INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED13
19-22 AUGUST 2013, SUNGKYUNKWAN UNIVERSITY, SEOUL, KOREA

THE LANGUAGE OF COLLABORATIVE ENGINEERING PROJECTS

Ben HICKS
University of Bristol, United Kingdom

ABSTRACT

The challenges of managing large, long-life, distributed engineering projects are discussed and the need to improve monitoring and control in order to reduce over-runs, improve productivity, better manage IP and monitor risk is highlighted. In order to achieve this, it is proposed that the outputs of the communications and the digital objects generated as part of the project are fundamentally related to performance and that analysis of their content can provide understanding, insights and predictions about the project. In order to explore this proposition an exploratory study of the content of a project email corpus and its relation to the project schedule and project performance is presented. This study demonstrates a series of eight trends (termed signatures) between longitudinal traces of problem solving, information transactions and management that corresponded to project states and modes of management intervention. Although the project-related rationale accounting for the signatures seen in this project may not accurately characterize all projects, their alignment with actual events points to the potential value of email content for improving project management.

Keywords: design management, knowledge discovery, collaborative design, content analysis, email

Contact:
Prof. Ben Hicks
University of Bristol
Mechanical Engineering
Bristol
BS8 1TR
United Kingdom
ben.hicks@bris.ac.uk

1 INTRODUCTION

The activity and practice of globally distributed design and manufacture has now emerged as a fundamental characteristic of modern engineering. As such engineering work is highly distributed, multi-national and heavily dependent upon electronic communication and digital objects that define the engineered product, the process by which it is designed and the process by which it will be manufactured.

One of the key drivers for this shift – hitherto based on cost alone – is now increasingly the geographic availability of expertise and skilled personnel. For example, in a recent address Tom Enders, Airbus's CEO, highlighted the relative intellectual disarmament of the UK/EU and the increasing importance of development teams in India and China (Enders, 2011).

It is not only the increasing globalisation of design and manufacture that complicates the delivery of engineering products. So too does the complexity that in a variety of forms is increasingly present in today's artefacts and systems, ranging from large, long-life, multi-domain engineering systems to consumer products and software.

This highly distributed nature of modern engineering combined with the complexity of today's engineering artefacts mean that a multitude of electronic communication tools and digital objects are now employed. The communication tools include email, instant messaging, video conferencing and social networking and the digital objects include, for example, spread sheets, CAD models and specialist simulation models. It is the basic premise of this paper that the outputs of these tools (the communications and the digital objects) are related in a number of fundamental ways that are not currently understood, but could provide insights and understanding which can aid engineering management.

By way of examples, a small machine or software project (<£1M) can involve 20+ contributors (engineers from various disciplines, customers, subcontractors, administrators, etc.) generate 20,000+ emails, 3,000+ reports and presentations, hold 500 meetings, generate 1,000+ models (versions) and 40 prototypes (Regli, 2010). In contrast, design, construction and commissioning of a building can span 5 years, involve 100s of project members, 100,000+ emails, 15,000 reports and presentations, 2,000+ meetings and 5,000+ models/representations (Watson, 2012).

In order to explore the aforementioned proposition this paper describes an exploratory study that examines the content of a project email corpus and its relation to the project schedule and project performance. The paper begins with background on modern engineering and the challenges of managing engineering projects. The paper then summarises the experimental method and results of the exploratory study. The findings are then considered with respect to the proposition and implications for further research as well as the next generation of project management and product data management systems.

2 MODERN ENGINEERING

As previously stated modern engineering is critically dependent upon electronic communication and digital objects, which have, exploded in terms of their: prevalence of use, volume of content, variety of type and overall numbers. While this explosion has been necessary and beneficial at the detailed application level, it has resulted in overload of information and communication, and fragmentation across individual and organizational digital objects and records with different access and ownership rights. Additionally, the communication and information evolve very rapidly and often, seemingly chaotically, across organizations and teams meaning that no individual or management group can be continuously up-to-date.

The consequences, in the context of complex engineering projects, are that: *potential issues can be almost impossible to identify early and mitigate; progress monitoring, control and performance measurement are all but impossible; and opportunities to innovate and maximize value are seldom pursued.* Thus, effective management and control of collaborative engineering projects and engineering work is highly challenging and problematic.

