Digital Curation Centre SCARP Case Studies ISSN 1729-586X

Curation of research data in the disciplines of Engineering

SCARP Case Study No. 7

Alex Ball

Colin Neilson

DCC SCARP INTERIM CASE STUDY REPORT
Deliverable B4.8.1.2

Version No.	1.1
Status	FINAL
Date	13 Jan 2010

Copyright

Catalogue Entry

Title Curation of research data in the disciplines of Engineering
Creator Alex Ball (author); Colin Neilson (author)
Subject Data curation; repository services; design disciplines; Engineering; disciplinarity; digital working
Description Case Study produced by the Digital Curation Centre (DCC) SCARP project, funded by the Joint Information Systems Committee (JISC).
Date 30 November 2009 (creation)
Type Text
Format Adobe Portable Document Format v.1.3
Resource Identifier ISSN 1759-586X
Language English
Rights © 2009 DCC, University of Bath

Citation Guidelines

Ball, A. and Neilson, C. (2010), "Curation of research data in the disciplines of Engineering", SCARP Case Study 7, Digital Curation Centre, Retrieved <date> from http://www.dcc.ac.uk/scarp

Contents

EXECUTIVE SUMMARY **5**

RECOMMENDATIONS **6**

1 INTRODUCTION AND BACKGROUND **7**

1.1 Digital objects in Engineering 7

2 ENGINEERING LANDSCAPE **9**

2.1 The character of Engineering as a discipline 9

2.2 Scope for this case study 9

2.3 How do practising engineers use information and data for research and design? 10

2.4 Information needs and information-seeking behaviour of engineers 11

2.5 Engineering Digital Repositories Landscape Analysis 13

2.6 Sharing Engineering Information 14

2.7 Disposition of research data 15

3 ENGINEERING RESEARCH DATA IN HIGHER EDUCATION **16**

3.1 Data Audit 16

3.1.1 Research Centre profile 16

3.1.2 Infrastructure profile 17

3.1.3 Data management policies 18

3.1.4 Profile of data holdings 18

3.2 Curation aspects 19

3.2.1 Research domain 19

3.2.2 Data creation 20

3.2.3 Working with the created data 21

3.2.4 Discovery, access and use 22

3.2.5 Curation practices 22

3.2.6 Sharing data 23

4 POLICY ON CARE OF ENGINEERING RESEARCH DATA **24**

5 CONCLUSIONS **26**

6 REFERENCES **28**

Executive Summary

This case study examines approaches to data deposit, sharing and reuse in engineering research fields within the UK Higher Education sector. In this sector engineering research is an extensive area, employing over 4400 full-time research staff making use of research funds of more than £2.3 billion.

The study reports the disposition and the types of research data found in a University engineering research centre at the University of Bath using a standard data audit methodology. The focus of the study is on data or documents upon which analysis is carried out so as to derive engineering research and design results. The data audit results are considered in relation to the curation aspects of (*a*) data creation, (*b*) working with the created data, (*c*) discovery, access and use of research data, (*d*) curation practices, and (*e*) data sharing.

The context of the data findings is considered in the light of previous ethnographic studies of how practising engineers use information and data for research and design as well as previous surveys reviewing the information and communication pattern of engineers.

The case study highlights the commercial and industrial context for engineering research which requires confidentiality and control of disclosure of information and data. This defines, in many cases, what 'appropriate' data sharing can be. There is a lack of open public repositories for engineering research data. The care and use of knowledge, information and data is embedded in the social knowledge networks that engineers use to carry out their work. Where information or data are held in a repository, engineers use colleagues to help locate them, understand their usefulness and check their trustworthiness in a professional context.

The findings of the case study are that for engineering research a simple prescriptive approach to curation issues such as data sharing is unlikely to succeed. Further work is needed to understand the creation and use of data over the wide and diverse fields of engineering research. The approaches taken to the sharing, reuse and preservation of data in fields like engineering research need to be appropriate if they are to influence researchers' choice of what to share, with whom and when.

Recommendations

Recommendation 1 (section 2.2). UK Higher Education Institutions should be supported in carrying out work to survey the disposition of engineering research data in their engineering research centres. The Digital Curation Centre should continue through case work and advocacy work its efforts to demonstrate not only the advantages of using standard audit methodologies such as the Data Audit Framework (http://www.data-audit.eu/) but also the potential benefits of providing a cross-institutional view of engineering research data in UK higher education, allowing comparison and consideration in an international context.

Recommendation 2 (section 3.2.6). The Digital Curation Centre should seek to collect examples of confidentiality agreements from UK higher education research projects that work with industrial partners, and use these to develop an understanding of the impact these have on data sharing practices over the research information lifecycle. This work should be used to inform the development of data sharing agreements and templates to supplement any DCC work on data management plans and templates.

Recommendation 3 (section 4). For engineering disciplines the role of informal sharing and the use of social knowledge networks should be further recognised by research funders and policy advocates in their policies governing research outputs and publications. More work should be funded to understand the multiplicity of contexts in engineering research which influence the appropriateness or otherwise of sharing, reusing and preserving knowledge, information and data. This work should be based on the study of activities and practices of work, such as mentoring, and supported by the engineering communities through which sharable use of information would be achieved.

1 Introduction and background

This DCC SCARP case study aims to provide some understanding of the approaches to data deposit, sharing and reuse in engineering research fields within the UK Higher Education system. The case study makes use of the existing literature covering how engineering professionals work with data, information and knowledge, relating this to the results of an audit of the types of research data found in an engineering research centre. Gaining an understanding of researchers' practice in producing, using and exploiting research data is an important requirement for supporting the uptake of digital curation techniques appropriate to the forms of digital working in each specialised field.

The broad themes of the SCARP case studies are as follows.

Policy drivers and barriers. Organisational and institutional factors including different skill levels, deposit and preservation policies and arrangements, willingness to use these and the relationships to incentives and reward structures within disciplines.

Stewardship practices. Research processes and methods, and how these relate to the digital objects created (e.g. data, drawings and models, software tools and methods, repositories and databases). How the digital assets produced (outputs) are used – including reuse and sharing – and linked to publication. Attitudes to doing this and to the usefulness of prior data, and the sustainability of collected digital information.

