

IDIMC 2014 Making Connections

Loughborough University
17th September 2014

Programme Book



Contents

Welcome.....	1
Programme	2
Invited speakers	3
Peter Cooke: Can you make “Agile” work with a global team?.....	3
Mark Harrison: Open Data and Linked Data – how can they help your organisation? (Opening Keynote).....	3
Sue Newell: A new Era of Knowledge Management? Reflections on the implications of ubiquitous computing.....	4
Jacky Swan: The role of social networks in mobilizing knowledge	4
Presentation abstracts	5
UK Data Service: creating economic and social science metadata microcosms.....	5
The implementation of Basel Committee BCBS 239: An industry-wide challenge for international data management.....	8
Adopting a situated learning framework for (big) data projects	11
Learning about information management across projects	13
Exploring different information needs in Building Information modelling (BIM) using Soft systems methodology	17
Using Big Data in Banking to Create Financial Stress Indicators and Diagnostics: Lessons from the SNOMED CT Ontology	18
The use of ontologies to gauge risks associated with the use and reuse of E-Government services.....	20
Assessing trustworthiness in Digital Information.....	23
Managing the risks of internal uses of Microblogging within small and medium enterprises.....	25
Re-purposing manufacturing data: a survey	27
Exploring vendor's capabilities for cloud services delivery: A case study of a large Chinese service provider.....	29

Poster presentation abstracts	32
The Applicability of Western Type E-commerce to Saudi Arabian Entrepreneurship.....	32
A digital Identity Management Model for the Jordanian online press and publication law to sustain the national e-commerce strategy.....	35
Developing an information security policy framework for Saudi Arabia and Middle Eastern countries.....	38
A framework for identifying suitable cases for using market-based approaches in industrial data acquisition.....	43
Developing Data Quality Metrics for a Product Master Data Model.....	45
A taxonomy of crowdsourcing systems.....	47
Balancing Big Data with Data Quality in Industrial Decision-Making.....	49

Welcome

Welcome to the inaugural International Data and Information Management at Loughborough University. IDIMC2014 is being hosted by the Centre for Information Management in collaboration with the British Computer Society's Data Management Specialist Group and it is our great pleasure to offer you an exciting and high quality programme.

This year's conference theme is 'making connections'. Academic work can be criticized for operating in silos and failing to bridge disciplines in a way that practitioners face on a daily basis. The information society and knowledge based economy rely on the organization and retrieval of data and information; the processes associated with knowledge creation; and the knowledge required to design, develop and implement solutions that enable the exploitation of knowledge, data and information. However, it is only when the contribution of these strands of important research are combined and integrated that their influence has the power to make breakthrough impacts on the information society and knowledge based economy. We hope that this conference will help to build bridges between academics and practitioners through sharing knowledge and exploring opportunities to connect disciplines and theories.

We would like to express our deep appreciation for everyone who has helped make the conference possible. First and foremost, to the organizing team Mark Hepworth, Claire Creaser, Philip Woodall, Louise Cooke, Kristin Meredith-Galley, David Gerrard, Alex Von Lünen, Lesley Chikoore for their hard work and in particular, to Sharon Fletcher, Ruth Cufflin and Ondine Barry for their crucial administrative support; to the members of the programme committee Christine Borgman, Guy Fitzgerald, Bob Galliers, Michael Myers, Reijo Savolainen, Philip Woodall, for their help with reviewing and promoting the conference; to the BCS Data Management Specialist Group and Facet Publishing for their support; and to you for being part of this new conference!

We look forward to having a successful IDIMC2014 with plenty of constructive and inspiring debates and, of course, fun and new friendships. We sincerely hope that you enjoy your visit to our campus, that you enjoy the conference, and that you take home many wonderful research ideas!

Tom Jackson, Director,
Centre for Information Management



Crispin Coombs, Deputy Director,
Centre for Information Management and Organizing Chair



Programme

09:00 *Registration and coffee*

09:30 Welcome to CIM *Prof Tom Jackson, CIM Director*

09:45 **Opening Keynote** *Mark Harrison, University of Cambridge*
Open Data and Linked Data – how can they help your organisation?

Parallel 1 – A (room 0.17)

10:30 The implementation of Basel Committee BCBS 239: An industry-wide challenge for international data management
Malcolm Chisholm

10:50 Using Big Data in banking to create financial stress indicators and diagnostics: Lessons from the SNOMED CT ontology
Alistair Milne

11:10 Re-purposing manufacturing data: a survey
Philip Woodall

Parallel 1 – B (room 0.18)

Assessing trustworthiness in digital information
Laura Sbaffi

Exploring vendor's capabilities for cloud services delivery: A case study of a large Chinese service provider
Gongtao Zhang

Exploring different information needs in Building Information modelling (BIM) using Soft systems methodology (SSM)
Mohammad Mayouf

11:30 *Refreshments and poster session*

12:00 The role of social networks in mobilizing knowledge *Jacky Swan, University of Warwick*

12:40 *Lunch and poster session*

13:40 A new Era of Knowledge Management? Reflections on the implications of ubiquitous computing *Sue Newell, University of Sussex*

14:20 Can you make “Agile” work with a global team? *Peter Cooke, Ford Motor Company*

15:00 *Refreshments and poster session*

Parallel 2 – A (room 0.17)

15:30 UK Data Service: creating economic and social science metadata microcosms
Lucy Bell

15:50 Adopting a situated learning framework for (big) data projects
Martin Douglas

16:10

Parallel 2 – B (room 0.18)

Managing the risks of internal uses of Microblogging within small and medium enterprises
Soureh Latif Shabgahi

Learning about information management across projects
Sunila Lobo

The use of ontologies to gauge risks associated with the use and reuse of E-Government services
Onyekachi Onwudike

16:30 Best paper and poster awards *Tom Jackson and Crispin Coombs*

Closing remarks

17:00 *Close*

Invited speakers

Peter Cooke

Global Enterprise Architect, Ford Motor Company

Can you make “Agile” work with a global team?

Agile development methodologies arose as a way to deliver value faster – but assumed that developers and customers were collocated. However, is it possible to get the benefits of Agile in a global environment where business stakeholders and IT personnel are geographically dispersed around the world?

This lecture briefly describes the history of Ford Motor Company’s experiences in global IS development before summarizing research into virtual requirements elicitation. The final section of the lecture will report on how this research is currently shaping new practices being deployed at Ford.

Mark Harrison

*Director Auto-ID Lab, Distributed Information and Automation Laboratory,
Institute for Manufacturing, University of Cambridge*

Open Data and Linked Data – how can they help your organisation?

(Opening Keynote)

Open Data and Linked Data are somewhat complementary to Big Data but are technologies that companies can already use today to tackle some of their Big Data challenges. Increasingly, government departments and agencies are publishing open datasets about government spending, local socio-economic information, geographic information, weather forecasts, roadworks, public transport, health, etc. and there are opportunities to combine this freely available data with our internal data to increase its value and help us to make better decisions, using the additional context information provided by these open data sets. There is also an opportunity to make the public-facing data about the products and services provided by your company available as structured open data on the web, so that your company and its products and services can be more easily discovered by search engines, smartphone apps and other software. Linked Data technology (also known as Semantic Web technology) provides the capability to publish structured data together with its semantics, to perform federated queries from multiple data sources, both local and remote – and to navigate end-to-end through a number of data linkages across those datasets. Some illustrations will be provided about how Open Data and Linked Data are being used today and in the future.

Sue Newell

Professor of Information Systems, University of Sussex

A new Era of Knowledge Management? Reflections on the implications of ubiquitous computing

This presentation will focus on how changes in IT (specifically the increasing use of social software and more generally the digitization of our everyday lives) is changing organizational approaches to knowledge management. For example, organizations are increasingly relying on 'the crowd' to perform the kind of knowledge work that was previously done internally and they are using big data to examine connections and from this make predictions rather than rely on expertise and understanding of knowledge workers. The presentation will focus on how this is impacting knowledge work and more generally learning within and across organizations.

Jacky Swan

Professor of Organisational Behaviour, University of Warwick

The role of social networks in mobilizing knowledge

The presentation will focus on the importance of understanding social networks and relationship between different kinds of social networks and knowledge processes (transfer, translation, transformation).

Presentation abstracts

UK Data Service: creating economic and social science metadata microcosms

Lucy Bell

Functional Director Data Access, UK Data Archive, University of Essex

Purpose or background

The UK Data Service is a comprehensive resource funded by the ESRC to support researchers, teachers and policymakers who depend on high-quality social and economic data. The Service disseminates data via around 67,000 downloads per year to its 22,000 registered users, worldwide.

In the context of open and linked metadata, this paper describes the technologies and metadata management processes that the Service employs across its four academic sites in order to ensure timely and effective information retrieval and compares this with some of the issues surrounding open data.

Design, methodology or approach

One of the Service's primary goals is to make the data it curates discoverable. Without the careful organisation and linkage of metadata, information retrieval suffers, and users may fail to discover data critical to their research.

Connecting metadata within a Knowledge Organization System mirrors the power afforded to the data themselves when linkages are made using variables; the difference with the latter being that careful consideration has to be given to the protection of individual data subjects' privacy when dealing with personal data. Linking metadata, on the other hand, has far fewer restrictions, and can support more powerful retrieval.

There are several initiatives these days which are pushing the open data agenda; the UK Data Service is engaging with many of these. It has released open data where possible and *made more open* those data which cannot be fully open. It has also embraced whole-heartedly the principles of open metadata through being a signatory to the discovery.ac.uk Open Metadata Principles¹, as endorser of the Joint Declaration on Data Citation Principles² and via distribution of its metadata (OAI, XML).

In being able to work freely and openly with metadata, the UK Data Service is implementing its vision to create a discovery tool with, borrowing from the health literature, a *whole systems approach* to metadata.

In the spirit also of FRBR, it promotes connections between related data, case studies of data use, publications and other outputs, citations, as well as investigating the inclusion of standardised means of identifying organisations and individuals (ORCIDiDs, ISNIs etc.). Through these activities it is creating metadata microcosms, representative of the wider data world that those metadata reference.

This whole system methodology is implemented within Discover, the UK Data Service's influential resource discovery application.

¹ <http://discovery.ac.uk/businesscase/principles/>

² <https://www.force11.org/datacitation>

Findings

In order to maximise information retrieval and to create such connections between the metadata, the UK Data Service employs a number of methods. This paper will describe these in the context of the Service's aim to create open metadata microcosms.

The paper highlights: the metadata schema used; the data citation methodologies that are a) employed and b) have been designed in-house; and the innovative, international SKOS-based thesaurus management development the service is currently undertaking via a 5-year ESRC award.

Discussion

The **metadata schema** the UK Data Service uses include:

- the Data Documentation Initiative Codebook (DDI-C) version 2.5;
- QuDex;
- Text Encoding Initiative (TEI);
- SDMX;
- Dublin Core (DC).

The primary reference schema is the DDI, which has become the *de facto* standard for social science metadata. The DDI-C schema contains 351 detailed elements to facilitate the accurate and cross-national description of the data.

Because of the need for cross-border information retrieval, the DDI Alliance recommends the extensive use of **Controlled Vocabularies** (CVs). The UK Data Service contributes to this aim both by sitting on the DDI Controlled Vocabularies Group and by implementing the resulting CVs.

The UK Data Service also creates and distributes its own CVs, through two linked thesauri: the Humanities and Social Science Electronic Thesaurus (HASSET) and its sister product, the European Language Social Science Thesaurus (ELSST). These tools are currently the focus of a large-scale development project, whose aims are a) to review and improve their structures and b) to re-develop their management applications so the two may be managed in tandem. The last of these aims is being taken forward in an innovative way, based on, but extending, the new ISO 25964, by mapping the concepts held in the two thesauri using extended equivalence relationships. This has resulted in metadata linkage between these two powerful products, allowing data users across the globe to search for and retrieve the data they require using twelve languages.

The last element of metadata linkage to which this paper will refer is **persistent identification**. The UK Data Service has been working with the British Library's ODIN project as a stakeholder to investigate how DOIs may be linked with ORCIDs to create more linkages for the benefit of data users.

The UK Data Service adds DataCite DOIs to all its data collections, with well-documented versioning practices. It is also developing methodologies for more granular data citation, at variable level. The first results of this work are visible via our qualitative material within the Quali

Bank, which allows a user to create a paragraph-level citation of interview material, thus providing an even more detailed, atomic view of the metadata.

The application of DOIs allows the Service's metadata to become part of the wider DataCite search as well. The Service encourages the academic community to support data citations wherever possible, for example via the inclusion of its metadata within the Thomson Reuters Data Citation Index. Through this and other means, it works to make data citation standard practice within academia, pushing for these citations to figure within the mesh of references published within the Citation Indexes.

Conclusions

Social science data provide a rich resource for policy-makers, academics and business in identifying trends and emerging issues in the economic and societal arenas.

Social science data are different to those from the hard sciences, because they are personal. Access to these types of data must be carefully managed. This must be balanced, however, with the requirement to make their metadata as freely and openly available as possible and, crucially, discoverable.

Linking social science data can create statistically disclosive variables, which all social science archives must prevent. Linking their metadata, on the other hand, can augment discoverability to the vast benefit of data users globally.

