

COMPARING THE INFORMATION CONTENT OF FORMAL AND INFORMAL DESIGN DOCUMENTS: LESSONS FOR MORE COMPLETE DESIGN RECORDS

Hamish McAlpine¹, Ben Hicks¹ and Stephen Culley¹
(1) University of Bath, UK

ABSTRACT

Even the most routine engineering design projects generate large volumes of formal and informal information by virtue of an ever-increasing variety of tools and systems. During the design process, the efficient creation and transformation of this information is critical to the success of the project. However, at the end of the design process there is a requirement to prepare formal project documentation necessary to support manufacturers and customers, and conform to legislation. This includes elements such as CAD models, performance and manufacturing data. In the process of preparing this documentation, only a small proportion of the information generated can be captured – if only because of limited resources. It is therefore unclear as to whether potentially valuable design information could be omitted from this formal record and eventually lost. In order to explore this issue, a comparative study of documentation from a design project has been undertaken. The study analyses the difference in information types and content of a complete set of formal and informal documentation. The results are then discussed with respect to the creation of more complete design records.

Keywords: design records, engineering logbooks, notebooks, information loss

1 INTRODUCTION

Design documentation represents information and knowledge that cannot be gained from the artifact alone, such as the design alternatives considered, rationale for decisions, and how to manufacture and maintain the artifact. However, to document even a small project in a way that is complete and allows efficient re-use is a considerable challenge for two reasons:

Firstly, it is complex because of the sheer variety of ways the information is stored, with a recent survey finding 105 electronic file formats in use between just 40 engineers [1]. This multitude of types and data formats exist partly because each has been developed to support different aspects of the design process particularly well. For example, CAD tools afford precise representations of very complex designs to be created and communicated quickly, whilst at the other end of the spectrum, logbooks (or notebooks as they are sometimes referred to) afford the quick recording of ideas through notes, sketching and quick calculations. Secondly, this problem is compounded as much of this information is not formally managed in a way that facilitates re-use or even retrieval. This is particularly true of the associated ‘informal’ or unstructured information (including emails, logbooks, meeting notes, and presentations which represent the emerging design), which are often inaccessible or discarded completely at the end of a project.

Of these, logbooks in-particular appear to contain a rich variety of informal design-related information [2] including sketches, contact information, design ideas/rationale and meeting notes. There is also evidence of logbooks being used to support design activities considered important, such as ‘self explanation’ and problem solving [3], sketching [4] and being able to provide evidence of thought processes and rationale [5]. However, they remain a “largely untapped resource” [6] for engineering organisations, with little research on them from a design records perspective and in-particular how they differ from the ‘formal’ record. This relative lack of research into the nature of - and differences between - informal logbooks and formal project records forms the motivation for this research. This paper therefore presents a comparative study of the informal information recorded in engineering design logbooks and the associated formal project record (project reports and CAD drawings). The

aim is to better understand the nature, content and relationship between formal and informal records so that lessons may be drawn for making the design record more complete and arguably more useful. Following a description of the methodology, the results are presented and discussed in detail in section 3. Section 4 builds on the results with a discussion and illustrative example. Section 5 then discusses these results with respect to their implications for the design record.

2 METHODOLOGY

This section first gives an overview of the methodology. The dataset used is discussed in 2.1, below. The classification schema created to analyse the dataset covered both the type of information and its nature, and is discussed in detail in 2.2. How the analysis was performed is then discussed in 2.3.