The challenges of collaborative engineering concern all sectors from civil, aerospace, automotive and pharmaceuticals, to the creative industries. Recent high-profile examples of engineering projects that have experienced problems include the new Royal Naval aircraft carrier which has been affected by contractual and cost issues (Anon, 2011). Cost overruns were also experienced in the development of both the Airbus A380 (2-year overrun and 2 billion euro overspend (Anon, 2009)) and the Boeing

Dreamliner (over two years late and $10 billion overrun (Drew, 2009)). The importance and impact of the challenges of distributed design and manufacture of complex products is set out in a recent report by the US National Science Foundation which reported that the total value of delay and cost overruns stands at $150 million each day for the US Department of Defense alone (NSF, 2010). While such figures are unavailable for the UK it is likely that a similar relative magnitude of cost is incurred by UK industry.

UK manufacturing is at a critical point in its evolution where collaboration involving distributed aspects of design and manufacture is demanded. These demands are felt by both OEMs who require a supply of expertise and resources and engineering service providers who manage projects and/or provide resources to overseas manufacturers. Further, the new Patent Box tax changes (HMRC, 2012) will start to shift emphasis in the manufacturing space to the generation of new products and associated IP in the UK. The consequence of this is that control and analysis of IP embedded in communication and the associated digital objects dealt with in this proposal will become evermore critical.

It follows that dimensions of management and control include but are not limited to: team cohesion; effectiveness of collaboration and co-creation of digital objects; the control of intellectual property; decision making and rationale capture; uncertainty and problem solving; interface negotiation and concessions; contractual agreements; risk; and costing. In addition there are also implications for completeness of design records, access and reuse of design records and learning from previous projects. It is the issues surrounding adverse performance/control across these dimensions that the research reported in this paper will begin to remedy.

3 EXPLORATORY STUDY

The aim of the exploratory study was to examine the content of project email with respect to project performance to explore potential relationships. In order to analyze the content of email, a directed, qualitative content analysis method was employed. The development of the methodology and corresponding coding scheme is described in detail in Wasiak (2010). The aim of the content analysis is to understand what topics email discuss, why email is sent, and how the content is expressed. The topics of email were classified into three categories: Product, Project and Company. These categories were derived from a literature review of generally accepted taxonomies for the classification of the lifecycle of engineering design. To categorize why email is sent, that is, communicative intent, three categories were used: problem-solving, information transactions and management transaction.

- Problem-solving activities relate to intent to respond to a problem by proposing alternative solutions, evaluating them, and making a decision.
- Information Transactions deal with intent to express, request, inform or record information.
- Management Transactions deal with intent to guide activities.

In order to examine how the content is expressed and its effect on social aspects that affect the project group’s ability to achieve its goals, Bales’ content analysis scheme for the analysis of interaction in small-group communication was used (Bales, 1950). Bales’ scheme focuses on the social and emotional content of communication, making it possible to take into account the positive or negative social effects of an email.

The overall coding scheme is given in Table 1, which highlights the top-level categories of What topics email contains, Why email is sent, and How the content is expressed. There are eight sub-categories: Product, Project and Company (What); Problem-Solving Activities, Information Transactions and Management Transactions (Why); and, Social Interaction (How). Within each of these sub-categories, there are up to a dozen descriptors. The definitions of the categories and descriptors are given in full by Wasiak et al. (2010) and are an important element in reducing the influence of subjectivity and improving inter-coder reliability. The dimensions of scope, reliability and validity of the coding scheme were assessed through iterative testing and refinement of the categories and sub-categories and their descriptive terms, until a comprehensive yet practically applicable coding scheme was achieved. The iterative testing involved the researchers applying the candidate schemes both independently and jointly to sample emails. The resulting coding scheme is summarized in Table 1.

Table 1. Summary of the categories of the email coding scheme

Top level categories									
What topics the email is about			*Why* the email has been sent			*How* the email content is expressed			
Subcategories									
1. Product	2. Project	3. Company	4. Information	5. Management	6. Problem Solving	7. Socio Emotional		8. Task Related	
Descriptors									
Features	Risk	Stakeholders	Informing	Managing, Confirming	Goal setting	Shows solidarity	Shows antagonism	Gives opinion	Asks for opinion
Function	Plans	Economic issues	Requesting		Constraining	Shows tension release	Shows tension	Gives suggestion	Asks for suggestion
Ergonomics	Team	Financial resources	Clarifying		Developing solutions (solving)	Agrees	Disagrees	Gives orientation	Asks for orientation
Cost	Quality	Human resources			Evaluating				
Performance	Cost	Physical resources			Decision making				
Materials	Time	Knowledge resources			Reflecting				
Manufacturing	Manufacture	Tools and methods			Debating				
Specification	Delivery	Practices & Procedures			Exploring				
	Milestones								
	Contracts								
	Documents & Resources								
	Administration								