Tools and infrastructure. The tools and facilities used to collect, deposit, find, cite, discuss and annotate the digital objects and to ensure persistence and preservation over long time periods.

Preserving context. How communities of practice and their knowledge bases can be characterised, and how understanding of the lineage and provenance of digital objects may be documented.

1.1 Digital objects in Engineering

The variety of forms of data *produced* in engineering research is illustrated in the IdMRC Data Audit (section 3). Those listed include engineering calculations and analyses, modelling systems, specifications and instructions, data and text mining tools, concept and business process mapping systems, simulation data, equipment measurements and logs, questionnaire returns and interview recordings. The kinds of data *used* are even more extensive. Systems supporting visualisation and modelling are extensively employed.

In looking at digital working we need to keep in mind that engineers work with both analogue and digital forms. For instance, the use of logbooks by engineers is well documented (McAlpine, Hicks, et al., 2006), and these may exist in paper or digital forms.

The pattern of hybrid working, with analogue and digital forms, is observed in professions such as civil engineering by Suchman (2000) who describes the use of Computer-Aided Design (CAD) systems to support integration of analysis, 3D visualisations, flattened renditions as sections and plans with on screen working complemented by extensive use of paper plans in consultation meetings.

2 Engineering landscape

2.1 The character of Engineering as a discipline

Engineering, as a discipline or profession, covers a very wide field of knowledge. It is generally thought of, in a technological context, as the application of scientific, mathematical or technical knowledge in the design or implementation of devices or machines, materials and structures in order to carry out useful work or to meet a useful purpose. Given this wide scope any attempt to generalise about how engineers work, including digital working, needs to be tempered by appreciation that there has been only a limited number of studies of how engineers develop and use knowledge in actual practice (Radcliffe, 2008).

The nature of engineering work practice varies with the different industrial sectors and with the specialised engineering disciplines as well as the various stages of the lifecycle of engineered products and systems. Empirical studies of how engineers work in carrying out research and design have focussed at least partly on how they manage information, knowledge and experience. A particular concern has been to understand how engineers share and reuse knowledge and information (Ahmed, 2007). Any promotion of digital curation techniques to engineering design and research practitioners needs to take account of the empirical findings on how engineers work with information, including differences and variation in practices, if the promotion is to be appropriate.

2.2 Scope for this case study

Organisations carrying out engineering research and design produce and work with a wide variety of digital assets. These include data, tools and instruments such as algorithms and software. In the context of this case study the focus is on data or documents upon which analysis is carried out so as to derive engineering research and design results.

In the UK higher education system, the use of assets such as learning objects and published research articles is supported through specialised institutional repositories and databases. There are off-the-shelf software products which support institutional repositories and records management. The ongoing care of data produced in engineering research, however, does not receive special attention in current off-the-shelf products. This raises several research questions: Does engineering research require specialised systems supporting data? How is engineering research data currently cared for? What are the types and properties of engineering research data? What does the research literature tell us about how practising engineers use information and data for research and design?

The UK Research Assessment Exercise 2008 broke engineering down into the following units of assessment: Electrical and Electronic Engineering; General Engineering and Mineral & Mining Engineering; Chemical Engineering; Civil Engineering; Mechanical, Aeronautical and Manufacturing Engineering; Metallurgy and Materials. [1] Engineering is an extensive area of research work in the UK Higher Education sector with over 4,400 fulltime research staff employed, making use of research funding (from all sources) of more than £2.3 billion.

For the purposes of this case study the results of an audit of the data holdings in a research centre in Mechanical, Aeronautical and Manufacturing Engineering were considered. The result gives an indication of types of research data held but to generalise to the whole of engineering (as defined in the RAE exercise) would require an ongoing programme of work beyond a single case study.

Further data audit work in engineering could be supported by encouraging those responsible for the care of research data to make use of self-assessment tools. Two such tools currently available are the Data Audit Framework (DAF),[2] which provides organisations with the means to identify, locate, describe and assess how they are managing their research data assets, and the Digital Repository Audit Method based on Risk Assessment (DRAMBORA),[3] for self-assessment of the effectiveness and trustworthiness of digital repositories.

Recommendation 1. UK Higher Education Institutions should be supported in carrying out work to survey the disposition of engineering research data in their engineering research centres. The Digital Curation Centre should continue through case work and advocacy work its efforts to demonstrate not only the advantages of using standard audit methodologies such as the Data Audit Framework but also the potential benefits of providing a cross-institutional view of engineering research data in UK higher education, allowing comparison and consideration in an international context.

2.3 How do practising engineers use information and data for research and design?

Recent ethnographic studies, for example Demian and Fruchter (2006) and Baird, Moore, and Jagodzinski (2000), found that information work in engineering is socially mediated and that reuse of knowledge is carried out through social knowledge networks. In short, engineers rely on people as sources of information. In an industrial organisational context, information held within an organisation is more critical to carrying out engineering work

[1] The Research Assessment Exercise units of assessment for engineering may be found at http://www.rae.ac.uk/panels/main/g/ (archived at http://www.webcitation.org/5mjFugImw on 2010-01-12). For a US college view of engineering majors, see Engineers Dedicated to a Better Tomorrow (2007).

[2] For more information about the Data Audit Framework, see http://www.data-audit.eu/

[3] For more information about DRAMBORA, see http://www.repositoryaudit.eu/

Table 1. Characteristics of company experts versus those of details designers. Source: Lowe, McMahon, and Culley (2004a).

Company 'experts'	'Detail designers'
Use large volumes of information	Use small volumes of information
Are heavily reliant on personal collections	Maintain small personal collections
Tend not to trust 'central' storage systems	Rely on group/company information sources
Are prepared to search widely for information	Are less prepared to 'search' for information
Most dependent on text based information	Use drawings and geometry most frequently
Are the 'authors' of design guidelines	Are users of design guidelines
Know whether relevant work has been done before, where to look and whom to talk to	Carry out design by 'look-a-like'

than sources external to the organisation (Hertzuim & Pejtersen, 2000). Where information is held in a repository the engineers use colleagues to help locate it and understand its usefulness. In a design context, by speaking to a colleague or mentor the experience used to make decisions for a particular set of designs may be recovered and discussed. The appropriateness of use or reuse of data (information) can be gauged in the context of professional judgement and trust. Baird, Moore, and Jagodzinski captured social behaviours that supported management of risks in engineering such as engineers routinely 'prefacing their contributions to team meetings and discussions with the name of the contributors of their data', a 'verbal system of information provenance'. Skills are traded by members of different design teams on the basis of peer recognition of expertise and trust. Demian and Fruchter report that 'one of the primary mechanisms for knowledge reuse is through mentoring relationships'. The care of engineering information and data is embedded in the social processes that support how engineers work together.