The implementation of Basel Committee BCBS 239: An industry-wide challenge for international data management

Malcolm Chisholm

Visiting Fellow in Financial Technologies at Loughborough University,
and President, AskGet.com Inc

Background

In January 2013 The Basel Committee on Banking Supervision (the “Basel Committee” or “BCBS”) released a final version of *Principles for effective risk data aggregation and risk reporting*. This document outlines a set of supervisory expectations - *de facto* regulations - that are to be adopted by Global Systemically Important Banks (“G-SIBs”) by January 2016, and which may at the discretion of national supervisory bodies be applied to Domestic Systemically Important Banks (“D-SIBs”). The term “BCBS 239” appears to come from the name of the file issued in January 2013.

It should be noted that BCBS 239 is only one of a set of regulations and initiatives that confront major financial institutions, e.g. LEI, Basel III, FATCA, COREP, Solvency II, FINREP, IFRS, and others. However, perhaps more than any other single regulation, BCBS 239 specifically addresses data governance and data management needs. These needs are mostly implicit in the other regulations and initiatives, and this is what makes BCBS 239 unique - it represents a maturation of global regulatory thinking about data governance and data management.

Although BCBS 239 only addresses risk management data aggregation and reporting, it would seem that if its principles are adhered to in general, then banks will be satisfying the more generally expressed requirements for data governance and management expressed in the other regulations and initiatives that are in effect or under development. This gives a powerful reason for adhering to BCBS 239. It is therefore reasonable for banks to ask what capabilities they need to be able to implement BCBS 239.

Approach

BCBS 239 provides 11 principles that must be applied to a bank's risk management data and all key internal risk management models, plus 3 other principles that apply to supervisors. In addition to the principles, 6 objectives are enumerated.

In December 2013, BCBS published the results of a survey it conducted amongst G-SIBs. The survey questionnaire crystallized out 65 specific requirements from the 11 principles that apply to the banks. This has provided additional specificity, which is enhanced by explanatory notes that accompany the principles. This provides a much more detailed foundation about what data governance and management practices the banks should focus on. It makes an analysis of capabilities required by the banks more feasible than dealing with the rather high-level principles alone.

In attempting to identify the capabilities banks must have to respond adequately to BCBS 239, four main groups of capabilities were considered:

1. **Leadership:** the capabilities needed to set vision, specify mission, earmark resources, and bear the final accountability for adherence to the full scope of BCBS 239.

2. **Organization:** the capabilities for implementing vision and mission, planning, assigning roles and responsibilities, mobilizing resources, recruitment, setting work plans, and ensuring work gets done for the scope of BCBS 239.
3. **Methodology:** the capabilities for applying knowledge in a structured and repeatable way to design data stores, design processes, assure quality, and carry out operations, for the scope of BCBS 239.
4. **Technology:** the set of automated tools that support the scope of BCBS 239.

Each of the 65 Questions/Requirements in the survey questionnaire was analyzed to determine what capabilities in the 4 categories listed above are required to implement it. This was done using knowledge of existing data governance and data management practices.

For each Question / Requirement, each of the 4 categories was considered as to the strength of the capability involved. A three-level rating was used.

0 = No to weak involvement

3 = Supporting role

5 = Required

Findings

The capability assessment does show what directly relevant capabilities are needed, and their relative scopes. As might be expected, the relatively high level Leadership category has the lowest number of capabilities. After all, Leadership involvement may be highly important, but generally is represented by relatively few considerations. More surprising is that the Methodology category is the most relatively complex, having the greatest number of considerations.

Methodology is rarely considered in great detail. It is often obscured by greater attention to Technology, and a presumption that the Systems Development Life Cycle (or some equivalent project management methodology) can supply all the capabilities required. However, this is not the case, and there is good reason for this. When it comes to data governance and data management, Technology can be tremendously supportive, but is often a container whose value is ultimately dictated by the quality of the metadata that it contains. This metadata has to be gathered automatically (e.g. database discovery) or input by human agents (e.g. definitions of data elements, construction of taxonomies). The Technology itself only creates part of the content, and even the quality of this is subject to Methodology.

The remaining Organization category contains many familiar capabilities, many of which are rather generic.

Discussion

The scoring used in the assessment of capabilities is a reflection of the number of considerations, in terms of capabilities, that pertain to each Question/Requirement. The scores do not reflect cost, effort, or importance, but are a rough measure of relative complexity. Even here, however, caution is needed. For instance, a technological capability may be indicated, but many non-functional capabilities are needed to implement such an item of technology (e.g. hardware acquisition, software acquisition, network adaptation) that are not reflected here.

With this caveat, however, it can be seen that a large number of data governance and data management capabilities are needed by banks in order to successfully implement BCBS 239. It is not easy to compare this set of capabilities with those currently possessed by banks in the areas of data governance and data management. However, if the agendas of well-attended conferences are examined, rather few of these capabilities are found. This indicates that banks have a major capabilities gap that they must overcome in order to implement BCBS 239.

Conclusion

BCBS 239 is the regulation with the greatest specific focus on data governance and data management. It is possible to identify the capabilities banks require to comply with it successfully. However, it appears that banks have a gap between their present capabilities and those needed to implement BCBS 239.

References

Basel Committee on Banking Supervision *Principles for effective risk data aggregation and risk reporting*, January 2013.

Adopting a situated learning framework for (big) data projects

Martin Douglas

Cranfield University, School of Management

Joe Peppard

European School of Management and Technology, Berlin

Background

While the human aspects of (big) data initiatives are recognized as important, they are relatively under-researched, and then usually at the organization level (e.g. culture) rather than at the project or initiative level, which is the common context for such initiatives.

Against this background, this paper seeks to determine the usefulness of adopting a Situated Learning framework when approaching such projects to supplement more traditional project frameworks, which tend to take requirements definition as their rather straightforward starting point. The paper emerges from a wider research project exploring human aspects of Data Insight projects, focusing on how participants in such projects think about or 'frame' such initiatives and how this impacts the progress and outcomes of initiatives.

The starting point for the study was recognizing that generating insight from data is essentially about discovering new knowledge, learning and indeed can be considered as research. It is therefore 'grounded' in or 'primed' by a comprehensive review of the social constructionist areas of Information Systems, Knowledge management, Sensemaking and Situated/Organisational Learning literatures.

Methodology

An abductive research strategy was adopted, using an ethnographic approach for the inquiry. An immersive study was undertaken in two organisations over a six month period, of three different data initiatives. It was exploratory and explanatory in outlook and mainly etic in focus. An engaged research stance was adopted.

Considerable data was collected in the form of a field diary of personal observation, in-depth interviews (14), attending project meetings (34), project artefacts (19) and additional incidental meetings and interactions. The vast majority of the interviews and formal project meetings were recorded and subsequently transcribed or summarized; detailed notes were taken for the remainder.

The collected data were subjected to several different forms of analysis, including: mapping communities of practice, comparative thematic analysis of interviews, exploratory narrative analysis, and detailed coding using Nvivo, in order to aid reflection, triangulation and reflexivity.

Findings

The paper describes an emerging preliminary explanatory model of the situated learning that occurs in such data initiatives. It illustrates its explanatory value for two related projects from one organization. It shows how practitioners from different communities interact, are bounded by prior knowledge, and highlights the challenges they face in generating insights and hence value from such initiatives.

The paper finds that adopting a Communities of Practice, situated learning framework is useful for researching such initiatives, and for the practitioner planning them. It identifies several

participant 'frames' for the projects studied, grounded in their prior knowledge and experience, as well as institutional or reified thinking, and suggests how these might hinder effective learning and generating insight. It also highlights boundary spanning activities and related sensemaking artefacts and how these may be important.

Discussion

The use of a situated learning framework brings the practitioners engaged in learning or generating insight firmly into research focus. Their pursuit of particular purposes in relation to a phenomenon of interest, using and mediated through data and related tools, presents a rich social context for investigation, and demonstrates the opportunity to use theory and approaches from knowledge creation and learning related disciplines.

The relationship between phenomenon and data surfaces as particularly important. While there has been considerable emphasis on the importance of quantitative skills for (big) data initiatives, this paper argues that it is research skills that are as important (qualitative and quantitative), e.g. to address questions of validity appropriately, especially when engaged in exploring social rather than physical phenomena.

The use and role of tools and applications also offered an opportunity to 'disentangle' related data and conceptualise it more clearly, especially where data is subsumed within third party software applications and becomes taken for granted in use. In this sense, embedded data can represent a 'snapshot' reification of knowledge about a phenomenon, in terms of the dimensions and distinctions considered important. This can frame use and result in bounding and anchoring impacts. This underlines the socially constructed nature of data and presents an opportunity to extend Orlikowski's (1991) structuration view of systems to more explicitly address related data.

Conclusions

This paper offers an explanatory framework for data initiatives, which researchers and practitioner may find useful. It demonstrates the usefulness of leveraging thinking from Knowledge Management and Organisational Learning disciplines when approaching research about generating insights from data. Moreover, it also argues for more clearly recognising the socially constructed nature of data when researching data initiatives.

Orlikowski, WJ (1991) The Duality of Technology: Rethinking the Concept of Technology in Organizations, *Organization Science*, 3(3), pp. 398-427

Learning about information management across projects

Sunila Lobo

University of Reading

Purpose or background

Digital technologies are transforming engineering design practices on major building and infrastructure projects. However, too much of the learning on technologies, and the processes and practices around these technologies are lost e.g. (Lyytinen and Robey, 1999) or there is 'significant degrees of reinvention and relearning' (Newell, Bresnen et al. 2006:175). There are also limits to the extent lessons are learnt and taken on, across projects (Kotnour, 1999).

In this paper, I draw on the literature on project-based learning to examine the categories of learning about information management that are taken and embedded across major projects, in one focal firm, Global Design, across borders and over time. Hence, this paper hopes to contribute to the project learning literature by attempting to answer the following research question:

Which categories of learning of one firm across projects get transferred to its other projects and becomes embedded into future project practice?

Method

Through contact with the firm 'Global Design', three infrastructure projects of theirs including the one, Motorway, which is the focus of this paper were selected as being especially innovative in their use of digital technologies for design, collaboration and project management. Two of the infrastructure projects, were completed UK-based projects, which were studied through interviews and secondary materials. The third, 'Motorway', was an on-going project in Australia (from approval in 2006 to completion, in 2010). The project involves a young team (average age, 24) at the other side of the world from the 'Global Design' central offices.

22 semi-structured interviews on Motorway (9 were telephone interviews conducted in October 2008 and 13 were face-to-face interviews conducted in March 2009 in Brisbane) were conducted. The interviews averaged more than an hour each. There was also a one to two hour set-up meeting with project managers via conference call; which was followed by feedback from the firm's senior engineers and managers after the first round of data collection. The feedback was used to clarify findings and collect further data in March 2009. I had the opportunity to visit the project site offices, getting to know the engineers and designers, taking some site photos and observing on-going work. Data collected included pre-interview questionnaires. Hence this research can be seen as interpretive, case study research (Eisenhardt 1989, Walsham 1995, Walsham 2006).

Setting

The Motorway upgrade project is divided into eight areas with 26 new bridges being built across its length. The dedicated design and project management organization comprises a 7-member Alliance Leadership team made up of the client, and the 4 member companies (2 design consultancies and 2 contractors), and Alliance Managers. The design consultancies engaged on this project formed a Design Joint Venture: The firm 'Global Design' is in charge of building 15 bridges and another firm 'Design International' was commissioned to design 11 bridges in Motorway. The overall design office had about 90 designers at its peak.

Findings

The categories of learning were derived from the analysis of the interview transcripts, documents, and through feedback received on the case report written up for Global Design in 2009. The areas within which learning transfer between projects and on to the Motorway project occurred can be categorised as learning about information management and data in aspects which enhance *coordination* and *control* across heterogeneous and dispersed teams, as well as, *practices*, (people and task). Table 1 briefly describes these categories:

Table 1: Categories of project learning which were the focus

Categories of Learning	Examples
Coordination	<ol style="list-style-type: none"> 1. Repositories - BIM Solution for (international) real-time collaboration was recommended and used on Motorway 2. IT strategy and set-up for efficient delivery across teams and disciplines seen to be crucial (some issues initially) 3. Co-location of project teams including design, construction, client 4. How to work with remote/distributed teams and with co-located teams- e.g. need to be more clear in mark-ups and build relationships
Control	<ol style="list-style-type: none"> 5. Rules – design management* 6. Processes – design review process 7. Procedures – milestone definition sheets 8. Standards - drawing and numbering standards, directory structures, indices 9. Objects – spread-sheets (standardized forms), PDF of final drawing for review
Practices	<ol style="list-style-type: none"> 1. Safety in Design – took learning from the UK, which is more advanced 2. Different ways of working – the 2 design teams had to adapt to each others' practices

*NOTE: There were no guidelines set-up for design work by Global Design on the Motorway project – since then, Global Design has set up a standard set of rules to be used on future projects.

A point to note is that in Bailey and Barley (2011)'s paper on learning by engineers, the 5 broad content areas where learning was sought or offered are aligned with some of the learning categories identified in Table 1.