3.1 The email corpus

The email corpus comprised 16,000 emails sent over a four-year period from 650 senders to 1080 recipients. For the purpose of analyzing the overall content of the email corpus and due to resource limitations, every 20th email was sampled in chronological order, equating to around 200 per year and one for each working day – assuming 225 working days per year. In total, 800 emails of between 19 to 900+ words were coded. On average, an email took between 5-15 minutes to read and code, equating to more than 200 hours of analysis. Sampling email in this manner enabled a reasonably uniform distribution of email volume from each stage in the project over its lifecycle.

Each email was coded by starting with the top-level categories of the scheme and progressing through the sub-categories with the aim of classifying units of information in an email, for which the smallest unit is a sentence and the largest unit is the entire email. Further, no sub-categories are mutually exclusive and multiple categories can be assigned to an email and to units of information.

Training in accordance with standard practices in content analysis was undertaken prior to the email analysis to ensure that the coders learned the categories and worked with each other to arbitrate differences. This included a one week training period and joint coding of a subset of 60 emails. Two researchers coded the emails according to the scheme, including duplicating sets to allow inter-coder reliability over time to be calculated. When the coders disagreed, they arbitrated their codes. The stability of the coding was found to have an acceptable coefficient greater than 0.8. The split half technique was used to confirm the statistical significance of the subset of emails. A coefficient of more than 0.8 was achieved. Finally, the inter-coder reliability was determined with two coders marking up the same batches of emails. Cohen's kappa coefficient (Cohen, 1968) for inter-coder reliability for the categories that constitute the basis of the results are reported in Table 2 (Wasiak et al, 2010) with values that are sufficiently high (>0.7) for the exploratory nature of this study (Lombard, 2002).

Table 2. Inter-coder reliability for sub-categories using Cohen's kappa coefficient

Code	Coefficient
Product	0.87
Project	0.85
Company	0.81
Information Transactions	0.82
Management Transactions	0.76
Problem Solving Transactions	0.71

The post processing of the content analysis involved summing the codes by frequency and then aggregating the codes by time, to reveal how the email dynamic changes throughout the project life-cycle.

3.2 Project performance data

In order to be able to compare the content of email with project performance data was collected from semi-structured interviews and from project documentation. This included the project schedule and the problems experienced during the project. Semi-structured interviews were conducted with all project members. A set of prepared questions was used during interviews to ensure that key areas were covered. In addition to exploring the roles of project personnel and background of the project, the interviews sought to elicit the difficulties experienced during the various phases of the project and their cause. Participants included engineers, middle and top tiers of project management.

3.3 Project characteristics

The project was a long-life systems engineering project for the marine sector. The project included design, manufacture, assembly, testing and commissioning (service) which necessarily demanded communication between different organizations and inter-disciplinary teams that were geographically dispread. The value of the contact was circa £6M. The contract included six sub-systems which were intended to be based on the same design, although some variation in the designs was required due to specific requirements of end users.

The six sub-systems were deemed to be highly novel and represent a reasonable risk to the company. The company sought to undertake such a venture to gain understanding prior to other competitors in the field. The company was familiar with the domain of operation, and much of the design work involved adaptations and variations to existing sub-systems. There were, however, some technical aspects which the company had not dealt with previously and as such these represented original design. The first unit required the most financial investment and time, taking three to four years from conception to service, while the five following variants lagged by around one to two years.

The project required interdisciplinary engineering design from a variety of domains including software, systems and mechanical engineering. Teams based at two geographically dispersed locations undertook the project, with assembly taking place at a third location. One location addressed systems and software design and the other mechanical design. Each location had its own Project Manager who was accountable to the Project Director.

Emails were normally sent to and from individual user's personal mailboxes. The company policy was to copy 'relevant' email into a central repository. The judgment of relevance was left to the individual. It was also possible for users to send email directly from the central repository thus creating a stored version therein. At the end of the project over 16,000 emails were contained in the repository, and it is this corpus that is the basis for this study.