2.4 Information needs and information-seeking behaviour of engineers

Lowe, McMahon, and Culley (2004a) introduced the concept of information profiles to adequately characterise the information work patterns of engineers, according to the different stages of the design lifecycle worked on and also in relation to the social environment in which the engineers work. The study, based on the case of design engineers in two different aerospace industry organisations, draws a clear distinction between engineers who can be considered company 'experts' and mainstream design engineers referred to as 'detail designers' (see table 1).

The role of personal information stores for the expert included the need to ensure custody of the full range of information and data needed to carry out early stage designs and feasibility studies. There was active distrust of organisation-wide repositories or document libraries. Personal stores included a large number of domain-specific technical reports and tended to exclude project information such as correspondence, minutes etc. which were stored in project files and/or local group filing systems. Lowe et al. (2004b, 2004c) found that around 20% of a designer's time is taken up with finding and absorbing information, with this increasing for those in a more technical role. Around 40% of design information requirements were met from personal information stores.

The study by Tenopir and King (2004) remains the most complete survey of information and communication patterns of engineers to date, synthesizing forty years of research into the topic. They found that engineers rely less on formal modes of communication such as journals, and more on interpersonal and information modes of communication than do scientists. The mix of formal and informal sources used by engineers is a result of the diverse range of activities that they participate in: not only research, development, teaching and management, but also design, production, construction and marketing. This is confirmed by Ward (2005) who also contrasted the explicit knowledge used by engineers, such as print and online sources, with the tacit knowledge held in the minds of the engineers, embedded in organisational practices and expressed in artefacts.

The variety of forms of knowledge used by engineers is also commented on by Fahy, Lervik, Easterby-Smith, and Elliott (2009), though they see the tacit–explicit dualism as something of an oversimplification. Their work is based on an approach of studying the actual practice of engineers in real work contexts; this approach focuses on the learning that takes place as people do things together. The theoretical perspective favours inter-subjective knowledge construction rather than regarding knowledge as something residing in an individual mind. The approach is illustrated in a growing body of 'practice turn' literature.[4]

Fahy, et al. apply this perspective in the study of the work of continuing product development engineers in the aerospace industry. For diagnosing problems with products, many forms of knowledge are important: expert knowledge, documentary records and drawings, experiments and calculations, 'community knowledge', 'knowing one's way around the archiving system', and having a 'feel' for problems. The engineers act as *bricoleurs*; that is, in order to accomplish tasks they assemble and organise knowledge from whatever material, mental, social and cultural resources are at hand.

In a study of South African consulting engineers, du Preez (2008) found that the variety of tasks in which engineers were involved led to a need for information from a wide variety of sources, few of them formal publications. An additional trend observed was for information

[4] For a brief review of 'practice-based studies' and the 'practice-turn' in organisation studies, see Corradi, Gherardi, and Verzelloni (2008) and also Geiger (2009).

needs to become less intense but more focused as a project progressed. Age, discipline and gender were found to have little influence on information needs and sources consulted, although older engineers relied more on their own experience and less on software. Carstensen (1997) lists the following as typical information resources: previous designs, similar products, design rationales, known issues with existing designs, component specifications, standards, norms, established procedures, production line characteristics, new materials and components, product documentation, personal expertise and literature.

There is in the literature a broad consensus that, despite variance by discipline, activity, country and personal characteristics (level of education, age), engineers are largely task-focused, and therefore prioritise accessible and easy-to-use information (Tenopir & King, 2004; Pinelli, 1991, 2001).

A study by Needham, Sidwall, Bevan, and Harrington (2002) confirmed that for technical information, engineers seek out the most accessible rather than the highest quality sources. They may not be aware of the full range of information sources available, and find technical reports hard to find. Kwasitsu (2003) also found that engineers' choice of information source was mostly based on availability, accessibility, relevance, currency and ease of use.

2.5 Engineering Digital Repositories Landscape Analysis

The concept of a repository, as a central place where documents and data are deposited and cared for by specialists, is well established within industrial engineering. For larger organisations at least, Product Lifecycle Management (PLM) systems are used to store and manage access to the design documentation of record produced by that organisation; many also facilitate the integration of design information into other systems and processes within an organisation. Repositories may be private to a company or project, commercial with access closed in the sense of being limited to subscribers, or open to public access without restriction. As described in the case study by Hertzuim and Pejtersen (2000), the repository in the sense of a corporate, company or project archive may consist of extensive paper documents and analogue artefacts and well as computer-based filing and retrieval systems.

Within academia, there is strong support for institutional repositories and discipline-specific data centres as curators of documents and data, often in relation to research output and published material. [5] The usage of such digital repositories in academia varies greatly between disciplines and between institutions.

[5] For some background on digital repositories in academia, see the Digital Curation Centre briefing at http://www.dcc.ac.uk/resource/briefing-papers/digital-repositories.pdf. For some case studies illustrating the development of higher education institutional information services based on repositories, see http://www.rsp.ac.uk/repos/cases and the other material available at the Repositories Support Project.

As part of the PerX project to develop a pilot repository cross-search service specifically aimed at engineering information, [6] MacLeod and Moffat (2005) produced a landscape analysis providing a review of digital repositories with content relevant to the engineering subject area. PerX excluded 'closed access' commercially produced repositories which do not allow free searching. This means commercial services such as traditional online databases services providing indexes, abstracts and access to published engineering literature (e.g. Inspec) were excluded.

Overall the finding of the landscape analysis was that the level of 'digital repository provision specifically for the engineering community appears to be relatively low'. PerX considered a wide range of resource types which repositories might provide: research data, pre- and post-prints of journal and conference papers, technical reports, electronic theses, learning objects, multimedia resources, and assessment materials; they also considered national initiatives related to such repositories, alongside subject repositories and open access journals. The report identified significant gap areas in the repository landscape, notably in research data, subject-based access, UK technical reports, open access journal articles and assessment materials repositories. In other areas, levels of repository provision were still relatively low, although there was better provision for metadata than full resources.