Discussion & Conclusion

The paper shows that cross-project learning transfer (DeFillippi and Arthur 1998, Newell, Bresnen et al. 2006) in particular categories of learning, has occurred across borders and over time, for a particular focal firm, Global Design. Unlike previous literature on cross-project learning which distinguishes between, for instance, product and process knowledge (Newell, Bresnen et al. 2006) and mechanisms of learning transfer (e.g. Hansen et al, 1999), this paper's focus is on the categories of cross-project learning about information management which gets embedded across major projects. It also addresses the question raised by Kotnour (1999) as to what lessons are actually used or taken advantage of, across projects.

This paper contributes to theory by showing how this learning becomes an organisation-wide capability, at the cross-project level.

Bibliography

- Ayas, K. and N. Zeniuk (2001). "Project-based learning: Building communities of reflective practitioners" *Management Learning* 32(1): 61-76.
- Bailey, D. E. and S. R. Barley (2011). "Teaching-Learning Ecologies: Mapping the Environment to Structure Through Action." *Organization Science* 22(1): 262–285.
- Bresnen, M., A. Goussevskaia, et al. (2004). "Embedding New Management Knowledge in Project-Based Organizations." *Organization Studies* 25(9): 1535-1555.
- Cacciatori, E. (2008). "Memory objects in project environments: Storing, retrieving and adapting learning in project-based firms." *Research Policy* 37(9): 1591-1601.
- Davies, A. and T. Brady (2000). "Organisational capabilities and learning in complex product systems: towards repeatable solutions." *Research Policy* 29: 931-953.
- Davies, A. and M. Hobday (2005). *The Business of Projects: Managing Innovation in Complex Products and Systems*. Cambridge, Cambridge University Press.
- DeFillippi, R. and M. Arthur (1998). "Paradox in project-based enterprise: the case of film making." *California Management Review* 40(2): 125-139.
- DeFillippi, R. J. (2001). "Project-based learning, reflective practices and learning outcomes." *Management Learning* 32(1): 5-10.
- Easterby-Smith, M., K. Golden-Biddle, et al. (2008). "Working with pluralism: determining quality in qualitative research." *Organizational Research Methods* 11(3): 419-429.
- Eisenhardt, K. M. (1989). "Building theories from case study research." *Academy of Management Review* 14(4): 532-550.
- Engwall, M. (2003). "No project is an island: linking projects to history and context." *Research Policy* 32(5): 789–808.
- Grabher, G. (2002). "Cool projects, boring institutions: Temporary collaboration in social context." *Regional Studies* 36(3): 205-214.
- Grant, R. M. (1996). "Toward a knowledge-based theory of the firm." *Strategic Management Journal* 17(Winter Special Issue): 109-122.
- Hansen, M., N. Nohira, et al. (1999). "What's Your Strategy for Managing Knowledge?" *Harvard Business Review* March-April: 106-116.
- Hobday, M. (2000). "The project-based organisation: an ideal form for managing complex products and systems?" *Research Policy* 29: 871-893.
- Huber, G. P. (2001). "Transfer of knowledge in knowledge management systems: unexplored issues and suggested studies." *European Journal of Information Systems* 10: 72-79.
- Kotnour, T. (1999). "A Learning Framework for Project Management." *Project Management Journal* 30(2): 32-38.
- Leonardi, P. M. and D. E. Bailey (2008). "Transformational Technologies and the Creation of New Work Practices: making Implicit Knowledge Explicit in Task-based Offshoring." *MIS Quarterly* 32(2): 411-436.
- Lindkvist, L. (2004). "Governing project-based firms: Promoting market-like processes within hierarchies." *Journal of Management and Governance* 8: 3-25.
- Lyytinen, K. and D. Robey (1999). "Learning Failure in Information Systems Development." *Information Systems Journal* 9(2): 85-101.

Newell, S., M. Bresnen, et al. (2006). "Sharing Knowledge Across Projects: Limits to ICT-led Project Review Practices " *Management Science* 37(2): 167-185.

Newell, S. M. and L. F. Edelman (2008). "Developing a Dynamic Project Learning and Cross-Project Learning Capability: Synthesizing Two Perspectives." *Information Systems Journal* 18(6): 567-591.

Prencipe, A. and F. Tell (2001). "Inter-project learning: processes and outcomes of knowledge codification in project-based firms." *Research Policy* 30: 1373-1394.

Scarbrough, H. (2008). *The evolution of business knowledge*. Oxford, Oxford University Press.

Scarbrough, H., J. Swan, et al. (2004). "Project-Based Learning and the Role of Learning Boundaries." *Organization Studies* 25(9): 1579-1600.

Szulanski, G. (1996). "Exploring internal stickiness: Impediments to the transfer of best practice within the firm." *Strategic Management Journal* 17: 27-43.

von Hippel, E. (1994). "'Sticky Information" and the Locus of Problem-Solving: Implications for Innovation." *Management Science* 40(4): 429-439.

Walsham, G. (1995). "Interpretive case studies in IS research: nature and method." *European Journal of Information Systems* 4(2): 74-81.

Walsham, G. (2006). "Doing Interpretive Research." *European Journal of Information Systems* 15(3): 320-330.

Whyte, J. and S. Lobo (2010). "Coordination and control in project-based work: digital objects and infrastructures for delivery." *Construction Management & Economics* 28(6): 557-567.

Exploring different information needs in Building Information modelling (BIM) using Soft systems methodology

Mohammad Mayouf and David Boyd

Birmingham School of Built Environment, CEBE Faculty, Birmingham City University

Sharon Cox

School of Computing, Telecommunications and Network, CEBE Faculty, Birmingham City University

Managing information for construction projects has been a crucial task where information technology (IT) was employed to tackle this issue. Building information modelling (BIM) is considered to be the first truly global digital construction technology; it is a process that aims to inform and communicate project decisions through creating and using an intelligent 3D model. BIM is claimed to be an effective tool for information exchange, which involves digital representation of physical and functional characteristics of the building. However, the involvement of interdisciplinary stakeholders within a construction project implies different data requirements, thus various information needs. The complexity of information and data required to deliver this information in construction projects is an on-going issue, thus understanding the nature of this complexity is needed. This paper aims to investigate different information needs from multiple perspectives, and raise awareness towards data requirements that BIM can incorporate. CATWOE as one of the analytical tools of soft systems methodology (SSM) will be used to demonstrate different data requirements. The data is obtained using interviews conducted on the facility management team and end-users from newly operated buildings. The paper concludes with a proposed road map suggesting different required data from design to operation. Further work is needed to check BIM capabilities in terms of integrating this data, and whether interoperability issues would require additional tools to support the BIM process. This paper provides essential awareness towards information needed from BIM, thus resulting in a more productive building.

Using Big Data in Banking to Create Financial Stress Indicators and Diagnostics: Lessons from the SNOMED CT Ontology

Alistair Milne and Paul Parboteeah

School of Business and Economics, Loughborough University

Purpose or background

Big data and the opportunity it presents for near real time access to large scale information sets from disparate sources promises to become a key tool for monitoring and anticipating systemic financial risks. But experience in a variety of fields (eg spatial data analysis and bio-informatics) on data modelling and visualisation indicates that a prior requirement for obtaining precise and interpretable results from big data is a clear understanding of underlying data structures and relationships.

In the context of systemic risk monitoring this can be understood by drawing a distinction between financial stress indicators and financial stress diagnostics. By financial stress indicators we mean timely but broad indicators of current or prospective financial stress. By financial stress diagnostics, we mean measures that reveal a specific mechanism that can potentially trigger a systemic problem and thereby suggest an appropriate policy response.

While big data may be useful 'as is' for the creation of financial stress indicators, using big data to create effective financial stress diagnostics that are revealing about the underlying causes of stress requires something more: a conceptual framework for understanding of the granular data on financial instruments and how these can be parsed and aggregated i.e. in modern computer science terminology an ontological understanding. It is therefore appropriate to try and draw lessons, from the development of ontologies in other domains. The aim of this paper is to examine the SNOMED CT medical ontology, drawing out the lessons for the use of big data for monitoring systemic financial risk.

Design, methodology or approach

This paper brings together previous research on the development and use of SNOMED CT, provides a critique on its successfulness to improve patient care/drug development and extracts the lessons that can be learnt. Stemming from this critique, implications are drawn for the creation of an ontology to support financial stress indicators and diagnostics. In more detail: the paper considers the current state of data analysis and modelling in financial services, exploring where and why an ontologies may be useful. It also explores where standardization has been developing and is starting to set the foundation for an ontology of financial activities.

Findings

After briefly reviewing SNOMED CT and its development, use, problems and role in supporting clinical decision making, lessons can be extracted from the SNOMED CT implementation and applied to the creation of financial stress indicators and diagnostics.

The first lesson from SNOMED CT is that developing an almost industry wide ontology is a large undertaking. Whilst it did take approximately 40 years to arrive at SNOMED CT, efforts now to produce a useful FIBO or LEI implementation should be substantially shorter: the technological environment is more advanced, the regulatory pressures are greater and ontologies are already a proven technology, with uses in, for example NASA. The SNOMED CT model of development does however act as a good practice for FIBO developers (FIBO being necessary for the semantic analysis need in financial stress diagnostics).

As noted by Heja et al. (2008) among others, SNOMED CT does contain numerous ontological errors that prohibit automated processing and reasoning. A further lesson to be learnt here is that sufficient importance needs to be attached to the rules governing the creation of ontologies in this space, including FIBO. Creating an organizationally correct ontology will then permit the semantic analysis to enable the financial stress diagnostics to take place because inferences will be able to be drawn and decisions taken.

In contrast to SNOMED CT however, creating a stress diagnostic system is more about mapping the financial system as a whole to identify relationships and dependencies. SNOMED CT on the other hand, has as its focus the drawing of better conclusions from large cross-sectional data and better understanding of which treatments or risk factors are more closely linked to disease and its progression.

Discussion and conclusion

In this paper we have sought to draw a distinction between financial stress indicators providing immediate warnings of imminent systemic problems and financial stress diagnostics which provide insight into the underlying causes of stress and therefore into the appropriate policy response. We have argued that while big data techniques can be used directly to construct financial stress indicators their use to create financial stress diagnostics requires an ontological understanding of data at the granular level. Only then will it be possible to effectively use big data to provide effective quantitative measurement of key systemic vulnerabilities of the financial system, such as network instabilities or common exposures to property and other asset markets.

There has been progress in the creation of 'common financial language', definitions, standards and at least in prototype form ontologies for finance. But, while travel has been in the right direction, the experience from medicine of the SNOMED CT ontology suggests this is still only the beginning of a long road. It will take a sustained effort, involving many stakeholders, to develop a complete financial ontology that can be used to support a complete mapping of financial sector vulnerabilities from big data. Achieving this end will depend on effective co-operation and governance, which is not always easily achieved.

The use of ontologies to gauge risks associated with the use and reuse of E-Government services

Onyekachi Onwudike, Russell Lock and Iain Phillips

Dept of Computer Science, Loughborough University

E-Government ontologies have successfully been developed for different levels of government over a number of years. However, the majority of these ontologies have had little or no impact on E-Government as a whole. The varying number of E-Government ontologies in existence that have been developed in isolation, without wider integration in perspective and the lack of reuse of components present serious challenges for the E-Government domain. Although the idea of reuse across ontologies seems to be a welcome idea with respect to the problem of interoperability, the risks and disadvantages associated with reusing existing solutions, as well as making certain functionalities sharable between E-Government services, is a relatively new area of research.

This reuse of existing solutions may potentially help to foster co-operation amongst E-Government departments, reduce costs and reduce development time as well as increase reliability and maintainability of such systems. This paper explores existing E-Government ontologies and assesses the assistance a suitably designed ontology could have in reducing system development and evolution risks. It presents the role of ontologies in overcoming risks that may be associated with service combinations such as overlapping of services, the uncontrolled reuse of components, monopoly of information across departments and areas of dependency resulting in conflict amongst others. The listed scenarios avail us the opportunity to make decisions from the ontology on whether some combination of services are beneficial especially in cases where there is service dependence amongst services. We conclude that the use of ontologies may play a significant role in gauging the risks associated with this.

Introduction

It is currently difficult to answer questions such as “what difficulties or threats can arise when information is exchanged across departmental boundaries?” or “can dependencies among services result in conflict?” In areas such as business and engineering, metrics are used to determine the health of a project and whether the dividends justify the costs. The threats that face any enterprise are critical to the advancement and growth of that enterprise. Looking at the E-Government domain as an enterprise especially with respect to the delivery of services, there is a possibility of making incorrect or unwise decisions when it comes to the reuse of service components, the threats such reuse and combination of components bring about as well as the possible countermeasures. It is one thing to chant the chorus of reuse; it is quite another to discover the effects these have on the E-Government domain. Since there is no precise definition of the adverse effects lack of reuse can bring about in the E-Government domain, there is confusion among Service Providers on which services can be combined and which components of these services can be reused. Therefore an ontology to gauge the risks and threats associated with this is a viable solution to this problem. The reason for this is that they would allow definitions of the entities in an ontology in a sound manner as well as represent relationships such as dependencies that exist amongst entities. Furthermore, the use of an ontology would give decision makers greater depth as to why certain decisions should be made when it comes to Electronic services in the E-Government domain. Having seen that decisions are mainly made by managers who have little or no knowledge about the underlying infrastructure on which the E-Government domain is built upon and who base decisions on intuitions rather than on defined metrics, the use of an ontology can be used to conquer this.