4 RESULTS

The results present in this paper are reported in full in Wasiak et al (2011) and summarized herein got the purpose of evidencing the existence of potentially relevant engineering signatures. The signatures are established by cross-referencing the email dynamics with project documentation and in particular the seven stages of the project. These seven stages included:

1. Specification - where the overall system design is specified with respect to agreed

functional, performance and physical requirements.
2. Sub-system manufacture - during which sub-system designs are finalised and manufactured.
3. Sub-system testing - involving performance and integrity testing of individual sub-systems.
4. Sub-system delivery - where sub-systems are delivered to a single location.
5. Assembly – involving the assembly and integration of sub-systems.
6. Testing – involving full systems testing within an operational context.
7. Service – full commissioning and in-service operation of the system.

For each of these stages a time-phased analysis of the number of emails relating to the dimensions of *What* and *Why* was performed. Figure 1 shows traces of the time-phased numbers of email concerning information sharing, management and problem-solving. Figure 2 shows traces of the time-phased numbers of email concerning the product, project and company.

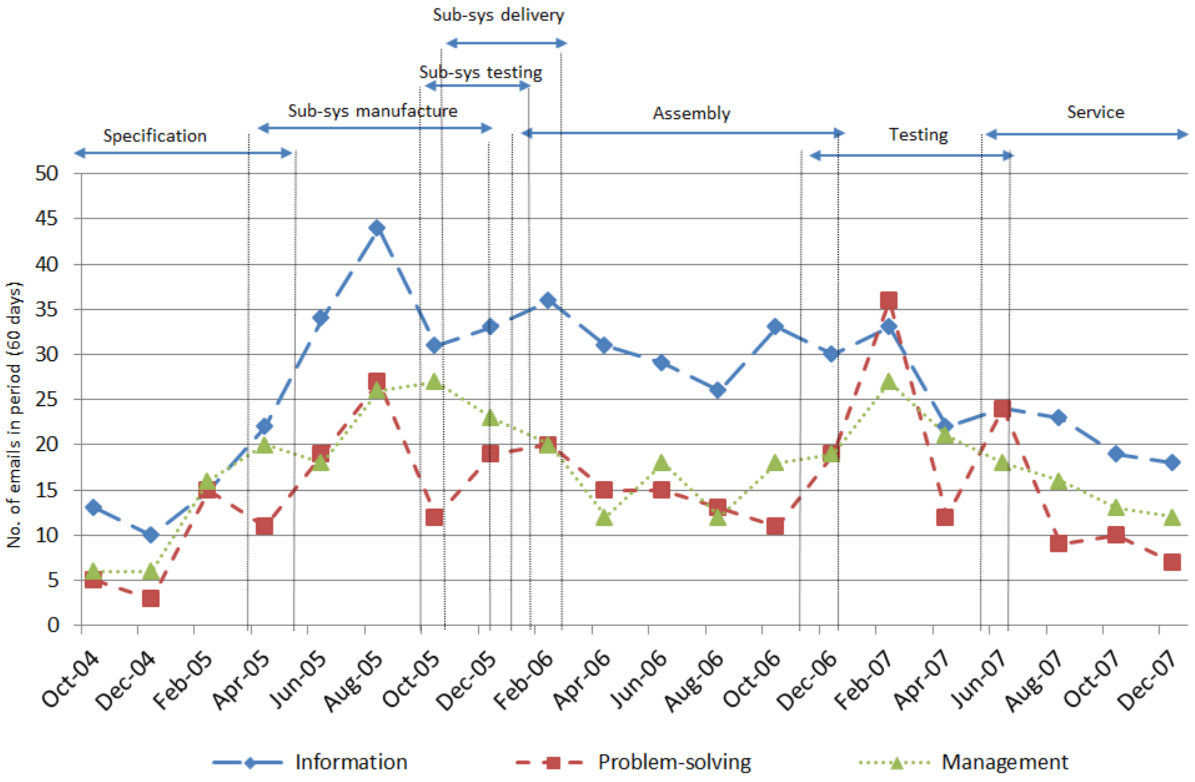

Figure 1. Time-phased numbers of email concerning information sharing, management and problem-solving

For the purpose of interpreting the email dynamic the traces can be used to infer meaning for project management. The approach taken in this paper is to examine the relationship between the traces and, in particular, their relative vectors of change over a time-phased period. These combinations of vectors of change (rising and falling) can be thought of as signatures. In the case of the three traces in Figure 1 there are eight possible combinations of traces and hence eight signatures. These signatures are listed in Table 3 and discussed below with respect to the project analysed and their possible interpretation.