2.6 Sharing Engineering Information

MacLeod and Moffat (2005) provide several reasons why engineering information is not as freely available as information in some other disciplines. Journals published by or on behalf of professional and learned societies tend to have a balanced mixture of industrial and academic readers and authors; introducing an author-pays model of publishing would likely exclude industrial authors from contributing. Trade journals tend to publish content that is largely ephemeral – news items and analyses, market information, news and reviews of products and software, job advertisements, and so on – and therefore not amenable to deposition in a repository. A sizeable proportion of engineering information cannot be published at all due to matters of commercial sensitivity. While this is clearly important in the industrial context, it also affects the academic context where the research involves collecting data from industry or carrying out research with industrial and commercial partners. Indeed, some academic engineering research has more of the character of consultancy, with the research outputs circulated only within the industrial organisation providing the funding. Finally, engineering data are often generated from private systems, and therefore do not have quite the universal applicability that scientific data do.

[6] For the Pilot Engineering Repository Xsearch (PerX) project, see http://www.icbl.hw.ac.uk/perx/; the objective was to 'develop a pilot service provides subject resource discovery across a series of repositories of interest to the engineering learning and research community': see http://www.engineering.ac.uk/

In order to explain the landscape findings MacLeod and Moffat conducted a literature review of three different aspects of engineering: the information and communication needs of engineers, the complexity of their information landscape, and the information seeking behaviour of engineers within the engineering discipline. The following are among the conclusions drawn by MacLeod and Moffat:

- ‘Few repository sources were identified for engineering research data. There is no engineering equivalent of the provision of bodies like UK Data Archive with coverage of Social Sciences data.’
- ‘Overall there is less emphasis on open provision to research outputs in engineering than there is in science.’
- ‘Repositories per se are [...] unlikely to become as important in engineering as they have become, or may become, in some other disciplines.’

2.7 Disposition of research data

Given the above description of the information landscape for engineering, one approach to exploring the practice of engineers in working with information and data is to examine the disposition of research data in an academic centre where engineering research is carried out and promoted. Whatever method is used to survey data in a particular institutional setting some care is needed (and perhaps a degree of scepticism) in generalising results to the whole of an academic discipline or to the whole area of professional practice.

For this case study use was made of the pilot audit carried out at the University of Bath as part of the JISC project to develop the Data Audit Framework for use in UK Higher Education institutions. The benefit of taking this approach is that other engineering research centres could employ a common tool, use the same audit methodology, and so over time help to build a fuller picture of the data disposition across differing engineering research centres and across differing organisational or institutional settings.

3 Engineering research data in higher education

Research data can be understood as research material in two senses: (*a*) as part of the material used or created in carrying out a piece of research work, and (*b*) as part of the outcome of the research, supporting the research results and research products. Engineering research is often understood in a design context and involves working with a complex continuum of disciplinarity, and may seek not only to generate knowledge but also to support the transfer of techniques and to apply results in a professional practice context. The work is often strongly linked with industry.

3.1 Data Audit

The purpose of a digital data audit is to cultivate awareness of the existence, location, condition and value of digital assets. There are five core questions:

1. What data assets does the organisation currently hold?
2. Where are these assets located?
3. How have they been managed to date?
4. Which of these assets should be maintained in the long term?
5. Do current data management practices place any of these assets at risk?

An audit of the Innovative Design and Manufacturing Research Centre (IdMRC), University of Bath was carried out over the summer of 2008. [7] The work formed part of the process of developing the Data Audit Framework (DAF), which IdMRC materially assisted in by agreeing to act one of the pilot sites for the audit methodology. The lessons learned and basic approach taken are described in the report on the DAF methodology by Jones, Ross, and Ruusalepp (2009).

3.1.1 Research Centre profile

The Innovative Design and Manufacturing Research Centre (IdMRC) is a research group within the Department of Mechanical Engineering at the University of Bath. It was set up in October 2001 with funding from the EPSRC's IMRC programme, and is one of sixteen such centres in the UK. It has four research themes:

- Advanced Machining Processes and Systems (AMPS) – Computer Numeric Control (CNC) machining and inspection systems interoperation, new manufacturing techniques.

[7] For more information about the IdMRC, see http://www.bath.ac.uk/idmrc/ . For the context of collaborative working with industry, see http://www.bath.ac.uk/idmrc/resources/industrial_collaboration.pdf.

- Constraint-Based Design and Optimization (CBDO) – incremental improvements to manufacturing processes and workflows.
- Design Information and Knowledge (DIAK) – knowledge and information management, advanced engineering documentation.
- Metrology and Assembly Systems and Technologies (MAST) – measuring and reverse engineering artefacts, testing metrology equipment and techniques.

Each theme is lead by an academic and administered by a Research Fellow or Officer. As at summer 2008, the Centre had fourteen academics, three research fellows, sixteen research officers and twenty research students. The IdMRC's work is widely supported by industry, especially from the aerospace and packaging sectors and with emerging strengths in shoe and electronics manufacture.

3.1.2 Infrastructure profile

Extensive use is made of computer file systems (using formal directory structures and file naming conventions) as a repository for research and other data. Each academic and researcher of the Centre has their own PC, and access to a private network drive hosted by Bath University Computing Services (BUCS). The academics within the Department of Mechanical Engineering also have drive space on the general University network drive (again hosted by BUCS) for use as shared drives by the projects they lead (hereafter referred to as the Research drive). For example, researchers working on the KIM Project, headed by IdMRC Director Christopher McMahon, have access to a shared drive under `.../Research/CAM/CAM-0001/`. A text index file under `.../Research/` indicates to which project each of these directories corresponds. A further index file in each academic's directory specifies to which academic the directory corresponds and, for each subdirectory, the corresponding project, the date the folder was created, the date when the was folder archived (if applicable) and a contact list of people chiefly responsible for the content; additional notes may also be provided if the directory uses any special conventions or configurations. Archived folders are kept for seven years unless otherwise specified in the index file.