Therefore, the main goal of this research is to develop an ontology that demonstrates the threats that endanger the E-Government domain in areas such as component reuse, component combination as well as asset procurement and which countermeasures can be taken to counteract the threats that would be identified or better still reduce the probability of fatality.

E-government ontologies

An ontology is defined as “an explicit specification of a conceptualization” (Gruber, 1993). Therefore E-Government ontology can be described as an explicit description of the E-Government domain containing a common vocabulary with shared understanding. In the context of this paper, the domain of an E-Government ontology comprises of issues that are government related. Therefore the purpose of the E-Government ontology is to facilitate adequate understanding of the E-Government domain by service providers so that issues relating to the integration of services as well as the risks associated with integration in the E-Government domain can be addressed as well as used as prediction tools for future governments.

Designing an ontology to gauge risks and threats associated with E-Government

The choice of OWL to model our ontology gives us the ability to easily build systems that are interoperable which would enable the production, reasoning and visualization of data in the E-Government domain (OWL, 2004). Considering that the E-Government domain is a large one and the amount of data associated with it is quite large; there is need to make use of reasoning components that are highly optimized which is made available through the use of off-the-shelf editors i.e. Protégé (Protege, 2005). The main goal of this research is to provide an ontology to support service providers and decision makers in the E-Government domain. The ontology would be able to highlight potential threats and risks which endanger service combinations, component reuse as well as asset procurement and what countermeasures can be taken to lower the chances of attacks and severity.

Application scenarios

Various scenarios exist for the application of E-Government services which make use of components that are reusable or integrateable. For example, the Driving License department issues Drivers licenses while the Passport Department issues passports. These are E-Government services that are common to any government. There are certain components that can be reused by these departments such as Authorisation/Authentication components which make use of information like Name, Date of birth, Nationality.

These services can actually reuse these components or be combined in order to increase the efficiency of the systems (Musen, 1992). Although it is good practice to reuse these components, can the risks associated with reusing these components be investigated with the use of an ontology?

Conclusion

With the use of an ontology, we would be able to investigate the efficiency and resilience of this tool against an existing Government by pointing out whether certain service combinations or integrations are useful. We would also be able to investigate with this tool the overlaps, failures and collisions that may occur as a result of this.

References

Gruber, T. R. (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing, 907–928.

Musen, M. (1992). Dimensions of knowledge sharing and reuse, 25(5), 435–467.

Singhal, A. (2010). Ontologies for Modeling Enterprise Level Security Metrics Categories and Subject Descriptors, 5–7.

The protégé ontology editor and knowledge acquisition system, <http://protege.stanford.edu>

W3C (2004) OWL Web Ontology Language Overview. <http://www.w3.org/TR/owl> features

Assessing trustworthiness in Digital Information

Frances Johnson, Laura Sbaffi and Jenny Rowley

Department of Languages, Information & Communications, Manchester Metropolitan University

Purpose

Within the field of Information Management there are diverse research communities. Connecting those that study information behaviour and those focusing on systems design has the aim to build the understanding of how knowledge of information behaviours can improve the design of interactive information systems. In the context of providing information in digital environments, the information behaviour of trust formation may be core as a predictor of a user's 'intention to use' a given piece of information, typically to resolve some underlying need or problem. As the design of information systems and/or information based applications become more interactive, enhanced with functionality enabled in the web environment (such as links to related information, recommendations, features for annotation and personalisation) it seems vital that such developments serve a functionality and, at same time, impact on the users' judgement of trust. With regards to the conference theme of 'making connections' the poster presentation will review our recent research on modelling trust with a discussion on some of the key findings which suggest the use of a trust scale in understanding the potential impact of the system design on the confident use of digital information.

Background

Trust in electronic environments has been variously studied (Chopra and Wallace 2003; Ivanov et al 2012; Kelton et al 2008; Shekapour and Ketebi 2010; Belanger & Carter 2008; Rowley and Johnson 2013) and specifically in the context of online information, the user has a particular need to fulfil, creating a state of dependence and precondition for trust to be formed. This suggests trust is a dynamic concept formed in the context of the information and the information need and furthermore is likely to be subject to various influences, thus multidimensional and difficult to measure directly.

Approach

Towards the development of an instrument for the study of trust in digital information, the aim in this research is to identify the factors that affect the formation of trust. The empirical study adopted a quantitative, survey-based research design, in order to develop measurement items and explore the relationships between variables in the process of trust formation.

Questionnaires were chosen to collect data as they were deemed suitable for gathering large amounts of data and collecting accurate information. The core of the questionnaire was a set of Likert-style statements, designed to investigate respondents' (in this case students') perceptions of the relative importance of various aspects of web health information in their evaluation of its trustworthiness. Specifically the aspects reflected the constructs of credibility, usefulness, content, authority, style, verification, brand, ease of use, recommendation and were designed with a 5 point scale. On the data collected, factor analysis was carried out to test a theoretical model of the core criteria on which trust is formed and to model the observed correlations among the influences, in this case the critical evaluation of the information and its context.

Findings

The findings of the study identify the key factors that influence trust judgements as information style, content, brand and ease of use, and modelling the data positions credibility and usefulness as the most important antecedents to trust. In addition, the findings from the study demonstrate that trust formation involves a range of factors, and is a complex process. Exploration of the descriptive statistics at the item level offers further insight into the critical

evaluation of the information particularly with a greater range of cues and indicators being brought to bear in judgements formed by students as they progress to the final stages of their studies. Their level of involvement in the information judged was also recorded according to their declaration of the information need as either triggered by a general interest in the health topic, or because they or a family member had a specific complaint. This variable was also taken to explore the impact of the user's dependency state on the importance of the factors adopted in forming a judgement of trust. In the data set analysed it appears that with a higher level of involvement with the information (that is, triggered by a specific complaint) a greater consideration was given to the information content and less given to the contextual indicators such as ease of use in forming the trust judgement.

Discussion

On the basis of these findings, recommendations for further research can be made to develop the use of the trust scale as an instrument in furthering understanding in the information behaviour of trust formation in digital information contexts. Not least further investigation is needed in a variety of contexts, both user and task contexts, to explore the nature of the differences in the factors influencing the formation of trust with the view towards obtaining a theoretical model. On a practical application the scale, developed for use in evaluating the impact of design on the user's trust formed, offers a potential to inform system designers and developers concerned with the impact on use and usability but also, and of equal importance, on the users' ability to form critical judgements on information presented and specifically on the particular dimensions or constructs of trust. The opportunity to present this large scale study of trust in digital information in the proposed poster session allows for reflection on its potential to inform and build collaborative work across relevant research communities in information management. In particular the study of trust formation and, when based on the critical evaluation of the information as a key behaviour, is subject to variation in different contexts, suggests its value in use in the development and evaluation of system interface design in enabling this critical user behaviour.

References

- Belanger F & Carter L.(2008) Trust and risk in e-government adoption. *The Journal of Strategic Information Systems*. 17(2), 165-176.
- Chopra K,. Wallace W A. (2003). Trust in electronic environments. *System Sciences. Proceedings of the 36th Annual Hawaii International Conference on. IEEE*.
- Ivanov I, Vajda P, Lee J S & Ebrahimi T (2012) In tags we trust: Trust modelling in social tagging of multimedia content. *Signal Processing Magazine, IEEE* 29(2).
- Kelton K, Fleischman K. R, & Wallace W. A. (2008). Trust in digital information. *Journal of the American Society for Information Science and Technology*, 59 (3), 363-374.
- Rowley J & Johnson F. (2013) Understanding trust formation in digital information sources: The case of Wikipedia. *Journal of Information Science*, 39(4), 494-508.
- Shekorpour S,& Ketebi, S D. (2010) Modelling and evaluation of trust with an extension in semantic web. *Web Semantics: Science and Agents on the World Wide Web*, 8(1) 20 26-36.

Managing the risks of internal uses of Microblogging within small and medium enterprises

Soureh Latif Shabgahi and Andrew Cox

University of Sheffield

Purpose

The current enterprise microblogging (EMB) literature was reviewed focussing on the internal uses of microblogging, rather than marketing. It was found that the risks of EMB and ways of managing the tools remain under-explored and require further inquiry. The aim of the research described in this paper is to investigate the risks of microblogging as perceived by companies and based on this evidence propose company policies to manage risk related to microblogging. It is hoped that the findings can help organisations reduce the possible risks of microblogging.

Literature Review

The most familiar example of microblogging is Twitter. Launched in 2008 and since 2012 owned by Microsoft, Yammer is another example of a microblogging platform, but one specifically designed as a private and collaborative tool for enterprise social networking. Most research to date on EMB has been typically the result of trials of microblogging tools within large organisations. Before conducting this study, it was expected that small to medium enterprises may be also engaging with these platforms for business purposes, but there is little research exploring this.

Examples of uses of EMB identified in the literature include creating or sustaining a feeling of connectedness (Zhao and Rosson, 2009) and sharing knowledge or information (Schöndienst et al., 2011). Much of the EMB literature also explores the risks of microblogging. For example Zhao and Rosson (2009) identified a threat to the privacy of employees i.e. having concerns about the safety of using microblogging. Case and King (2010) also suggest there is a risk of confidential or private information leaking to outsiders through microblogging.

Several researchers have suggested the need to introduce policies to organisations, in order to mitigate such risks. For example, Case and King (2010) suggest that rules and guidelines may be required in order to balance the benefits of using EMB tools such as Twitter, with the risk of leaking private data. Similarly, Schöndienst et al., (2011) claimed that issues of privacy could be minimised by introducing rules on who can receive contents, and for how long they will be shared.

Although the current EMB literature has revealed the need for policies in organisations, aspects related to ways of managing the tools remain under-explored and are the focus of this study. Having consulted other relevant literature about organisational policies i.e. websites which educate people about new technology and managing the workforce, it was again recognised that employers need a comprehensive and well-defined policy to prevent abuse. For example, there is a need to clearly state the benefits of using the platforms and of having a policy, to provide a clear platform for educating employees, taking into consideration any legal consequences, as well as referring to proprietary and confidential information at risk.

Design/methodology

Empirical data was collected through conducting face-to-face semi-structured interviews with 23 employees of SMEs in the area of South Yorkshire, UK. The participants were selected by directly contacting the organisations by phone or email. The majority of participants were the director or manager of their organisation and they had influenced the decision to adopt

microblogging. The first set of questions asked background information about the company and process of microblogging adoption. The second and third section focused on the benefits and risks of microblogging and how microblogging could be improved, to help minimise issues associated with using the tools. Thematic analysis was selected as the approach to analysing the data. The '6-phase guide to performing thematic analysis' was followed (Braun and Clarke, 2006).

Findings

All of the organisations used Twitter and a few also used Yammer. Based on the qualitative data collected, a visual representation of the risks of microblogging, as well as of management strategies for EMB was developed. "Internal Risks" and "External Risks" were identified. The internal risks were: "Bullying Colleagues", "Loss of Personal Privacy" and "Losing Valuable Information". The main themes associated with the external risks were: "Breach of Confidentiality", "Damage to Reputation", "Misleading Information", "Negative Media Coverage" and "Computer Security Risk". A number of participants associated: "No limitations/ Risks" with EMB. Reasons include: "Risks exist around Usage and not Tools". Eleven main types of policies were revealed in the data. They include: "Policy on Talking about Colleagues", "Policy on Who Should Engage with Microblogging" and "Have work reviewed by Others". A small number of participants did not introduce rules or policies. Their reasons include: "Rules turn people off using Microblogging".

Discussion

There are both similarities and differences between the findings of this research and previous work. Previous EMB literature, for example, identified threats to the privacy of employees i.e. having concerns about the safety of using microblogging (Zhao and Rosson, 2009). Interviewees in this research also identified this as an internal risk: "Loss of Personal Privacy". The differences between this research and previous work are that although EMB literature has explored the risks of microblogging, it has not differentiated between internal and external risks. In this study, threats of microblogging have not only been explicitly identified as internal or external, the relationships between the risks have also been explored. For example there are threats which have been considered as main risks, and some are influenced by or lead to other risks.

Several researchers have also explored the introduction of policies, in order to mitigate risks. However, as more risks have been identified in this study so correspondingly a wider range of potential policies have been identified.

Conclusions

Before this study, most research to date has been typically the result of trials of microblogging tools within one large organisation. Previous studies were not necessarily a reliable guide to how other types of organisations such as SMEs might use microblogging, and the types of policies they could introduce to suit this type of workplace. The visual representations of the data can be a useful guide for other researchers and organisations to explore the company policies surrounding microblogging. They can help organisations avoid or minimise possible risks of microblogging.

Re-purposing manufacturing data: a survey

Philip Woodall and Anthony Wainman

University of Cambridge

Purpose or background

Manufacturing organisations currently use data for a variety of purposes and in many companies operational data is stored in an Enterprise Resource Planning (ERP) system. This may include data about suppliers, deliveries, customers, sales etc. Operational uses of this data span the selection of suppliers for particular materials by the procurement department to sending out marketing materials by the marketing department.