The first signature represents a simultaneous rise in all three traces. This occurred in two phases of the project: June to August 2005 and January to March 2007. Over these periods product related emails also increased (Figure 2). These periods were identified as key pressure points in the project. The simultaneous and sudden rise of the three traces combined with increasing rates of product and project email can be considered a signature of increasing work activity necessary to address the problems experienced during the transition from design to manufacture.

The second signature is represented by a rise in management and information sharing email, and a decrease in problem solving email, and occurred during the periods September to November 2006 and February to April 2005. When considering the Project Schedule these periods correspond to phases of steady work progress or completed work. The period involved assembly and test set-up, and as such there was a significant amount of email discussion compared to information sharing and coordination of work. Notwithstanding this, a decrease in problem-solving email can be seen at the end of each stage. This is expected since at the end of each stage there is a need to share information of what work had been completed and for management to verify milestones and contractual obligations.

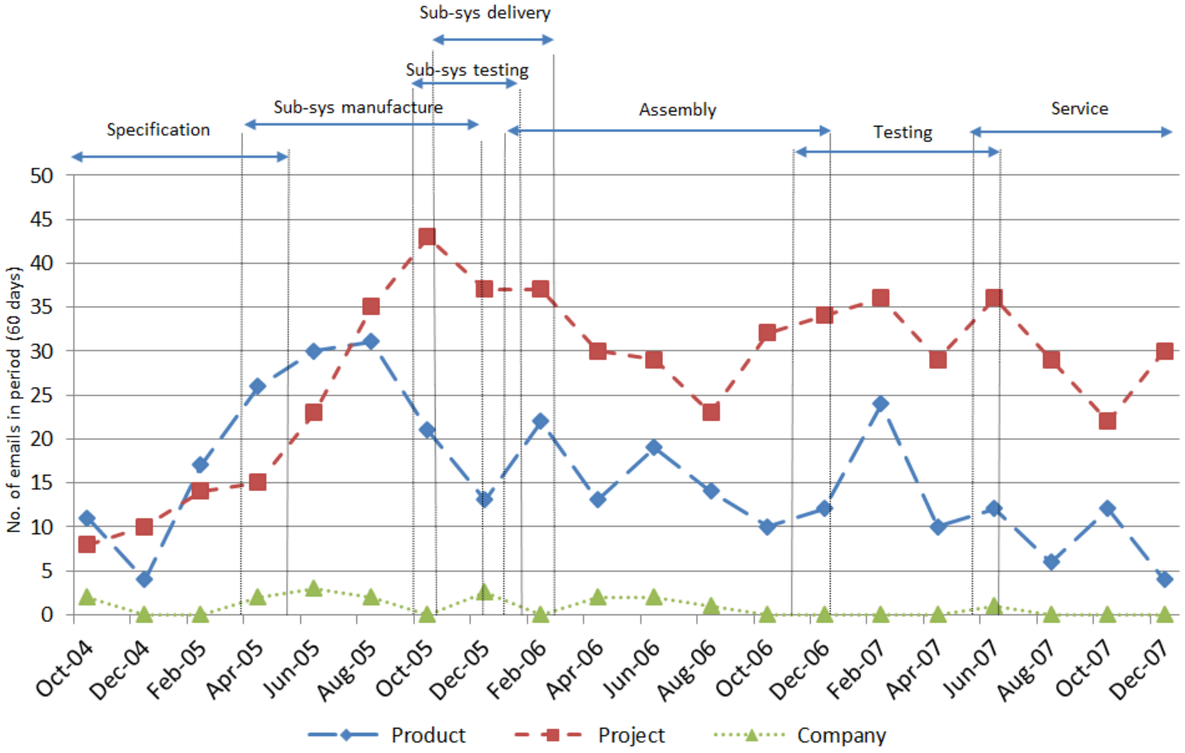

Figure 2. Time-phased numbers of email concerning the product, project and company

The third signature is represented by an increase in the level of management email while information sharing and problem solving traces are decreasing. This occurred for a brief point between May to July 2006. The project performance data indicated that management intervention was necessary at this time to direct the project.