In addition, the Centre runs a pair of desktop PCs acting as servers, with one machine backing up the other (hereafter referred to as the IdMRC drive). Each member of staff has a personal area on these latter machines, while each theme has a shared area accessible to all staff working on that theme. Several researchers use external hard drives to store and transport their data, and a number back their data up on CD-R. The Centre has a lab for conducting experiments with manufacturing and measurement equipment. Most of the pieces of equipment are controlled by PCs, which also act as the primary repository for data generated by the machines. The University provides several additional tools relevant to data management. It has its own Subversion repository, primarily for collaboration on software development although it can be used to manage any plain-text-based project. It also has its own Wiki system; several researchers are using this as an alternative to sending attachments by e-mail, and for collaborating on documents.

3.1.3 Data management policies

Aside from the aforementioned policies and procedures for project data on the Research drive, there are no formal data management policies in place within the Centre, although some ad hoc principles are observed. For example, all data and documents exist in at least two separate locations, and in many cases a further copy is lodged with the supervising academic. More specific conventions are observed within each theme. For example, within DIAK, any raw material deposited on the IdMRC drive – videos, sensor data – is kept unchanged; any processed or cleaned versions are stored separately and do not overwrite the original data. Within MAST, a formal directory structure and file naming convention has been employed. The Research drive is a recent innovation, resulting from negotiations with BUCS; of the fourteen IdMRC academics, five have project folders in this area. While researchers have been encouraged to use it as the primary location for storing their research data, take-up is slow. There is no specific budget for data management.

3.1.4 Profile of data holdings

The Centre does not have a wealth of past data assets to draw on. This is partly because the Centre is young – just under seven years old at the time of the audit in summer 2008 – but rather more because of the nature of the research carried out.

A substantial proportion of the research performed by the Centre is carried out in partnership with one or more private companies – as mentioned above, the Centre has strong links with the aerospace and packaging industries as well as shoe and electronics manufacturers. This means that many of the data held by the Centre are either supplied by private companies or derived from data, information or knowledge held by them; consequently, these data are protected by confidentiality agreements that may forbid their reuse or even demand their destruction following the conclusion of the research. In such a climate, there is little incentive or scope to share data across the Centre, and so the prevailing culture is of each researcher looking after their own data on their own drive spaces. For the audit, this meant that data collected and used by former researchers were largely invisible – though in some cases copies were held by supervising academics – giving a picture of mainly youthful data.

The assets predominantly fall into one of the following types.

- *Generated data.* The data assets produced by the AMPS and MAST themes are largely generated from the manufacturing and metrology machinery held at the IdMRC.

- *Data supplied by industry.* The three themes other than MAST work extensively with data sets supplied by industry. The long term importance of these data depends in no small part on the confidentiality constraints placed on them: some may be kept for future research or reference, while others must be deleted following completion of the primary research. Three of the assets identified by the audit were held off-site by the external organisations to which they belonged.
- *Survey data* (questionnaire responses, interviews). These appear only within the DIAK theme, and in a variety of textual and audiovisual forms.
- *Software.* The work of the Centre, across all themes, requires a certain degree of technological innovation and so the Centre has produced custom software (and indeed hardware) in the course of research. Most of these assets are still in use but not currently curated.

The IdMRC has a number of other digital assets which fell outside the scope of the data audit: for example, collections of researchers' best papers, collected journal articles and conference proceedings, and spreadsheets of contact information.

3.2 Curation aspects

For the purpose of this SCARP case study the findings of the IdMRC data audit were examined and considered in relation to the following curation aspects:

- Research Domain
- Data Creation
- Working with the created data
- Discovery, access and use of research data
- Curation practices
- Data sharing

3.2.1 Research domain

Engineering research is carried out both by academics and within industry. Academia generates some data that are not research data, while in industry the vast majority of data are not what one would consider research data, although some classes of data appear in both research and development contexts: for example, data arising from testing aspects of designs such as maximum loads and aerodynamic behaviour.

The academic study of engineering has several branches:

- developing experimental technologies;
- improving design and manufacturing processes, tools and techniques;
- understanding current engineering practice;
- providing consultancy services to engineering firms.

Industrial work is best characterised by product lifecycle: design (subdivided into requirements, embodiment, and detailed design stages), manufacturing, in-service maintenance, re-design and upgrading, decommissioning/disposal.

3.2.2 Data creation

In the course of its research, the IdMRC generates data of the following types:

- Computer-aided design (CAD) models in many different incompatible and ephemeral formats.
- High speed (MPEG) and normal (MPEG, QuickTime, WMV) video captures of, for example, manufacturing machines in action.
- Ontologies (RDF, HTML, DAML, PPRJ).
- Questionnaire returns (regular Office formats).
- Interview recordings (Wave, MP3, WMA, proprietary Sony formats).
- Data and text mining software and output data (XML, CSV).
- Finite element analysis inputs (BDF Bulk Data Files) and outputs (Nastran OP2).
- Digital pen image files (plus GIF transformations).
- Mind maps (Inspiration).
- Product structure maps (Studio MDL).
- Matlab methods, macros and calculations.
- IDEF0 process diagrams.
- Topic maps (XML).
- Rationale maps (IBIS, DRed).
- Photographs (JPEG) of, for example, equipment layouts.
- Constraint modelling software, macros and data outputs (SSW).
- Simulation data (Witness).
- Measurements and logs from experimental equipment, measurement equipment, manufacturing equipment (various formats, including XML and proprietary text/binary files).
- Locally developed software.

In addition the Centre had access to many different corpora of industrial documents: some within custom database/retrieval systems held by industrial partners, some held on campus in regular Office document formats. For the most part these documents are commercially sensitive.

These data are produced for the four reasons outlined above:

- Proving the concept of innovative products and experimental techniques; for example, indoor GPS for use in manufacturing, new modes of transport.
- Improving existing design and manufacturing processes, tools and techniques; for example, improving the reliability of packaging machines.

- Understanding current engineering practice; for example, how manufacturing costs are estimated.
- Providing consultancy services to engineering firms; for example, determining possible improvements to their internal processes and manufacturing methods, providing specialist metrology services.