With the advent of Big Data, and the effort needed to manage large amounts of organisational data, companies are keen to ensure that maximal value is extracted from their data. The operational use of data takes priority and the data must be fit for this purpose. However, many organisations are starting to invest in data analytics programmes where the operational data is also used by data scientists to reveal business insights that can give a company a competitive advantage (Davenport and Harris, 2007). Unfortunately, a major problem is that many data scientists spend inordinate amounts of time getting data into a form where it can be used for their analysis (Kandel et al., 2012). The data must be transformed in some way before it can be input into data analytics tools and algorithms such as data mining applications. Typical transforms include:

- conversion of units such as times and dates (e.g. from two week intervals to 1 month intervals) or units of measure (e.g. centimetres to metres)
- coding of values and abbreviations such as United Kingdom to UK
- conversion of aggregated quantities (e.g. 1 box of items may be equivalent to 500 parts)
- entering values for missing values (e.g. entering a default value, or inferring the missing value based on other data)

In many cases, organisations create and use satellite systems, such as spreadsheets, that ERP data must be extracted into and then manipulated manually in order to carry out the various transforms needed to get data into the necessary form for analysis. A problem that this causes is that this data often then becomes the “real source of truth” and ends up being better quality than the data in the ERP system. The spreadsheet becomes the holder of the data that is managed and looked after.

People in other departments, who need the data, end up using the old and dirty data in the ERP system and have no visibility of the transformed data in the satellite systems. In this scenario, there is the potential for other departments to leverage these transforms for other purposes and therefore recoup some of the costs of transforming this data.

Design, methodology or approach

We therefore address the problem of how can manufacturers re-purpose ERP data, which has already been used for a business operation, for other tasks/decisions in different parts of the business?

Note that re-purposing in this case refers to the use of data for a completely different task/decision than the original purpose. This is different from reuse, which would be using the data again for the same or very similar task.

In order to tackle this research question, the following two sub-questions will be answered:

Sub question 1: What are the situations where data can be repurposed and what is the demand?

Sub question 2: How does the data need to be transformed to be fit for purpose for the new target audience?

The approach to answer these questions will involve interviews with manufacturing companies to discover how they currently use their ERP data and what transformations are needed when the data is re-purposed.

Using this information the overall aim is to develop a system that can automatically present ERP data to any department and enable the user to select and execute existing transforms that have already been implemented. This way, the users do not need to re-implement transforms themselves and they can simply pick and choose from the available transforms ones that will modify the data as they require.

This is an ongoing study where we expect to report the findings within the next two months in the full paper for the conference.

References

Davenport, T.H. and Harris, J.G. (2007). *Competing on Analytics: The New Science of Winning*, Harvard Business School Press.

Kandel, S. et al. (2012). Enterprise data analysis and visualization: An interview study. *Visualization and Computer Graphics, IEEE Transactions on*, 18 (12), p.pp.2917–2926. Available at: [Accessed March 20, 2014].

Exploring vendor's capabilities for cloud services delivery: A case study of a large Chinese service provider

Gongtao Zhang, Neil Doherty and Mayasandra-Nagaraja Ravishankar

School of Business and Economics, Loughborough University

Introduction

Cloud computing is a service-based model of computing resources provision (Marston *et al.*, 2011; Venters and Whitley, 2012; Willcocks *et al.*, 2014). Recently, the phenomenon has been the subject of much interest in practitioner and IS research. For example, in 2010, Amazon's annual revenue of cloud services was estimated between \$500m and \$700m (Economist, 2010). Forrester predicts that global market for cloud computing will grow to \$240bn by 2020 (Dignan, 2011). Cloud computing has been viewed as an effective solution which improves operational efficiency, simplifies management process and motivates innovation (Boss *et al.*, 2007; Hayes, 2008; Willcocks *et al.*, 2013). The cloud computing literature suggests that the technology can be exploited for better value creation in different industries (Marston *et al.*, 2011; Willcocks *et al.*, 2013). Yet, due to complex and rapidly changing environments, cloud computing for value creation remains a challenging problem.

There are two important gaps in extant research on cloud computing led value creation. First, the process of how cloud computing can be deployed to achieve benefits, has not been examined in great detail. In other words, most of the current prescriptions for acquiring benefits with the use of cloud services have not been empirically validated (Garrison *et al.*, 2012), and thus appear to be overly abstract. Secondly, the vendors' perspective of cloud computing has hardly been explored. Understanding the vendor perspective in IT service engagement is crucial, since vendors' capabilities determine the quality of service delivery (Levina and Ross, 2003). In this study, we address these gaps by understanding and examining how vendor capabilities deliver cloud services, through a thorough case investigation of a large service provider in China.

The resource-based view (Barney, 1991) of the firm and the notion of dynamic capabilities (Teece and Pisano, 1997) are applied as theoretical lens in this study. We used a combination of narrative analysis and visual mapping strategy to organize the empirical data. Based on the emerging data, we categorised the events, activities and decisions into three distinct phases. The insights from the data are presented by a process model of vendor's cloud service delivery.

Design/Methodology/Approach

The case study approach is adopted for two reasons. First, it is usually used to answer 'how' and 'why' questions (Yin, 2003). Second, the case study is a well-established method in IS research, particularly for “*sticky, practice-based problems*” (Benbasat *et al.*, 1987). The investigation of how vendor capabilities deliver cloud services satisfies all these criteria.

There were two stages of data collection, namely pilot interviews and case investigation. A total of 23 interviews were conducted in June 2013. The first stage aimed to formulate proper interview guidelines, test the feasibility of data collection, and gather general market information. Then, a further step was conducted to bridge, improve and shape our ideas, in a new context: China. The second stage was designed to conduct detailed case study. Secondary data was also collected to supplement interview data, such as internal publications, memos, newspaper articles, industry reports and information from the company website.

Findings

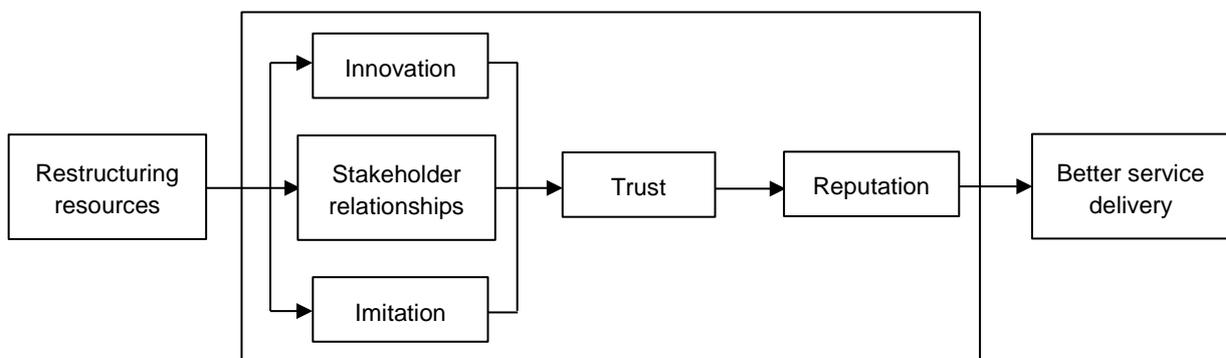
We developed three distinct time-based phases to demonstrate the evolution process of the capability development in the case. The three phases are: (1) acquiring primary firm-specific resources, (2) obtaining new capabilities, (3) reforming core capabilities. In each phase, we highlighted the motivation, process and consequences of strategic activities. Our analysis demonstrates the important roles of six core capabilities which are relevant to delivering cloud services. The six capabilities are: innovation, stakeholder relationships, reputation, imitation, restructuring and trust. After elaborating each capability, we further develop the implications of each capability, in terms of how they influence the delivery of cloud services. Finally, these findings are interpreted in light of the extant literature.

Discussion and Conclusion

We have developed a process model (see Figure 1) to highlight the relationships between each capability, and clearly demonstrate how each capability influences the service delivery. The model also contributes to existing literature by opening the black box of value creation, from the vendors' perspective.

This study makes several important theoretical contributions. First, by examining how vendor delivers cloud services, this study contributes to a resource-based perspective of cloud service delivery. Second, by further examining how vendor's capabilities are developed, the study sheds light on how a firm sustains its competitive positioning in cloud service industry. Third, the study has provided evidences which reveal how value is delivered, by in-depth case investigation. The case study clearly explains how a vendor firm shapes, develops and reforms its capabilities to deliver cloud service in a rapidly changing and intensively competitive context.

Figure 1: A process model of vendor capabilities



In terms of practical implication, this study sheds important light on cloud service delivery, particularly for vendors. This study also provides details on Chinese context, and empirically supports what capabilities are crucial to deliver public services. Vendors may use our model to assess whether they have developed essential capabilities. Through investigating the case, they can have a better understanding of how to develop these capabilities and identify possible risks.

Reference:

- Barney, J. B. (1991). Firm resources and sustained competitive advantage. *Journal of Management* **17** (1), 99-120.
- Benbasat, I., Goldstein, D. K. & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly* **11** (3), 368-386.
- Boss, G., Malladi, P., Quan, D., Legregni, L. & Hall, H. (2007). Cloud Computing, IBM Technical Report: High Performance on Demand Solutions (HiPODS).
- Dignan, L. (2011). *Cloud computing market: \$241 billion by 2020* [Online]. <http://www.zdnet.com/blog/btl/cloud-computing-market-241-billion-in-2020/47702>. [Accessed 3 June 2014].
- Economist, T. (2010). *Tanks in the cloud* [Online]. www.economist.com/node/17797794. [Accessed 3 June 2014].
- Garrison, G., Kim, S. & Wakefield, R. L. (2012). Success Factors for Deploying Cloud Computing. *Communications of the ACM* **55** (9), 62-68.
- Hayes, B. (2008). Cloud computing. *Communications of the ACM* **51**, July (7), 9-11.
- Levina, N. & Ross, J. W. (2003). From the vendor's perspective: exploring the value proposition in information technology outsourcing. *MIS Quarterly* **27** (3), 331-364.
- Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J. & Ghalsasi, A. (2011). Cloud computing - The business perspective. *Decision Support Systems* **51**, 176-189.
- Teece, D. J. & Pisano, G. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal* **18** (7), 509-533.
- Venters, W. & Whitley, E. A. (2012). A critical review of cloud computing: researching desires and realities. *Journal of Information Technology* **27**, 179-197.
- Willcocks, L., Venters, W. & Edgar, A. W. (2013). Cloud sourcing and innovation: slow traing coming? - A composite research study. *Strategic Outsourcing: An International Journal* **6** (2), 184-202.
- Willcocks, L., Venters, W. & Whitley, E. A. (2014). *Moving to the cloud corporation: How to face the challenges and harness the potential of cloud computing*, Hampshire, Palgrave Macmillan.
- Yin, R. K. (2003). *Case Study Research: Design and Methods*, Third Edition, California, SAGE Publications.

Poster presentation abstracts

The Applicability of Western Type E-commerce to Saudi Arabian Entrepreneurship

Khulood Al-Mani

Department of Computer Science, Loughborough University

Aims

The poster outlines Khulood Al-Mani's PhD, which commenced in October 2013.

Prior studies have examined and proposed theories regarding the drivers and barriers influencing entrepreneurial adoption of e-commerce in developed Western countries. However, limited research on this topic has been conducted in the context of developing high-income Arabian Gulf economies, particularly in the Kingdom of Saudi Arabia; especially as the adoption of e-commerce in Saudi Arabia is a recent phenomenon. Therefore, there is potential to explore this area by comparing the Saudi environment to Western economies in terms of factors impacting the success or failure of entrepreneurial Internet start-ups. The examination of particular factors impacting e-commerce practices in the context of Saudi Arabia is important because some factors are likely to be unique and directly related to the Saudi environment but viewed as marginal in developed Western nations. Examples include the high cost of ISPs, the lack of sufficient infrastructure, the absence of legislations and governmental frameworks alongside Saudi-specific issues relating to culture and gender.

This research attempts to fill this gap in identifying the factors that affect the development of successful entrepreneurial e-commerce businesses and to explore how e-commerce is adopted by Saudi Arabian entrepreneurial firms. This poster aims to introduce the literature behind entrepreneurship and how entrepreneurship in Saudi Arabia may be being influenced by the barriers and drivers affecting e-commerce.

The following sections introduce the salient points arising from a literature study looking at e-commerce –related entrepreneurial activity in Saudi Arabia.

Background Literature: Entrepreneurship

Entrepreneurship, According to Frank H. Knight and Peter Drucker is about taking risk. The behaviour of the entrepreneur reflects a kind of person willing to put his or her career and financial security on the line and take risks in the name of an idea, spending much time as well as capital on an uncertain venture. Schumpeter suggested that innovative entrepreneurship skills in pursuing opportunities lead to economic development, including the introduction of “new combinations”, ideas or new organizational methods. According to Varadarajan and Yadav, E-commerce is “a networked information system that serves as an enabling infrastructure for buyers and sellers to exchange information, transact, and perform other activities related to the transaction before, during, and after the transaction”.

Two types of entrepreneurships have been proposed: “necessity entrepreneurship and opportunity entrepreneurship”. Necessity entrepreneurship results from failure in finding a satisfactory job; entrepreneur thus is being forced into self-employment. This may not lead to economic growth or it may even create underdevelopment. Opportunity entrepreneurship occurs because of capitalizing on unexploited or under-exploited opportunism, which in turn may lead to economic growth.

