1zYu

[19] *Final NIH Statement on Sharing Research Data*, February 2003 (National Institutes for Health, Bethesda, MD). URL: http://www.webcitation.org/5vYytnpmW

[20] *Proposal and Award Policies and Procedures Guide*, October 2010 (National Science Foundation, Arlington, VA). URL: http://www.webcitation.org/5vYyzhZeY

[21] SQW Consulting and LISU. *Open Access to Research Outputs*, 2008 (Research Councils UK, Swindon). URL: http://www.webcitation.org/5vYz2h03N

[22] Stodden V. Enabling reproducible research: open licensing for scientific innovation. *International Journal of Communications Law and Policy*, 2009, 13, 1-25.

[23] *Open to All? Case Studies of Openness in Research*, September 2010 (Research Information Network and National Endowment for Science, Technology and the Arts). URL: http://www.webcitation.org/5vYz6UqHu

Contact: Chris McMahon
University of Bath
Department of Mechanical Engineering
Bath, BA2 7AY, UK
Tel: Int +44 1225 384026
Fax: Int +44 1225 826928
Email: c.a.mcmahon@bath.ac.uk

Chris is Professor of Engineering Design in the Department of Mechanical Engineering at the University of Bath. He teaches and researches in engineering design and computer-aided design. He is interested in many aspects of design and computing, in particular how computer aids can assist designers in the organization and management of the information used in design.