The data tend to be used on a short term basis, that is, to support a single paper, report or thesis, rather than to build up a resource over time. Data were kept in order to enable the findings of a paper to be justified if necessary, without any expectation of reuse.

At the time of the audit, the IdMRC did not collect preservation or curatorial data. Some ontologies and classifications were produced, but not necessarily with direct relevance to a particular dataset. In general, effort was not dedicated to adding value to data beyond the cleaning and analysis performed as part of preparing the associated paper, report, or thesis. There was no culture of fully documenting data, meaning that the significance of data produced by the Centre was and is in danger of being lost as researchers move on to other things.

3.2.3 Working with the created data

The IdMRC uses several tools and technologies for analysis. The following is not an exhaustive list:

- Waypoint: faceted classification
- Clemantine: data mining
- Nastran: finite element analysis
- SSW: constraint modelling
- Matlab: calculations

Researchers tend to work singly or in very small teams on a particular research question. In CBDO and MAST, researchers tended to work as lone consultants; in the former case, examining a production line for possible improvements, and in the latter case, performing measurements using specialist equipment. Within AMPS, a number of small projects were being run, including some standards work. Within DIAK, at the time of the data audit most of the work was associated with a single research project: a collaboration with ten other universities, and involving both engineering and management researchers.

Some of the raw data that researchers worked on was so confidential the researchers had to use it at the industrial partner’s site, and take only aggregated/anonymised derived data away. The majority of data used or produced by the Centre can be kept on campus, but is still commercially sensitive and cannot be released; some data may not even be shared with other researchers in the same team or theme. The only data found by the audit to be published in any form was the software, firmware and data supporting the RepRap self-replicating rapid prototyper (or '3D printer'), which is available on SourceForge.[8]

The data handled by the Centre is either owned by industrial partners or generated by researchers, often from resources owned by industrial partners. As already mentioned, this places tight restrictions on publication and use. The audit discovered no evidence of third parties requesting access to the Centre's data assets, and due to the issues above such requests would likely be refused in most cases.

3.2.4 Discovery, access and use

The *constraints* on discovery, access and use at the Centre are as follows:

- Commercial sensitivity (as noted above) places significant restrictions on discovery, access and use.
- With a few exceptions, most of the specialist software in use in the Centre is licensed on a time-limited basis, meaning that the readability of data associated with that software relies on (*a*) the vendor continuing to offer the licence, and (*b*) the Centre continuing to pay for it.
- There are few repositories of engineering information of any kind, and certainly there are no repositories dedicated to engineering research data.

The corresponding *enablers* are as follows:

- The Engineering Repository Cross Search Demonstrator pilot service (http://www.engineering.ac.uk) provides a cross-search service for what little Engineering information is available in repositories; this is mainly useful for discovering teaching and learning materials rather than research data.
- Some of the research within the Centre is looking at how to make industrial data more integrable and interoperable, so some of the proof-of-concept data exhibit these characteristics.

3.2.5 Curation practices

Quality assurance was largely left to individual researchers. No particular procedures were imposed or commonly used. To a large extent, quality was determined through attempting to use the data as the basis for published work. In one particular strand of research, a sample group of students was recorded performing a design task. This needed to be repeated (with different groups, design tasks and recording methods) a few times before a usable set of data was created.

[8] For more information on RepRap, see http://reprap.org/ or the project's SourceForge page: http://sourceforge.net/projects/reprap/

Most of the activity in the IdMRC relating to the long-term care and preservation of data is concerned with storage. All data and documents within the IdMRC exist in at least two locations, although a standard for which locations are used has not yet been established: individual researchers choose from desktop hard drives, machine control computer hard drives, removable media (hard drives, CDs, flash drives), and private and shared network drives. There has been a move to get project data onto a dedicated networked Research drive; management procedures, a formal directory structure and a retention policy have been implemented for this drive, the details of which are given at section 3.1.2.

3.2.6 Sharing data

Sharing data is impeded largely by commercial sensitivities, and the resultant confidentiality agreements. Where these do not apply, there is a mixed attitude to sharing data, with some groups happy to make data available, some more cautious (perhaps looking to the possibility of a spin-out company) and to some the possibility of others being interested in the data, or of being able to use it, does not occur.

Recommendation 2. The Digital Curation Centre should seek to collect examples of confidentiality agreements from UK higher education research projects that work with industrial partners, and use these to develop an understanding of the impact these have on data sharing practices over the research information lifecycle. This work should be used to inform the development of data sharing agreements and templates to supplement any DCC work on data management plans and templates.

4 Policy on care of engineering research data

Why bother to keep and care for data? Among the purposes served are:

- to validate published research results, promoting challenge, scientific consensus, trustworthy scholarship;
- to retain unique observational and measurement data impossible to re-create;
- to retain data where cost of maintenance is cheaper that cost of re-generation;
- to act as reference data for future research, including collection-based research to generate new understanding in science and technology;
- for use in teaching and public understanding of science and technology;
- to comply with legal, regulatory and ethical requirements including exploitation of intellectual property;
- to permit informed choice, audit and record in decisions regarding data destruction and deletion at appropriate stages of the information lifecycle;
- to manage the persistence of the cultural and scientific record allowing for future technological exploitation, scholarship and public understanding.

The Engineering and Physical Sciences Research Council (EPSRC, 2009) *Policy on access to research outputs* is an echo of the Research Councils UK position on access to research outputs, and so supports the principles that ideas and knowledge derived from publicly-funded research must be made accessible for public use (as far as practicable) and that outputs must be preserved and remain accessible for future generations. Jones (2009) provides an analysis of the UK research councils' policies in regard to curation requirements.

The EPSRC policy focus is on research outputs and publication; secure storage of the primary data used as a basis for publication is a requirement. The EPSRC does not provide or support a repository service for the deposit of engineering research publications or data. The researcher's institution is expected to provide the institutional or subject-based repository services required. This means in engineering research that provision for the curation of knowledge, information and data will vary greatly according to the facilities and resources of individual institutions and the pattern of local uptake of the services offered.

The focus on research outputs and publication means that research data, understood as research material used or created in carrying out a piece of research work, is not really in the scope of the policy. Rather the emphasis is on the specific sets of data that support the results and products summarised or referred to in a research publication (journal paper, conference paper), hence the concern with holding a copy securely in case checking is needed to investigate a challenge to the findings. In engineering research there may be less need, or demand, for data sharing in research publications.