Most scholars study entrepreneurship by linking it to economic development and innovation. Schumpeter attempted to link entrepreneurship with economic development, while Frank Knight distinguished between risk and uncertainty, by proposing that risk is, a measurable quantity while uncertainty is “unquantifiable estimate or judgment”, at the same time he identified the role of the entrepreneur as the role of discovery. In addition to job creation, other scholars have emphasized the role of human capital (Becker), motivation (Miniti et al.), profit and survival (Wagner and Block), and gender (Clain). Mueller was one of the first authors to examine the contribution of e-commerce to agricultural economics. Amit and Zott (2001) attempted to look into various theories of value creation and extended these theories to e-business, while Hasan and Harris (2009) found synergies between entrepreneurship and innovation in three leading e-companies.

Background Literature: Entrepreneurship, e-Commerce and Saudi Arabia

The adoption of e-commerce in Saudi Arabia is a recent phenomenon. Drivers for e-commerce and entrepreneurial activity in Saudi Arabia are strong. The country is a rapidly growing market with 13,000,000 users in 2012 (up from 200,000 in 2000). It has the largest population of the GCC (Gulf Cooperation Council) countries, recorded at 30 million people in 2013, and expected to reach more than 30 million in 2014, with 49% of the population using the internet. While the concept of linking entrepreneurship and e-commerce is novel, this researcher may adopt some of the drivers and barriers of e-commerce which have been identified by some scholars and apply them to Saudi e-commerce start-ups.

Five major themes will be examined to identify potential drivers to e-commerce adoption among entrepreneurs in the country, which would enhance their “opportunity entrepreneurship” such as organizational factor, technological factors, and governmental support. Organizational factors include: innovative attitude, strategy and acceptance of change. Technological factors include: information and Communication Technology (ICT) infrastructure, high speed Internet access, secure online payment system and reliable delivery of goods and services, and governmental support which include: entrepreneurship supporting organizations, financial support through start up fund loans, educational and training institutions and incubators. Other potential barriers such as lack of legal framework which includes: e-commerce laws and regulations to protect the rights of both parties (vendors and customers) including credit-card fraud protection. Additional barriers could be some cultural and social factors which include: lack of customers' trust and security, gender norms, customers' readiness and risk aversion. All of those factors are instrumental in opening vistas for new and innovative Saudi business, and not necessarily impelling those entrepreneurs into those ventures out of necessity.

Proposed Methodology

While this research is still in its early stages, the intention is to take a multiple case study approach, based on a number of Saudi entrepreneurial start-ups. The research will focus on a small number of established and successful cases where entrepreneurial activity has been enhanced by the opportunities afforded by e-commerce, and to compare these with a number of smaller entrepreneurial enterprises to examine the similarities and differences between them. Case studies will be embraced targeting two successful established entrepreneurial organizations and three start-ups. The criteria of selection of particular organizations are to meet the research objectives. Therefore, purposive samples are intended to be the underlying sampling technique representing both opportunity and necessity driven entrepreneurial firms. Furthermore, sampling will consider cultural and social factors associated particularly to the Saudi environment such as gender and infrastructure along with legislation related issues. Then

intention is for semi-structured in-depth interviews to be utilized by the researcher to identify the most important drivers and barriers to e-commerce adoption by entrepreneurs. This will allow the researcher to collect primary data based on open type questions with several people in each organization representing different sectors starting from top management (CEOs), moving to technical and financial staff allowing people to express their ideas in their own words, raising issues and concerns of the topic being discussed for in-depth explanation and understanding of the research problem. Furthermore, this will enable the researcher to gather information and incorporate elements related to both e-commerce and entrepreneurial businesses contributing to the study trying to explain the relationships between those elements.

Thus, an interpretive qualitative research approach will be the underlying methodology of this thesis.

The ultimate goal is to develop a model that would describe and explain the factors influencing e-commerce related entrepreneurial activity in Saudi Arabia and suggest recommendations that could enhance entrepreneurial activity in similar environments in other developing countries. These recommendations could be targeted at entrepreneurs, as well as at authorities in developing an e-commerce-related infrastructure, training and policies that support entrepreneurial activity in the region.

A digital Identity Management Model for the Jordanian online press and publication law to sustain the national e-commerce strategy

Adi Al-Waked

School of Business and Economics, Loughborough University

Purpose and Background

The purpose of this research is to provide an example of a case study that involves a technical database organisation, especially under a political unrest, which might entail passing some laws that could contradict with human rights or the e-business development.

The case study of this research is the Jordanian Press and Publication Law amendments that were expanded recently to include the online media publications, these modifications caused a big controversial between the Jordanian government and the civilian rights organisations, including the international human rights groups that frequently called the government to revoke these modifications, so far, without any positive response, as a result, Jordan was classified as a “Not Free” Internet Country in 2013, comparing to a “Partly Free” status in 2012 according to “Freedom House Organisation” Latest Reports.

The law amendments specify new business and technical requirements, the business requirements treated the online media in the same manner as the traditional publications, and obligate the online media websites to obtain a license from the government in order to work in the online media, while the technical modifications state that the online media websites should keep an archived record for their users identities, and should refrain from publishing any kind of comments that are deemed offensive; moreover, the website owners and the online commentators can be held accountable for publishing any abusive content online (from the commentators or the publishers).

The website owners objected against these new requirements since they don't see them practical, more specifically in the technical requirements, such as the identification of the online commentators identity, especially with the absence of a formal standardised digital Identity Management Model in the country as a whole, a model that can be considered as a benchmark through which the violations and breaches can be compared against, therefore, the online media publishers are convinced that this law targets the media freedom and doesn't support the online journalism in the age of internet development.

The Importance of the Issue

The competition in the internet media depends on several factors, such as the level of freedom, reliability, and independence of the online media, which are the main attractions for the readers to become frequent visitors to an online news website, therefore, when limiting the freedom of expression on both the users and the publishers, the number of daily hits on the websites will decrease, which will eventually impact the online business for the online media websites as a whole, and will reduce the profits from the advertisements as a result.

This research sees that the current law didn't produce any standards for the technical solutions, and that the government didn't contribute in facilitating the process, since the identity of the online users can easily be manipulated through different IP circumvention tools, and this can hold the wrong people the responsibility of something they didn't do.

The research proposes a mechanism to organise all the different online websites through one standardised database with a model that will identify the online users identity without any error, or at least with a low error margin, this will lead to a better organisation to both the business and the technical data, and will support the e-business of the online media, as well as the online freedom rights of the netizens.

Methodology

The validation of the proposed model will be done by assessing the stakeholders satisfaction (the website owners) on the proposed model, a simple qualitative approach will be followed, the reason for limiting the validation on the online media publishers is that the publishers have the power to suggest proposals to modify the regulation, since some parliament members asked the online publishers to propose new amendments on the law in order to discuss them in the parliament sessions.

The research also explores the opinion of the online authors, such as the columns writers, and whether they are willing to cooperate in the proposed digital identity management model.

Findings

About 80% of the explored publishers sample agreed on having an identity management model as long as it can be implemented and solves the issue of the online comments, and the rest 20% thought that the abusive online comments are easily to be detected and don't form an issue, but they see the whole law to be impractical, however, they don't mind to have an identity management model as a final choice if it can retain their readers.

The writers sample proved better results, as all the interviewees didn't mind having their online identities managed and recognised, moreover, they believe this can reduce the risk of abusive articles that are written by fake names, however, they expressed some concern that the online writers might be less inclined in writing articles, since they might feel that the government is targeting their online freedom.

Discussion

So far, no single study has investigated the effect of the technical data organisation in the online media regulations on the interrelated factors, such as e-business and human rights, furthermore, according to the author's knowledge, no research addressed the issue of digital identity management of the online news websites users.

One study was done on a similar law that was passed in Uzbekistan, the study highlighted the effects of this law on human rights and freedom of expressions, still no study addressed a technical solution yet for such a law, therefore, this research aims to produce a new piece of knowledge that addresses this issue and similar issues in the online related laws that are passed during the political unrests and don't consider the technical related aspects, such as the digital identity management.

The suggested federated identity circle in this our study includes all the licensed websites, so that the users can login by using one login credentials to all the websites in the circle, therefore, a future research can focus on the ability of the government to keep the licensed websites database transparent and without being violated.

Conclusions

Implementing a standardised federated digital identity management model for the online media websites in Jordan helps in managing the right identity of the online abusers and solves the issues of the online accountability; moreover, it helps to sustain the online media business and the e-commerce national strategy as a whole.

Developing an information security policy framework for Saudi Arabia and Middle Eastern countries

Hend Alkahtani

Loughborough University

Introduction

The purpose of this study is to implement an information security policy framework customised for regional use in Middle Eastern countries focussing principally on Saudi Arabia. The research uses the case study of Princess Nora Bin Abdulrahman University (PNU) to develop and apply a framework designed to take into consideration Saudi Arabian culture, laws and regulations and any local university regulations and culture. Information security policy clearly defines and documents all the needed standards, guidelines, procedures, rules and responsibilities of an organisation to help reach its information security goal and should be implemented to satisfy the organisation's legal, professional, social and ethical security requirements.

The main objective of information security policy is the security of all information system data, whether it is printed or written on paper, stored electronically, sent by post, fax or electronically, or presented in video or audio format, and the security of the information infrastructure such as computers, programs, networks, facilities, people and processes. The literature and standards relating to information security shows this information security policy should consist of:

1. A clear definition of the information security scope, objectives and the need for information security.
2. Management support and endorsement.
3. Information security implementation requirements including:
 - a) Compliance
 - b) Training and education
 - c) Business continuity management
 - d) Action required following the violation of information security
4. Responsibilities specification.

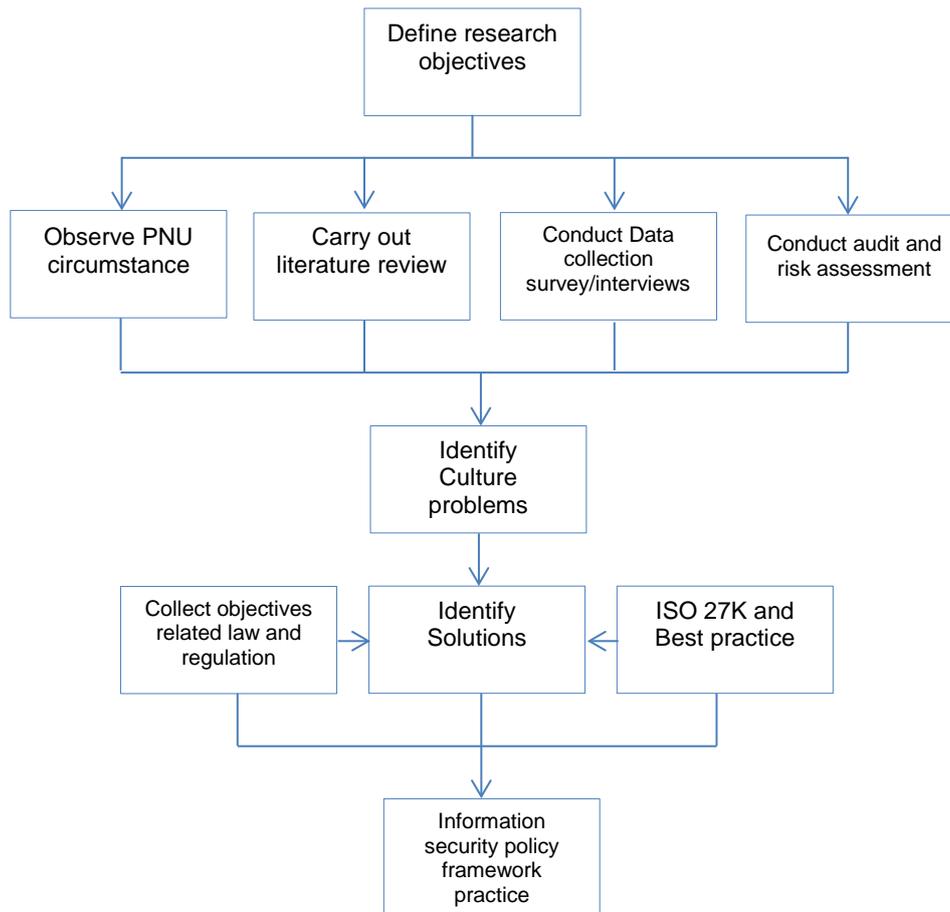
Writing an information security policy is not a simple task and must be conducted with due care. The need to involve employees to help in implementing and writing policy is important if it is to be successfully integrated into an organisation such as PNU. Therefore, the research aim is to develop an information security policy framework and workflow that is tailored to the Saudi Arabian culture and environment that can then be used by institutions, such as PNU, to implement its own security policy.

The development of the information security framework at PNU

The aim of the case study at PNU is to gather the requirements of a Saudi Arabian institution to create a framework consisting of a collection of controls, processes and standards for information security, along with guidelines and workflows to indicate where, when and how these should be applied. Of particular interest will be shaping the framework to work within the traditions and cultural constraints of Saudi Arabia. The cultural constraints are paramount at PNU as it is a university with all female academic staff and students but with a male dominated IT service department. PNU will be used to test the derived framework before it is applied to other institutions in Saudi Arabia.

The initial step was to identify and establish the research objectives and carry out an in-depth literature review to identify and review the role of information security policy and best practice in securing an information system. This was followed by primary data collection through a survey and interviews to investigate the state of information systems at PNU. Other implementation requirements include the considerations of the Saudi Arabian Ministry of Higher Education, and relevant laws and regulations. International and regional standards, such as ISO 27k and international best practice were considered and adapted in the implementation and development of the information security policy and framework (see figure 1).