The fourth signature is represented by all traces (problem-solving, management, information sharing and product and process) decreasing. This occurred during the Assembly stage and is arguably the most desirable state, given that management email is also decreasing i.e. it is a predictive signature that the project is progressing steadily and without concern.

The fifth signature is represented by a rise in information and problem solving traces and a fall in management email. This occurred from October 2005 to February 2006 and again from May 2007 to June 2007. These periods were identified from the project performance data as ones of effective working, where individuals communicated only to complete their work and there was correspondingly minimal management intervention.

The sixth signature is represented by brief increases in the management trace and problem-solving trace. These were evident throughout the project and reflect management intervention necessary to ensure that contractual obligations were being met, even when all necessary information for work to proceed was available.

The seventh signature concerns an increase in problem-solving without a corresponding increase in management or information sharing email and was observed between October and November 2007. At this point, the product was entering service delivery, and the email traffic confirms that the product met contractual obligations with only minor commissioning problems. This suggests that the direction was clear and that there was sufficient work to progress.

The eighth signature is represented by an increase in the level of information sharing with a corresponding decrease in problem solving and management. This signature was not observed and is likely due to the fact that management and problem-solving emails generally also share information.

Table 3. Signatures - rises and falls in information, management and problem solving transactions and their interpretations

	Management	Information	Problem Solving	Interpretation	Date observed
S1	↑	↑	↑	Pressure point, with real coordinated work achieved.	6/05 – 8/05 & 1/07 - 3/07
S2	↑	↑	↓	Work completing, and status updates being sent.	9/06 – 11/06 & 2/05 – 4/05
S3	↓	↑	↑	Work is being completed with little management coordination.	10/05 - 2/06 & 5/07 – 6/07
S4	↓	↓	↓	Steady mode of working.	3/06 – 9/06
S5	↑	↓	↓	Management impetus to control and coordinate work.	5/06 – 7/06
S6	↑	↓	↑	Information sufficient for work to progress but management control needed.	11/06 – 1/07
S7	↓	↓	↑	Direction clear and information sufficient for work to progress.	10/07 – 11/07
S8	↓	↑	↓	Information required to continue work requiring management intervention to improve information flow.	Not observed

5 DISCUSSION

Although the project-related rationale accounting for the email traffic signatures seen in this project may not accurately characterize all projects, their correlation with actual events in this case study point to the potential value of email content in a project management context. This finding support, in part, the original proposition of this paper, that there exists a relationship between the output of engineering work activity and project performance and that this relationship is implicitly embodied in the content of communications and digital objects. Although the content of digital objects are not considered it would seem logical that their analysis and combination would add further utility to the analysis of the content of email. By understanding the relationship between communications and digital objects it may be possible to establish characteristic signatures which represent a point in time or an aggregation of traces over a period, e.g. for an activity or phase. These signatures may themselves embody or enable understanding, or allow insights to be achieved. However, it is contended that *patterns* of relationships between signatures of communication and signatures of digital objects could provide the greatest insights and learning. For example, it may be possible to generate a project profile through aggregation of signatures over stages that affords the ability to compare and contrast 'live' projects with previous projects. An extension of this concept is that it may be possible to drive a generic set of *signatures* and *patterns* applicable to all engineering domains, as well as sets of *signatures/patterns* applicable to particular engineering domains.

While the ability to extract such signatures in real-time has yet to be achieved from 'big data', emerging new approaches for sense-making are beginning to enable real-time analysis of communication. Such approaches include Natural Language Processing, semantics-based inference, collaborative and social technologies, and collective intelligence. A recent example includes the use of Twitter feeds in conjunction with a mood-gauging algorithm to predict money markets and box office takings (Tech Review, 2012).

It is further contended that generating the aforementioned understanding could provide the basis for the creation of new ICT approaches such that representations of the content and co-evolution of the record (communication and digital objects) can be generated, interacted with and interpreted by project stakeholders to enable advance notification/warning of issues, improved management, control, increased productivity (individual and collaborative) and ultimately improved design and manufacture of the product.

This proposition also has implications for the next generation of 'intelligent' product data management systems and radically different project management methodologies that are agile in terms or their process monitoring and control.