Recommendation 3. For engineering disciplines the role of informal sharing and the use of social knowledge networks should be further recognised by research funders and policy advocates in their policies governing research outputs and publications. More work should be funded to understand the multiplicity of contexts in engineering research which influence the appropriateness or otherwise of sharing, reusing and preserving knowledge, information and data. This work should be based on the study of activities and practices of work, such as mentoring, and supported by the engineering communities through which sharable use of information would be achieved.

5 Conclusions

In areas like Engineering the evidence of ethnographic studies and research on communication and information patterns of engineers is that the care and use of knowledge, information and data are embedded in the social processes that engineers use to do their work. The issues of data deposit, sharing and reuse need to be informed by some understanding of the exigencies faced by engineers in their research work, according to their local circumstances and resources. The pattern is that sharing information is conducted firstly through personal contacts and that the rationale, methods and analysis used in producing data are just as important to share as the data itself; this communication activity forms part of professional consideration before reuse can be made of 'data', or the techniques used to produce data, in the different research contexts. Much of this communication activity is not reflected in formal channels such as research journal publication.

Research communities working in the wide fields of knowledge made subject to an engineering research approach do not share a universal set of general disciplinary characteristics that would support a simple prescription or fresh set of procedures or policies, mandating the sharing or reuse of data. The approaches taken to the sharing, reuse and preservation of data in fields like engineering research need to be appropriate if they are to influence researchers' choice of what to share, with whom and when.

Factors such as confidentiality in working with commercial partners or clients strongly influence the appropriateness of what can be shared. The engineering and research practitioners, and the managers of engineering research and design programmes, bring to bear their own professional understanding of the material interests in each of the specialised areas of work made subject to engineering approaches. This includes experience of working to satisfy client requirements and protect client interests. The professional understanding and judgment of what needs to be done to manage the material interests of the cooperating parties is a major factor in ensuring success of a research or design project.

In considering the digital assets (data, models, software, systems of computing, digital instruments) used to support engineering research work, the majority of digital assets may remain private to the creator, research team or owning institution. Only a subset of this information will be selected as a basis for collaboration with partners. In a commercial context, the methods used in engineering research and design may be regarded as private intellectual property to be protected so as to ensure exploitation. The eventual publication of results, or the public availability of a research product, is achieved through stages and processes that include the appropriate control of disclosure of information.

The information rights associated with digital assets require management. In making use of a repository for digital assets (whether simply a computer file system with manual care of assets or a purpose-designed repository system using standard protocols to support interoperability and federation across individual repositories) there is a need to enforce the policy governing information and data sharing agreements. The objective is the support of appropriate sharing within a secure and trusted digital environment providing policy-based information rights management at use. In manual systems of working (e.g. repositories based on computer file systems) this is managed by the researchers acting as gatekeepers within an understood system for authorising significant disclosure to third parties (e.g. agreement by the research director or a senior engineer).

Treloar and Harboe-Ree (2008) describe a curation continuum approach which provides an account of the different domains (private, shared, public) over which repository services and digital data stores for research data might be expected to function. They demonstrate how curation needs (e.g. for metadata to be provided at a certain breadth and depth) do vary according to the different domains served by repository services. This variation is particularly strong within engineering, where the perceived implications of information being in one domain rather than another are huge. This factor, combined with the economic pressures faced by industry, has a strong influence on whether and how much research information can be made publicly available and how much in the way of resources and time can be expended in making such information readily and safely reusable.

6 References

Allard, S., Levine, K. J., & Tenopir, C. (2009). Design engineers and technical professionals at work: observing information usage in the workplace. *Journal of the American Society for Information Science and Technology*, *60*(3), 443–454. Retrieved November 17, 2009, from http://dx.doi.org/10.1002/asi.21004

Ahmed, S. (2007). Empirical research in engineering practice. *Journal of Design Research*, *6*(3), 359-380. Retrieved November 17, 2009, from http://dx.doi.org/10.1504/JDR.2007.016389

Baird, F., Moore, C. J., & Jagodzinski, A. P., (2000). An ethnographic study of engineering design teams at Rolls-Royce Aerospace. *Design Studies*, *21*, 333-355. Retrieved November 17, 2009, from http://dx.doi.org/10.1016/S0142-694X(00)00006-5

Bracewell, R., Wallace, K., Moss, M., & Knott, D. (2009, March). Capturing design rationale. *Computer-Aided Design*, *41*(3), 173–186. Retrieved November 17, 2009, from http://dx.doi.org/10.1016/j.cad.2008.10.005

Carstensen, P. H. (1997). Towards information exploration support for engineering designers. In S. Ganesan (Ed.), *Advances in Concurrent Engineering, CE97* (pp. 26–33). Lancaster, PA: Technomic. Retrieved November 17, 2009, from http://www.it-c.dk/people/carstensen/pdf-files/CE97-explo.support.pdf

Corradi, G., Gherardi, S., & Verzelloni, L. (2008, April). Ten good reasons for assuming a 'practice lens' in organization studies. Paper presented at the Organizational Knowledge, Learning and Capabilities conference, Copenhagen, Denmark. Retrieved November 17, 2009, from http://www2.warwick.ac.uk/fac/soc/wbs/conf/olkc/archive/olkc3/papers/contribution312.pdf

Demian, P., & Fruchter, R. (2006) An ethnographic study of design knowledge reuse in the architecture, engineering and construction industry. *Research in Engineering Design*, *16*(4), 184–195. Retrieved November 17, 2009, from http://dx.doi.org/10.1007/s00163-006-0010-x

du Preez, M. (2008). *Information needs and information-seeking behaviour of consulting engineers: a qualitative investigation.* Unpublished master's dissertation, University of South Africa. Retrieved November 17, 2009, from http://etd.unisa.ac.za/ETD-db/theses/available/etd-07312008-131924/unrestricted/dissertation.pdf

Eng, N., Bracewell, R., & Clarkson, J. (2008, May). The role of narrative in evolving engineering design documentation. In *Proceedings of the 10th International Design Conference (Design 2008)* (pp. 769–776). Dubrovnik, Croatia.