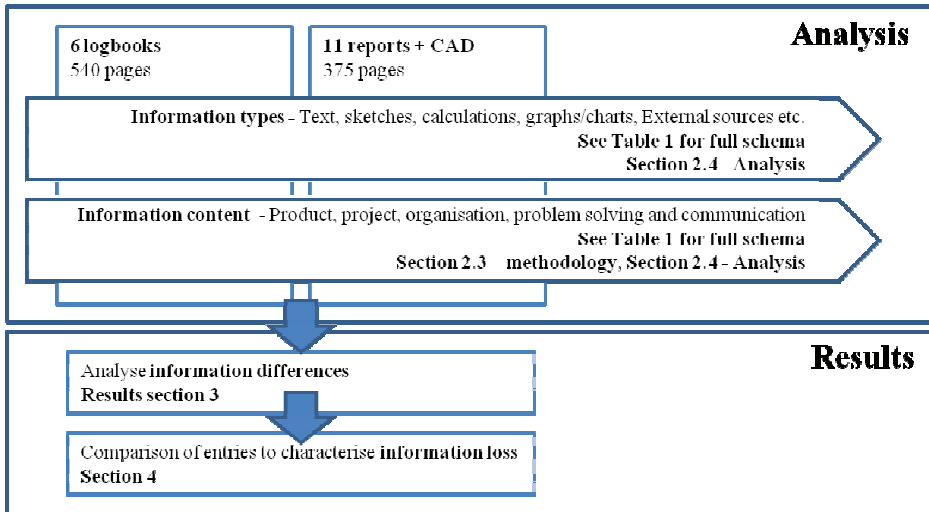


Figure 1. Overall methodology

2.1 Data set

The documentation used in this study was generated by six trainee engineers studying mechanical engineering at the University of Bath. It was a three-month project carried out in conjunction with a large UK-based engineering organisation. The sponsoring organisation's role was as the customer: they gave the team a design brief, and provided resources and access to facilitate the project. The brief was to re-design a module for a large packaging machine in order to reduce changeover times. The six team members had all previously worked for one year in a variety of engineering organisations. One team-member was assigned the role of project manager and it was their responsibility to organise the other team members in terms of tasks, roles etc.

This project was chosen for several reasons: Firstly, the dataset was relatively complete and self-contained. Whilst there were some associated emails and work on a whiteboard, it represented a large proportion of the physical record. Secondly, the engineers did not know in advance that the documents would be analysed, which was essential if meaningful comparisons were to be drawn. Finally, it represented a 'real' industrial problem, with engagement from a range of stakeholders and tasks spread across various design stages – from defining the problem to relatively detailed design. The authors therefore believe it was an accurate reflection of a common engineering scenario.

2.2 Classification schema

Following lessons learned from a previous attempt [7], a new information classification schema was created [8]. In summary, the previous approach used a schema originally developed to illustrate information loss in design review meetings [9]. The schema had four categories: *rationale*, *lessons learned*, *decisions* and *actions*, and was used to produce visual maps centred around topics. However, whilst it was used very successfully to analyse the discourse in design review meetings and compare it

to the formal minutes, it was difficult to apply it to logbooks and reports. Fundamentally, the information in logbooks did not correspond to single parts of the formal documentation, which meant that comparing the maps (and thus exploring the information loss) was almost impossible.

Therefore, a new information classification schema was created from an extensive review of the literature on classifying information in various fields, including engineering design, sociology and organisational behaviour. The categories attempted to comprehensively categorise information, covering *How* the information is presented, *What it is about* (whether it is product or process-related) and also *Why it is being created* (in terms of problem solving activities and intent of the communication). The rationale for the schema terms is discussed in detail in [8] and the full schema is presented in Table 1, overleaf.

For the purposes of this research, two modifications were made. Firstly, the analysis included the information classes (such as written notes, meeting notes, calculations, sketches etc) previously identified in logbooks by the authors [2]. Secondly, the classification schema originally included a 'communicative acts' category. This was intended to classify the type of interaction where two-way communication between people was involved (e.g. meetings or emails) and covered the way in which language was used in dialogue – for example, if the author of an email used language that *agreed* or *disagreed* with another team member. It was thus not appropriate to include it for this particular research, where the communication was essentially one-way (team member to logbook or report) and not expressed in the form of dialogue that could be analysed in that way.

Whilst the exploratory nature of the research means that certainty about the completeness of the schema cannot be claimed, it has been used extensively to mark-up other design documentation (primarily emails) covering a variety of types of design project, and has been through several iterations to ensure reasonable completeness of coverage with respect to the aims of the research. For this study, virtually all entries were marked-up with at least one term from each top-level category and all entries could be classified into one of the 13 previously identified information classes.

2.3 Analysis

For the purposes of marking-up the documents, the information was split into appropriate 'chunks'. Logbooks were split into entries, which were generally headed with a date or subject and ended with a terminating line, or the start of a new entry and were very easy to identify. The reports were split according to their numbered sub-sections. Whilst not perfectly analogous, it was felt that report sub-sections did correspond to logbook entries, as they both dealt with one aspect of the design or process, thus making the mark-up more manageable and the analysis meaningful. If, for example, the reports were analysed by entire sections, or logbooks in week-long chunks, so many terms from the classification would likely apply as to render any meaningful analysis unlikely.