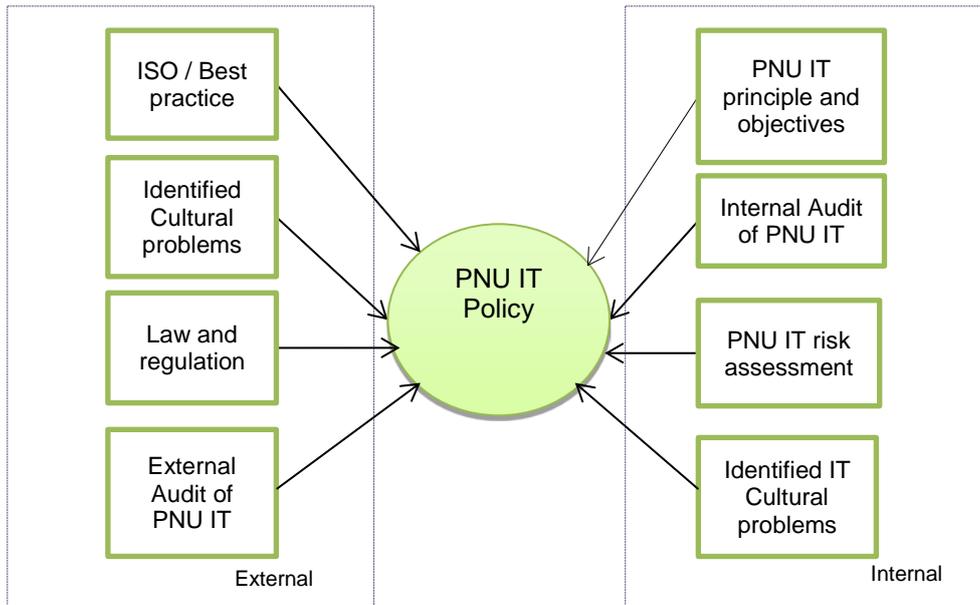
Figure 1: The development of information security framework



Formulation of information security policy at PNU

According to the international standard, BS ISO/IEC 27002:2005, organisations must identify their information security requirements prior to the implementation of information security policy. The main sources of requirements to establish the PNU information security policy are presented in figure 2.

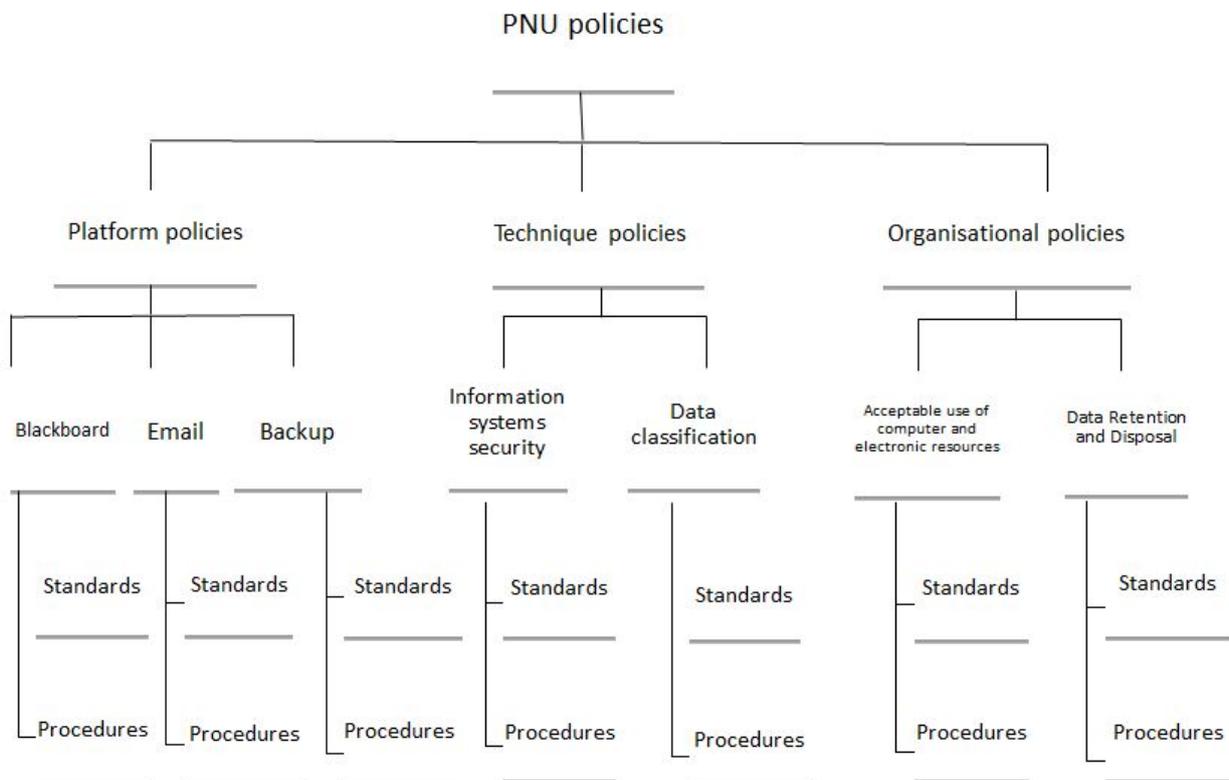
Figure 2: Requirements for information security policy structure at PNU



PNU required policies

PNU needs policies in three areas as seen in figure 3. Platform policy is one of these areas focussing on email policy, backup policy and Blackboard (the online learning environment used at PNU) policy. Other policy areas required at PNU are technique policy, such as data classification and information systems security, and organisational policy, such as acceptable use of computer and electronic resources policy and data retention and disposal policy.

Figure 3: PNU policies Framework



PNU information security framework

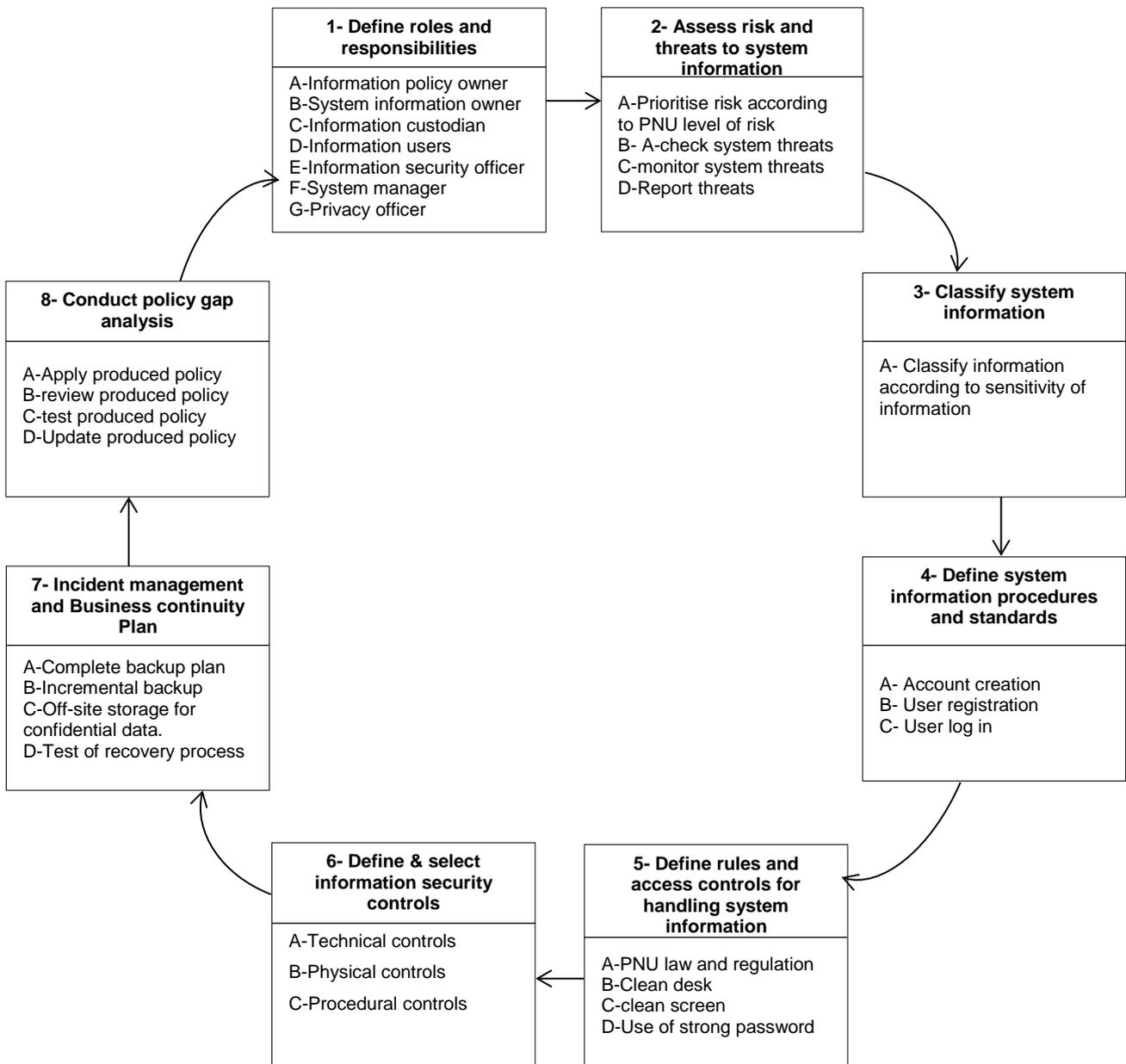
The first steps of implementation of the information security policy framework are to:

- Help employees at PNU understand the importance of the information they are handling.
- Clarify to all PNU employees the danger of misusing this information.

Specify the employee responsibilities, cooperation, compliance, and involvement in the making of the IT policy.

Figure 4 is a workflow that can be used by PNU Information system owner to implement its own information security policy.

Figure 4: Information security policy framework workflow



After the information system owner has created a clear definition of roles and responsibilities, then the information system owner and employees should assess risk and threats to system information. After that they should classify system information according to the sensitivity of

information: public, open, confidential and strictly confidential. The information system owner and employees should define system information procedures and standards. They should also define rules and access controls for handling system information. Next, the IS owner and employees should define and select information security controls. They should introduce incident management and a business continuity plan. Finally they should conduct policy gap analysis by applying the produced policy, reviewing produced policy, testing the produced policy and updating the produced policy.

The information system at PNU may technically resemble other systems in international use. However, it is managed, processed and used by people with unique culture and environment constraints that may not be seen in any other country. Therefore, building an information security system needs to involve these people and they should contribute to the making of the policy for their IT system. The principle behind this approach is that information security policy produced at the end of the process should give the best possible fit to the system requirements since it is mainly implemented, reviewed, tested and updated by the owner and employees of that system.

Current status of the research and future work

A trial model of the information security policy framework has been created, based on the steps indicated in figure 4. The next stage in the research will be to apply the framework at PNU to derive an information security suitable policy appropriate to that organisation. It is expected that this will lead to further development and refinement of the framework. Once the framework has been developed and implemented at PNU to a level that satisfies all stakeholders, then the policy framework will be applied to other universities in Saudi Arabia with the overall aim of getting the information security policy framework adopted by the Saudi Arabian government as a standard approach to be used by all universities in the country.

A framework for identifying suitable cases for using market-based approaches in industrial data acquisition

Torben Jess, Philip Woodall and Duncan McFarlane

Department of Engineering, University of Cambridge

Introduction

Industrial companies when working with datasets often face the problem of making sure the right user is using the right information for his decision-making. They often have various internal datasets, which have high maintenance costs and are also considering using open data or buying additional external datasets. This leads to a problem of inner company resource allocation, where the right data resources need to be allocated to the right users. Markets have been shown to work well for resource allocation problems. In order to test and further develop a market-based approach for industrial data management, suitable test cases need to be identified. This paper explains what characteristics need to be fulfilled so that a market-based approach for industrial data management can provide benefits. Using interviews to identify industrial cases and literature reviews to identify typical criteria for using a market-based approach, we developed a qualitative framework to identify possible cases to use markets in data management.

Sample cases

Data acquisition can be a complex problem due to various factors like varying utility to different data users but also due to complexity of large organizations, for example. The following two examples show this problem for two realistic industrial example use cases.

Example A: Company X is a manufacturing company with 100,000 employees. These employees work in a broad range of products and functions. Within this company employee H. Smith in the IT division has the option to acquire dataset containing credit data for suppliers for the costs of 10,000 USD every year. He has a sample data set from the company for one month to test the data set in his organisation. H. Smith thinks that this data could be relevant to some employees in the company. He knows that the supplier risk analysis division of 10 employees could use this data to predict potential problems with the supplier. An initial estimate from each of the employees tells him that this data set would save 800 USD per employee of the supplier risk division each year. However, H. Smith thinks that additional divisions like the procurement division with 50 employees could use this information, in order to negotiate additional discounts from their suppliers. Therefore H. Smith is unsure whether he should acquire the data set. Due to the large number of employees, H. Smith cannot ask all potential divisions in the company, due to the large number of employees, distances between employees and due to the fact that he does not know the complete organization.