Engineering and Physical Sciences Research Council. (2006). *Guide to good practice in science and engineering research.* Retrieved November 17, 2009, from http://www.epsrc.ac.uk/CMSWeb/Downloads/Other/GoodPracticeGuideSciEngRes.pdf

Engineering and Physical Sciences Research Council. (2009, June 5). *Policy on access to research outputs.* Retrieved November 17, 2009, from http://www.epsrc.ac.uk/AboutEPSRC/AccessInfo/ROAccess.htm

Engineers Dedicated to a Better Tomorrow. (2007, February 1). *A brief guide to engineering majors.* Retrieved January 12, 2010 from http://www.dedicatedengineers.org/Resources/Engineering_Disciplines_Handout.pdf

Fahy, K. M., Lervik, J. E., Easterby-Smith, M., & Elliott, C. (2009, April). Bricolage and alignments: a practice lens on knowledge work in multi-disciplinary engineering teams. Paper presented at the International Conference on Organizational Learning, Knowledge and Capabilities, Amsterdam. Retrieved November 17, 2009, from http://www.feweb.vu.nl/olkc2009/RoundTable/RT1Fahy.pdf

Geiger, D. (2009). The Practice-Turn in Organization Studies: Some Conceptual and Methodological Clarifications. In A. G. Scherer, I. M. Kaufmann, & M. Patzer (Eds.), *Methoden in der Betriebswirtschaftlehre: Tagungsband der Kommission Wissenschaftstheorie* (pp. 187–205). Wiesbaden, Germany: Gabler. Retrieved November 17, 2009, from http://dx.doi.org/10.1007/978-3-8349-9473-8_11

Hertzuim, M., & Pejtersen, A. M. (2000). The information-seeking practices of engineers: searching for documents as well as people. *Information Processing & Management*, *36*(5), Sept. 2000, Pages 761-778. Retrieved November 17, 2009, from http://dx.doi.org/10.1016/S0306-4573(00)00011-X

Jones, S. (2009, March 30). *A report on the range of policies required for and related to digital curation* (Deliverable H1.1). Glasgow: Digital Curation Centre. Retrieved November 17, 2009, from http://www.dcc.ac.uk/docs/reports/DCC_Curation_Policies_Report.pdf

Jones, S., Ball, A., & Ekmekcioglu, Ç. (2008). The Data Audit Framework: a first step in the data management challenge. *International Journal of Digital Curation*, *3*(2), 112–120. Retrieved November 17, 2009, from http://www.ijdc.net/ijdc/article/view/91/109

Jones, S., Ross, S., & Ruusalepp, R. (2009, May). *Data Audit Framework Methodology* (Draft for discussion, version 1.8). Glasgow: HATII. Retrieved November 17, 2009, from http://www.data-audit.eu/DAF_Methodology.pdf

Kwasitsu, L. (2003). Information-seeking behavior of design, process, and manufacturing engineers. *Library and Information Science Research*, *25*(4), 459–476. Retrieved November 17, 2009, from http://dx.doi.org/10.1016/S0740-8188(03)00054-9

Lowe, A., McMahon, C., & Culley, S. (2004a). Characterising the requirements of engineering information systems. *International Journal of Information Management*, *24*(5), 401–422. Retrieved November 17, 2009, from http://dx.doi.org/10.1016/j.ijinfomgt.2004.06.008

Lowe, A., McMahon, & C., Culley, S. (2004b, March/April). Information access, storage and use by engineering designers: part 1. *Engineering Designer*, 30–32.

Lowe, A., McMahon, C., & Culley, S. (2004c, May/June). Information access, storage and use by engineering designers: part 2. *Engineering Designer*, 23–25.

MacLeod, R., & Moffat, M. (2005, November 10). *Engineering digital repositories landscape analysis, and implications for PerX.* Retrieved November 17, 2009, from http://www.icbl.hw.ac.uk/perx/analysis.htm

McAlpine, H., Hicks, B. J., Huet, G., & Culley, S. J. (2006). An investigation into the use and content of the engineer's logbook. *Design Studies*, *27*(4), 481–504. Retrieved November 17, 2009, from http://dx.doi.org/10.1016/j.destud.2005.12.001

Needham, P. A. S., Sidwell, K., Bevan, S. J., & Harrington, J. (2002, October). *The MAGiC project: Managing access to grey literature collections.* Retrieved November 17, 2009, from http://hdl.handle.net/1826/58

Pinelli, T. E. (1991). The information seeking habits and practices of engineers. *Science and Technology Libraries*, *11*(3), 5–25.

Pinelli, T. E. (2001). Distinguishing engineers from scientists: the case for an engineering knowledge community. *Science and Technology Libraries*, *21*(3/4), 131–163.

Research Councils UK. (2006, June 28). *Access to research outputs.* Retrieved November 17, 2009, from http://www.rcuk.ac.uk/access/

Research Information Network. (2007, January). *Research funders' policies for the management of information outputs.* Retrieved November 17, 2009, from http://www.rin.ac.uk/files/Funders%27%20Policy%20&%20Practice%20-%20Final%20Report.pdf

Radcliffe, D.F. (2008). Contingent Immersion: a naturalistic methodology for discovering socio-technical realities of engineering practice. Paper presented at the Research in Engineering Education Symposium (REES), Davos, Switzerland. Retrieved November 17, 2009, from http://www.engconfintl.org/8axabstracts/Session%203A/rees08_submission_22.pdf

Suchman, L. (2000). Embodied Practices of Engineering Work. *Mind, Culture, and Activity*, *7*(1), 4–18. Retrieved November 17, 2009, from http://www.informaworld.com/smpp/content~content=a785309362~db=all

Tenopir, C., & King, D. W. (2004). *Communication patterns of engineers.* Hoboken, NJ: Wiley-Interscience.

Treloar, A., & Harboe-Ree, C. (2008). Data Management and the curation continuum: how the Monash experience is informing repository relationships. Paper presented at VALA 2008, Melbourne, Australia. Retrieved November 17, 2009, from http://www.valaconf.org.au/vala2008/papers2008/111_Treloar_Final.pdf

Ward, M. (2005). Information and the engineer. In R. A. Macleod, & J. Corlett (Eds.), *Information sources in engineering* (4th ed.). Munich: Saur.