The inter-coder reliability for the schema was not formally assessed for this research due to time constraints and because the reliability of the schema had already been assessed when applied to other documentation [8]. Although this assessment found levels of agreement between coders at the lower-level terms did vary, there was "near perfectly consistency" between coders at the higher (product, project, organisation) categories. A further, larger study of an email corpus with the same classification schema carried out at the University of Bath revealed that the inter-coder reliability for two coders as measured by the kappa co-efficient, was greater than 0.7.

The analysis included marking up the information types contained in each entry, such as written notes, meeting notes, calculations, sketches etc overleaf. It was felt that this would allow more insight to be drawn about the differences in types of information as well as their nature. This was important as some types of information (such as sketching and calculations) have a special importance in an engineering context. The actual number of sketches, calculations and CAD drawings (representing discrete information types that could be compared directly in a meaningful way) were also counted.

The results for the information content categories (product, project, organisation, problem solving and communication activities) are presented as the percentage of entries that contain each category and sub-category. For example, if 50 of 100 logbook entries contained some aspects of product performance, the percentage of entries would be 50%. As entries can (and often did) contain multiple categories of information content, the graph percentages do not add up to 100%.

What			Why
Information classes	Product	Project	Problem Solving
Written note	Functions Things the product must do, e.g. 'Be fast'	Risk Assessing likelihood and weighting implications	Goal Setting Identifying where the design is, and to where it needs progressing
Meeting note			Clarifying Clearing up misunderstandings (both requesting and in response).
Sketch	Performance How well the product achieves its functions	Planning Management of phases, activities and tasks	Debating Discussing opposite views
Table of figures			Informing Sharing, presenting or distribution information with others. No response is required. It is passive
Calculation	Feature The quality or characteristic with which the function is achieved	Team Team selection and development	Solving Encompasses one or more of the following stages: searching, gathering, creating and developing solutions
Contact Details	Operating Environment Objects that interact with the product	Quality Management Quality, standard or expectations	Evaluating Judging the quality, value and importance of something
Graph/Chart	Materials Materials selection and characteristics	Cost Financial arrangements at the level of the project, rather than specific component costs	Deciding Discussing possibilities and ideas, invoking suggestions
Diagram/Mindmap/Flowchart	Manufacturing Consideration of manufacturing, assembly and transport	Time Timescales, deadlines	Managing Includes arranging, directing and instructing. Implies action (such as a response) needs to be taken
External Documents			
Annotated External Docs			
External - CAD			
	Stake Holders Share holders, customers, directors; and their culture and politics	Financial Resources Cash, Assets, Borrowing	Decision Making Considering key factors from evaluation and possible compromises to form a decision
	Economic Issues Costs and efficiency, market and product selection	Knowledge Resources Current ability and stored information	
	Human Resources People, availability, allocation, training etc.	Tools and Methods Specific testing and modelling techniques	
	Physical Resources Ranging from offices to equipment	Practices & Procedures Accumulated by the company, often developed through experience	

Table 1. Classification schema

3 RESULTS

This section presents the main results from the study. First, the key characteristics of the documents are presented, before the other schema categories are examined in detail.

3.1 Information types and top-level categories

Table 2 shows some key characteristics of the documents analysed, along with the number of instances of sketches, calculations and drawings:

Table 2. Key characteristics

	Logbooks	Reports
Number	6	11
Total page volume	540	375
Total Entries	372	405
Average length of entry (pages)	1.45	0.93
% of entries with 2+ info types	33%	18%
Average info types per entry	1.45	1.21
Number of Sketches	124	34
Number of Calculations	52	21
Number of CAD drawings	0	30

The differences in the distribution of information classes is illustrated in Figure 2, below. Where the information class was present in less than 5% of the entries, they have been grouped into ‘all others’ for clarity of presentation:

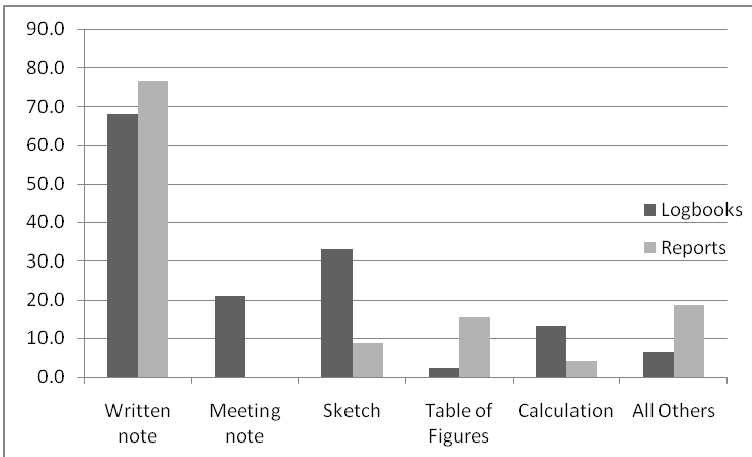


Figure 2. Percentage of entries containing various information classes

It can be seen from the above table and graph that there is a complete absence of meeting notes in the reports (compared to over 20% of logbook entries being meeting notes). Other significant losses can be seen in the number of sketches (although arguably many of the sketches will be manifested in the CAD drawings) and amount of entries containing calculations. It was observed that tables of figures were used more often in the reports to summarise the results of calculations and this is manifested in Figure 2, although this of course means the method used and any possible errors are not apparent in the reports. The ‘richness’ of the entries also differed significantly, with many more logbook entries containing two or more information classes (33% vs. 18% for reports). This is also evident in the figure for the average number of information types per entry (1.45 vs. 1.21 for reports).

However, these statistics alone do not afford a full understanding of the differences between the logbooks and reports. Starting with the top-level terms from the schema, the percentage of entries that contain these terms are shown in Figure 3:

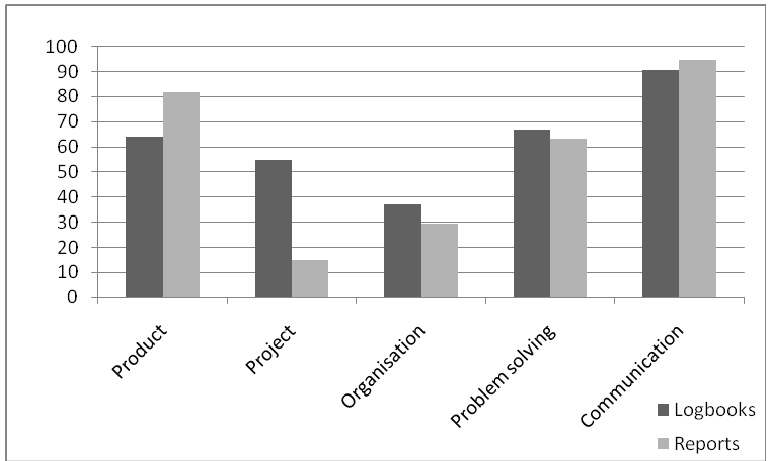


Figure 3. Percentages of entries containing top-level categories

From this, it would appear that the trends between the logbooks and reports are remarkably similar, with the exception of the relative lack of project-related information in the reports, 55% vs. just 15% of entries for reports. As can be seen in Table 1, project-related information covers topics such as managing risk, planning/task allocation, timescales etc. Therefore, it is necessary to drill down into each of these categories to give additional insight into the reasons behind these trends.

3.1 Sub-categories

Taking each top level category in turn, Figure 4 shows the breakdown of product sub-categories:

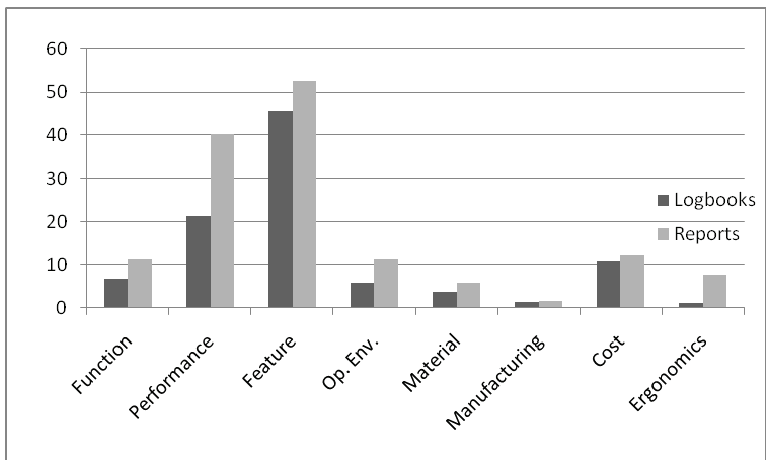


Figure 4. Percentage of entries containing product-related information

The trends across the sub-categories are broadly similar, although proportions of performance and ergonomics related entries (which are related to performance) are significantly higher in the reports. This is not counter-intuitive, as the traditional role of formal reports is to describe the final design

(including its performance), whereas logbooks have a greater role during the *emergence* of the design. The product cost information was observed to be present in both logbooks and reports. However information related to ergonomics was virtually absent from the logbooks, suggesting this was only considered at the end of the project.

The breakdown of project sub-categories also reveals more detail about the large differences in the number of project-related entries mentioned above, as illustrated in Figure 5:

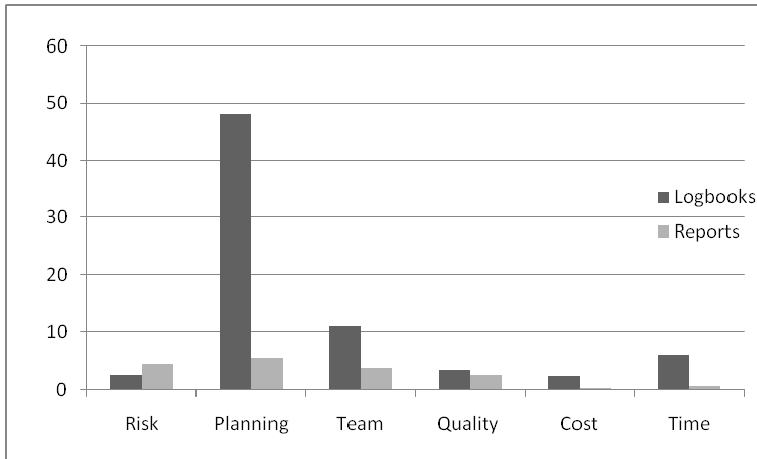


Figure 5. Percentage of entries containing project-related information

It can be seen that most of the difference arose from the relative absence of ‘planning’, ‘team’ and ‘time’ entries in the reports, with nearly 50% of logbook entries containing some elements of planning, compared to only around 7% of report entries. These differences can be attributed to the following factors:

- A significant amount of planning and task allocation (referring to team members) was recorded in meeting notes, which were completely absent from the reports (20% vs. 0%).
- Logbooks were all chronological and often resembled a diary. Thus the logbook was effectively a living document, making planning an integral and natural part of many entries.
- Logbooks were frequently used to track tasks outstanding for the individual in ‘to-do’ style lists at the beginning of entries.

Similar trends existed for organisational-related entries. Economic information was present in both logbooks and reports, as such information was provided by the stakeholders in a relatively clear form during meetings and simply copied from logbook to report. For example, the target cost of the machine was communicated to the engineers in the meeting, and was then transferred in to the formal requirement specification. There is also a clear loss of human resources (HR) related information. Such information was often very similar in nature to ‘team’ information (i.e. dealing with issues related to team members and their roles) and was therefore absent for similar reasons - namely that such issues were often discussed in meetings, records of which are absent from the formal reports:

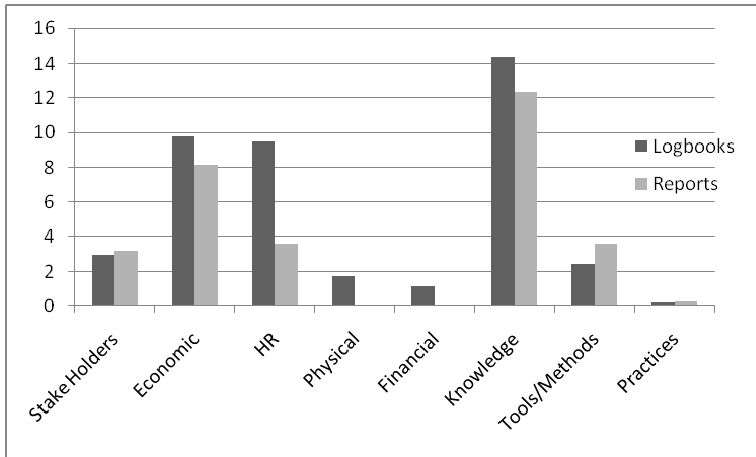


Figure 6. Percentage of entries containing various organisational information

Moving to the problem solving sub-categories (Figure 7, below) it can be seen that entries containing ‘goal setting’ and ‘constraining’ activities are at broadly similar levels. However there are significant differences between the amount of entries containing ‘solving’ and ‘evaluating’:

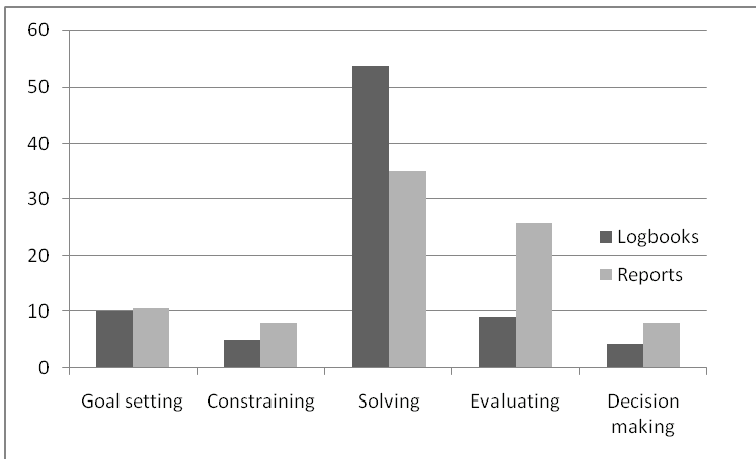


Figure 7. Percentage of entries containing various problem-solving activities

The majority of ‘goal setting’ was observed to be in the requirement specification and scoping section at the beginning of the logbook, which are also repeated in the reports, hence these percentages are very similar. The significant differences in the proportions of entries containing elements of solving and evaluating are probably for a similar reason as the difference in performance-related entries shown in Figure 4, namely that logbook entries are more likely to contain elements of solving – i.e. search, gathering and developing solutions in the earlier stages of design, whereas the reports emphasise the final design, including its evaluation.

The overall percentages for entries containing the top-level ‘communication activity’ (Figure 3, above) was – unsurprisingly – almost 100%. This is simply because virtually all entries can be classified as communicating *something*. However, drilling down into the sub-categories reveals that the *nature* of this communication differs significantly between the logbooks and reports (Figure 8):

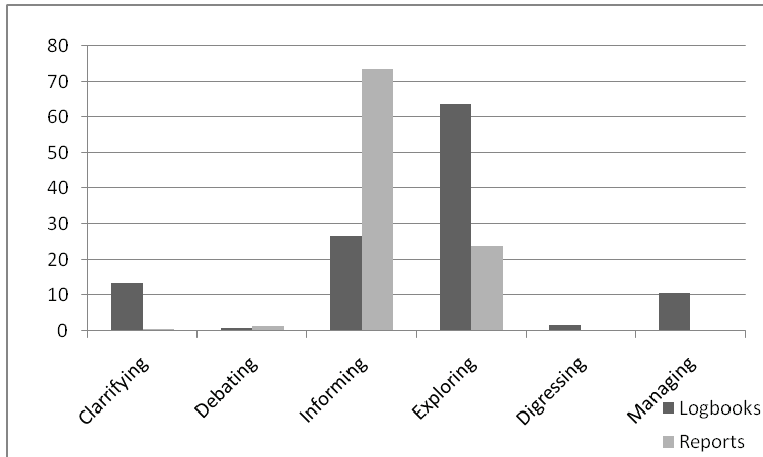


Figure 8. Percentage of entries containing various communication activities

Firstly, there is a lack of ‘clarifying’ and ‘managing’ in the reports. In a similar trend to the planning categories, this may be explained by the observation that clarifying and managing often occur during collaborative work. Such sessions would be recorded in logbooks as meeting notes, which are absent from the reports.