6 CONCLUSIONS

The challenges of managing large, long-life, distributed engineering projects have been discussed and the need to improve monitoring and control in order to reduce over-runs, improve productivity, better manage IP and monitor risk has been highlighted. In order to achieve this, it was proposed that the outputs of the communications and the digital objects generated as part of the project are fundamentally related to performance and that analysis of their content can provide understanding, insights and predictions about the project. In order to explore this proposition an exploratory study of the content of a project email corpus and its relation to the project schedule and project performance has been presented. This study demonstrated a series of eight trends (termed signatures) between longitudinal traces of problem solving, information transactions and management that corresponded to project states and modes of management intervention. Although the project-related rationale accounting for the signatures seen in this project may not accurately characterize all projects, their alignment with actual events in this study point to the potential value of email content in a project management context. While the study did not consider the analysis of the content of the digital objects, it would seem logical that supplementing these signatures with an evaluation of, for example, CAD model activity, would add utility and further value for project monitoring and management.

It is arguable that such knowledge discovery is the logical next step in product data management systems and may provide the foundations for new more agile models of engineering project management.

ACKNOWLEDGMENTS

The research reported in this paper is derived from two Engineering and Physical Sciences Research Council funded projects (Grant refs: EP/E00184X/1 and EP/K014196/1). In particular, the results of the exploratory study are extracted from the PhD thesis of James Wasiak who worked on the Project entitled: Understanding the information content of emails for improved information management and collaboration in large engineering projects. The project was led by Professor Ben Hicks with Professor Andy Dong (University of Sydney) and Dr Linda Newnes (University of Bath). The hypothesis of engineering management relevant signatures is being investigated under EP/K014196/1 entitled The Language of Collaborative Manufacturing. This project commenced in 2013 and is led by Professor Ben Hicks and Chris McMahon (University of Bristol) with Professors Steve Culley, Peter Johnson and Steve Payne, and Drs Linda Newnes and Leon Watts (University of Bath). The Author would like to express his thanks to both the EPSRC, the aforementioned colleagues and the industrial collaborators.

REFERENCES

Anon. (2011) "Queen Elizabeth class aircraft carrier", Dec. 2011, http://en.wikipedia.org/wiki/Queen_Elizabeth_class_aircraft_carrier.

Anon. (2009) "Airbus A380," Wikipedia, retrieved 2 September 2009, http://en.wikipedia.org/wiki/Airbus_A380.

Bales R. F., (1950) "A Set of Categories for the Analysis of Small Group Interaction," *American Sociological Review*, vol. 15, pp. 257-263.

Cohen, J., (1968) "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 70, pp. 213-220, October 1968.

Drew, C., (2009) "A Dream Interrupted at Boeing", The New York Times, 5 September 2009, http://www.nytimes.com/, ret. 6 Sept 2009.

Enders, T., (2011) "Airbus CEO address at SAE Aerotech Conference", Toulouse 18 October 2011.

HMRC (2012) Corporation Tax Reform: Patent Box, http://www.hmrc.gov.uk/budget2012/tiin-0726.pdf.

Lombard, M., *et al.*, (2002) "Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability," *Human Communication Research*, vol. 28, pp. 587-604, 2002.

NSF (2010) “NSF workshop on engineering systems design”, Feb 22nd-24th 2010, retrieved Jan 2012, http://vddi.org/ESDW-report.pdf.

Regli, W., (2010) “Digital Engineering Archives“, January 2012, http://www.digitalpreservation.gov/partners/documents/digarch-regli.pdf.

Tat, A., (2002) “Visualising human Dialog”, *Proceedings. Sixth International Conference on Information Visualisation*, 2002, pp. 16 – 21.

Tech Review (2012) “Twitter Mood Predicts The Stock Market”, MIT retrieved 2012, http://www.technologyreview.com/blog/arxiv/25900/

Wasiak, J., Hicks, B., Newnes, L., Dong, A. and Burrow, L., (2010) “Understanding engineering email: The development of a taxonomy for identifying and classifying engineering work”. *Research in Engineering Design*, 21 (1), pp. 43-64.

Wasiak, J., Hicks, B., Newnes, L., Loftus, C., Dong, A. and Burrow, L., (2011) “Managing by E-Mail: What E-mail Can Do for Engineering Project Management”. *IEEE Transactions on Engineering Management*, 58 (3), pp. 445-456.

Watson, J [Director of Global Research Arup] (2012) “Keynote address at the University of Bath”, July 2012.