The other significant difference is the contrast between entries that are informative in nature (27% for logbooks vs. 74% for reports) and those which are exploratory, where the trend is almost exactly reversed (64% for logbooks vs. 24% for reports). ‘Informative’ in this context means that the entry states one position, and does not consider options, alternatives or the rationale for the statement, whereas exploratory entries do consider alternatives.

4 COMPARING ENTRIES

There were clear indications of potential ‘information loss’ in terms of information classes, and in particular the numbers of sketches, calculations and meeting notes in this dataset. In the case of meeting notes, much of what was lost is project or organisational-related information (specifically planning, team/HR and managing activities). Other categories or information classes (‘product’, for example) did not appear to show much of a ‘loss’. Indeed, some types of information such as those related to product performance and evaluation, were more prevalent in the formal reports. It is suggested that the greater proportion of entries related to performance and evaluation could be attributed to such activities naturally occurring later in the process when reports are being written, and also that it is the function of a report to concentrate on the final design – and particularly evaluating its performance. This is further supported by the observation that significant differences still exist in the *nature* of these entries, with informal logbook entries containing performance or evaluation aspects still more likely to be exploratory in nature, or relating to the performance of alternatives and not the chosen solution. Overall, then, logbooks appeared to show more of the emergence of the design, as manifested through more logbook entries containing elements of ‘solving’ and ‘exploring’. They were also richer in terms of the number of types of information, the amount of rationale, and especially the amount of project and process-related information. In contrast, reports – as one might expect - placed more emphasis on the evaluation of the performance of the final product/artifact, largely through factual, textual description, tables summarising information and CAD drawings. These differences are clearly illustrated with a representative example in Figure 9, overleaf, which show corresponding entries from the logbook and the report:

5 IMPLICATIONS FOR DESIGN RECORDS

Despite the scenarios discussed above that may reduce the disparity (or even blur the boundaries) between formal and informal records, it is argued that in a significant number of cases, neither the informal nor the formal records are likely to represent a complete record on their own. Reading the formal reports alone (the usual scenario) may lead to an over-emphasis of the performance of the chosen solution, at the expense of rationale about decisions on, for example, the merits of alternatives and the process by which a solution was arrived. There are two main scenarios where a more complete record could be useful:

1. Information retrieval from an organisational perspective – for audit or to support other business processes.
2. Re-use of design information by engineers, either during the project or for another project in the future.

From an organisational perspective, the more complete records of planning and team/HR information are clearly useful for audit and intellectual property management purposes, even in its current form. Such information is particularly significant, as even where meetings are formally minuted, it has been shown that there is a large loss of information, and specifically actions arising [9]. This means that the informal records are often the only source of much of this information and could provide vital evidence of, for example, who was present at meetings, who was responsible for a particular task, or when a certain method or solution was first discussed. It could also be used to support other more routine business processes such as project and knowledge management activities. For example, the project categories presented in Figure 5 shows that information related to ergonomics was largely absent from the logbooks. Whilst this could just be a consequence of the nature of this particular project, routinely identifying such differences could be used to pro-actively identify possible deficiencies in the decision making process.

For the second scenario – re-using design information either during or after a project – being able to assess its value and relevance is a critical factor, which is in turn made possible by being able to understand the context in which it was created. It is arguable that logbooks go some way to providing such additional context - and particularly assumptions and models on which the analysis depends - as illustrated in Figure 9, above. Therefore, as well as being useful during a project to ensure the reports are an accurate, balanced reflection of the work done, re-use of information from past projects could be considerably improved. This is because of the potential for better access to the *product* context and rationale contained in informal sources, as well as through learning lessons from the *process*, which is impossible with the reports alone.

Linking informal records to their more formal counterparts would then appear to be a sensible way forward. However, establishing the best way to create such links is not straightforward. These and other future research issues are now discussed.

6 CONCLUSIONS AND FURTHER WORK

This research was concerned with information contained in informal and formal records. A detailed comparative study was carried out on complete set of representative documentation from a design project about reducing changeover time for packaging machinery. It comprised of six logbooks and 11 corresponding formal reports, plus CAD drawings. These documents were classified against a comprehensive mark-up schema covering information types, product and process-related categories, problem solving and communication intent.

The results revealed a number of significant differences. There was evidence of both loss of some specific types of information such as sketches, and virtually all planning, team/HR and task information. Importantly, there were also significant differences in the *nature* of the entries: logbooks were much more exploratory in nature, with an emphasis on ‘solving’ activities, whilst reports were much more concerned with factual evaluation of the design’s performance.

The clear implication was that access to both sources would produce a more complete - and arguably more useful - design record for both organisational and individual/project re-use in this case. However, it is noted that there are a number of factors relating to the nature of the project and organisation that could affect the usefulness of informal records, as well as other issues that require further investigation. For example, whilst there is clear evidence of ‘information loss’ and examples of where such ‘missing’ information may be useful, quantifying the value of the lost information is difficult –

i.e. is its capture worth the additional effort? It is also acknowledged that whilst in this case, the logbooks, reports and CAD made up the vast majority of the physical record, this may not always be the case. For example, distributed teams often rely heavily on email or other communication and recording tools, the analysis of which was outside the scope of this paper. Finally, how such links between formal and informal records may be established in a manner that is both cost effective and acceptable to the engineers is also unresolved. The richness of logbook entries, combined with their very different nature makes the direct linking of one entry to another in a report problematic – as does the relative inconsistency in the presentation of informal records – and this is an area that requires significant further work.

ACKNOWLEDGEMENTS

The work reported in this paper is part of a PhD sponsored by the EPSRC Innovative Design & Manufacturing Research Centre (IdMRC) at the University of Bath (grant reference EP/E00184X/1) and supported by an industrial CASE award. The authors would like to thank the engineers who provided the logbooks and reports, and James Wasiak and Craig Loftus of the IdMRC, who developed the schema adapted for use in this research.

REFERENCES

- [1] Hicks BJ, Dong A, Palmer R & McAlpine HC. Organizing and managing personal electronic files: A mechanical engineer's perspective. *ACM Transactions on Information Systems*, Vol. 26(4), 2008, pp 23:1-23:40.
- [2] McAlpine HC, Hicks BJ, Huet G & Culley S. An investigation into the use and content of the engineer's logbook. *Design Studies*, 2006, 27 (4), pp.481-504.
- [3] Trafton JG & Trickett SB. Note-taking for self-explanation and problem solving. *Human Computer Interaction*, 2001, 16, pp. 1-38.
- [4] Purcell AT & Gero GS. Drawings and the design process. *Design Studies*, 1998, (19), pp.389-430.
- [5] Grenier AL & Schmidt LC. Analysis of engineering design journal sketches and notations. In Proc. of the ASME 2007 Int. *Design Engineering Technical Confs. & Computers and Information in Engineering Conf.*, September 4th-7th, 2007, Las Vegas, Nevada, USA
- [6] Topi H, Lucas W & Babaian T. Using informal notes for sharing corporate technology knowhow. *European J. of Information Systems*, 2006, 15, pp. 486-499.
- [7] McAlpine HC, Hicks BJ & Culley S. Attempting to investigate the design information loss between logbooks and formal engineering reports. In 'Design rationale: Problems and progress' workshop held during the *2nd Conf. on Design Computing and Cognition (DCC'06)*, 9-13th July, 2006, Eindhoven, The Netherlands.
- [8] Wasiak JO, Hicks, BJ, Newnes L & Dong A. Emails In Engineering Design Projects: An Approach For Analysing Their Information Content. In *5th Int. Conf. on Digital Enterprise Technology*, 22-24th October 2008, Nantes, France.
- [9] Huet G, McMahan C, Sellini F, Culley S & Fortin C. Knowledge loss in Design Reviews. In *Int. Conf. on Integrated Design and Manufacturing in Mechanical Engineering (IDMME)*, 17-19th May 2006, Grenoble, France.

Contact: Hamish McAlpine
University of Bath
Innovative Design & Manufacturing Research Centre (IdMRC)
Dept. of Mechanical Engineering
Bath, BA2 7AY
United Kingdom
Phone: +44(0)1225 384 166
H.C.McAlpine@bath.ac.uk
<http://www.bath.ac.uk/idmrc>

Hamish McAlpine is currently completing a PhD on the subject of improving the re-use of informal engineering information through electronic logbooks.