Example B: Company Y is a large machine manufacturing company with 150,000 employees. The service division has 15,000 employees and offers mainly provision of spare parts but also repair services to the companies owning the large manufacturing products of company Y. The customers of Company Y asking for potential spare parts or specific services typically rely on Company Y for support. However, Company Y has little insight into their customer's machines and processes and is therefore unsure about orders of spare parts - orders from customers often "surprise" them. Company Y would benefit from further insight into their customers' data. The problem is that customers would only sell their data for very significant costs of 300,000 USD on average per machine as a customer because the data has high value to them. Manager J. Anderson wants to run a trial activity by acquiring data from 5 customers for 1.5 Million USD. He knows that the supplier forecasting division of 10 employees would benefit from

this data, but he is unsure to which extent this data would improve their operation. Each employee in the operation division will be able to give him a rough estimate for the value that this data created for him. He also does not know where else in the organization this data might be useful.

These problems illustrate the type of problems potentially addressed by markets. They form the basis for the problem space identified in the next section.

Characteristics of the problem space for a market-based approach in data acquisition

Generally a market-mechanism is applicable in a data acquisition context to situations where the value of a piece of data is required in order to make better data acquisition decisions. Therefore the following conditions should always be satisfied if a market-based approach is considered applicable:

- User is using data from multiple (possibly changing) data sources
- User has a set of offers (from data providers) to acquire more or different data to improve his decisions
- User knows the value of a certain piece of data or a combination of data pieces in terms of contribution to a decision
- Data provider knows the costs associated with its provision

These characteristics describe all situations where a market-mechanism could potentially be deployed. However, in order to provide benefits to industrial companies, a market-based approach also needs to outperform existing approaches. Therefore, additional characteristics focusing on the difficulties that exist within data acquisition and other applications are required. These characteristics are the following:

- Partial information with data users and/or data providers
- Heterogeneous environment for data users and/or data providers
- Distributed decision making between data user and data provider

These characteristics can then be checked against our scenarios to see if they fit our description for the problem space.

Characteristics of a market mechanism to address the problem space

A market has various characteristics that can help to address this problem space because of the following characteristics:

- Allocation of resources given value and costs
- Formalism and abstraction of problems
- Faster and/or more accurate computation of solutions
- Flexibility and extensibility in the calculation process
- Incentivises participants and increases efficiency

These different characteristics can be further classified to show a more detailed framework. When combining them with the sample cases from section 2, it shows that the market can help address the problems described within our problem space.

Future work

Future work aims to use this framework for the identification of scenarios that could then feed into the further development of market-based approaches in data management.

Developing Data Quality Metrics for a Product Master Data Model

Marhendra Lidiansa

Delft University of Technology, The Netherlands

Purpose or background

Master data management (MDM) is implemented to increase the quality of core business data by having a single managed repository. Like any other IT projects, there are failures in the implementation of MDM. Several main causes of failures in MDM implementation are related to a missing data quality process, for example, a lack of proactive data quality surveillance (Sivola et al., 2011) and a lack of data quality measurements (Haug, 2013). An important phase in the data quality process is the measurement phase that exercises the data quality metrics. In accordance with Elsevier's plan to implement product master data, the main objective of this study is to identify, collect, analyze, and evaluate the quality metrics for a product master data; to allow quantifying and improve their value.

Design, Methodology, Approach, and Findings

In order to meet the main objective, this study needs to address these three questions: (1) What is the type of methodology that should be used to develop business-oriented data quality metrics? (2) How appropriate is the methodology for a practical case? (3) What are the data quality metric specifications for a case study in Elsevier? There are four phases in this thesis work to develop and answer those questions. In the first phase, the introduction phase, the main objective and research questions are formulated with several considerations, particularly the scientific and practical benefit of the study, and the boundaries of the projects.

In the second phase, the conceptualization, we need to select or construct the general process framework (GPF) to develop business-oriented data quality metrics as the answer to the first question. This study selects methodology developed by Otto, 2009 as the GPF to develop the data quality metrics. The selection process is conducted by comparing the methodology with other methodologies—like AIMQ, TIQM, DQ-P, and ORME-DQ—on several features, for example, the process model, metamodel, business needs consideration in its data quality criteria, and the focus of the method. Other studies in data quality (DQ) requirements, DQ metrics specification, DQ metrics requirements integration, data modelling, and process modelling are also used to ensure that the process model and metamodel in the selected GPF are adjustable for the case study.

The background of this thesis is related to the MDM system which function is to provide numerous enterprise applications with high quality critical business objects. Thus, we need to make sure that the developed data quality metrics meet the requirements of several business applications (Loshin, 2010). This thesis uses the process model developed by Wang et al., 1995 and Batini et al., 1986 as the GPF to integrate data quality metrics from several applications into the product MDM repository. The activities include developing the appropriate data models, making the schemas conformed, and conflict resolution using qualitative criteria—completeness and correctness, minimality, and understandability. The result is a list of feasible data quality metrics that meet the needs of several applications.

Here the thesis work uses the GPF as the first version of the developed solution to address the main objective. The main processes in the GPF are identifying the business problems and data defects, specifying the data quality requirement and metrics, verification of the result, and integrating the data quality metrics.

The thesis work addresses the third phase, the validation, by executing the GPF for a case study in Elsevier. Each process is adjusted with the case and analyzed for the required alterations. The activities in this process consist of literature study and workshop/interview with the domain experts. The result of this activity is the altered GPF as the developed solution to address the main goal. The changes consist of an alternate configuration for the process model and the tangible objects for the components in the metamodel, for example, interview questionnaires, data quality requirements, data quality attributes, business problems–data defects matrix, and the data quality metrics. These results are used in formulating the answers to the second question. This thesis also conducts the testing activity by assessing the developed metrics with the criteria in the data quality metrics requirements.

The developed and filtered data quality metrics are feasible for the study case in Elsevier and include several data quality dimensions, namely completeness, syntactical correctness, absence of repetition, absence of contradiction, and accuracy. Those data quality metrics are the answers for the third question.

Conclusion

This thesis addresses its main objective by having two main results, namely a list of data quality metrics for eCommerce and product MDM system in Elsevier, and a process model to develop data quality metrics for a product MDM. The process model is developed on the basis of the works by Otto, 2009; Wang et al., 1995; and Batini et al., 1986, and considered practical, valid, complete, and resilience (Woodall, 2013). This study also provides several lessons, for example, the critical success factors for each phase in the process model, recommendations for data quality process in Elsevier, and updates for product MDM data model. Furthermore, studies on the same issue with several other data/process domains are needed to get other possible configurations of the process model.

A taxonomy of crowdsourcing systems

Helen Thomas

Loughborough Design School/Ordnance Survey

Ordnance Survey, who are part-funding this research, are interested in the potential offered by crowdsourcing which involves geographic data, and wish to have this investigated from a user-centred design perspective. With the emphasis of the PhD being centred so heavily on people, the intention is that data collected and investigated will mainly be qualitative in nature.

Jeff Howe, one of the authors of the term crowdsourcing, defines it thus: “Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large, group of people in the form of an open call.” Crowdsourcing is proving very successful in the world of business – how is it succeeding in the world of geographic scientific or societal activities?

The aim of this research is to first identify the motivations, barriers and enablers for crowdsourcing. This knowledge will then be used to produce guidelines for future design of crowdsourcing platforms to enable them to be targeted towards appropriate contributors, to optimize usability, and to design the system as a whole in order to maximize contributions.

The investigation is being carried out by:

- A literature review
- Immersive study by the researcher in crowdsourcing platforms with a geographic basis which cover a range of dimensions. This study is being carried out by a mixture of task analysis, heuristic analysis and subjective assessment
- Work to understand different crowdsourcing dimensions
- Testing knowledge gained in the immersive study with participants, in order to understand better the different factors which act as barriers and enablers to crowdsourcing contributors, and how much they vary dependent on the dimensions of the crowdsourcing. This will hopefully lead on to the formulation and testing of guidelines for future crowdsourcing design.

The research is nearing the end of its first year, and work is currently in progress to produce a taxonomy of crowdsourcing with a geographic aspect. It is anticipated that such a taxonomy will be useful as a way to categorise different types of crowdsourcing site (something which has not yet been done in the literature). This in turn will provide a framework to structure a more detailed investigation of the barriers and enablers to crowdsourcing. For example, one of the dimensions in the taxonomy is whether the activity only happens at a fixed moment in time or whether it is anytime/ongoing. Of the pool of potential crowdsourcing contributors, some may find the defined time to be an enabler, some may find the opposite. So as the project develops the taxonomy dimensions will be related to user motivation and needs and the guidelines tailored to meeting those needs.

It is felt that it would be beneficial to share this early-stage taxonomy with others who have experience of working with crowdsourcing to gain feedback and criticism to help to further refine and improve it by the end of the PhD.

Research so far has indicated that there is a relatively low take-up of crowdsourcing contribution, particularly if the nature of the crowdsourcing is more technically detailed. However, the people who are willing to be contributors to crowdsourcing vary depending on the characteristics (dimensions) of the crowdsourcing. The RSPB Big Garden Birdwatch, (an example of an activity at a certain period in time) for example, attracted around 400,000 contributors this year, which is a very high success rate for an event that takes place during just one weekend, and the contributions come from a wide range of different people. OpenStreetMap (an example of an on-going collection), on the other hand, has a very clearly defined contributor demographic of predominantly men aged between 20 and 40 with professional qualifications.

This research aims to understand better how future crowdsourcing initiatives can be designed to maximize contributions from, and hence the benefits of, the 'power of the crowd'. The taxonomy is the first step towards a structured process by which to achieve this.

Balancing Big Data with Data Quality in Industrial Decision-Making

Philip Woodall and Maciej Trzcinski

Department of Engineering, University of Cambridge

Purpose or background

The volume and availability of datasets are increasing massively in today's industrial organisations. Despite its critical role in the past, however, the problems of data quality are sometimes being dismissed in this Big Data world as being irrelevant, and the importance of improving data quality is being discussed (Mayer- Schonberger and Cukier, 2013). One important question is "as we scale data, will the effects of errors in the datasets be reduced to an extent where data quality assessment and improvement activities are no longer needed" (Woodall et al., 2014)?

For example, consider the question of determining the performance of a particular supplier to a manufacturing company (see (Woodall et al., 2014)). A series of transactions with the supplier could be recorded in a database indicating whether the supplier delivered parts to the company on-time or not (*see Table 1*). The company could calculate the supplier's performance using the number of on-time deliveries over the total number of deliveries for that supplier. If, however, the supplier is referred to by two different names (the quality error being that they should be consistently referred to), then some of the deliveries for that supplier may not be counted. This can happen when a query to the database requests all records where company is "Air parts Ltd." and this query would thus not return records for "Air pts". The result is that this may give an inaccurate performance figure, as the example in Table 1 shows: the performance of the "Air parts Ltd" supplier would be counted as 75% (3/4). Correcting the error in the supplier names ("Air pts" to "Air parts Ltd.") would yield the actual performance to be, instead, 50% (3/6).

If the company has a policy to keep suppliers above a threshold of 70%, then correcting data quality in this case actually changes the decision: without correcting data quality the company would continue business with the supplier, whereas after correcting quality the company would cancel the contract with the supplier according to their policy of 70% delivery performance.

Table 1: Supplier transactions

Company	On-time delivery?
Air parts Ltd.	y
Air parts Ltd.	y
Air parts Ltd.	y
Air pts	n
Air pts	n
Air parts Ltd.	n

Aims/objectives

The question we aim to address is: whether industrial decision makers can arrive at the correct decision by only increasing the size of the dataset, without having to correct data quality?

In particular, whether collecting more data records (in the spirit of the Big Data era) will lead to a result where a few erroneous entries will not inflict such a large bias on the overall results, and the result will tend to the correct supplier delivery performance. In the broadest sense, this is asking what effect Big Data has on data quality in industrial decision-making scenarios.

Table 2: after obtaining more data records

Company	On-time delivery?
Air parts Ltd.	y
Air parts Ltd.	y
Air parts Ltd.	y
Air pts	n
Air pts	n
Air parts Ltd.	n
Air parts Ltd.	y
Air parts Ltd.	y
Air parts Ltd.	y
Air pts	y
Air pts	y
Air parts Ltd.	n
Air parts Ltd.	n
Air parts Ltd.	y
Air parts Ltd.	y
Air pts	y
Air pts	n
Air parts Ltd.	n

If we take the example from table 1 and include more data records (possibly from another data source, or by recording more deliveries etc.) and re-calculate the supplier performance in the same way, we obtain that the performance of “Air parts Ltd” is 66% (8/12). The company therefore arrives at the correct decision to cancel the contract with the supplier because this value is below the policy threshold of 70%, and this has been achieved without spending time correcting data quality. Note that in this case the “actual answer” (after correcting data quality in table 2) is 61% (11/18).

In this motivating example, it is clear to see that the situation has been conveniently setup to reveal a desirable outcome because the errors “cancel each other out”. In real situations this may/may not be the case and it is our intention to determine whether, and in what particular cases, increasing data volume is an equivalent or improvement over correcting data quality and for what types of decisions etc.

Design, methodology or approach

In order to address our research question, we focus on the volume aspect of big data and investigate the effects of the completeness and inconsistency data quality dimensions (Wang and Strong, 1996) on the result of scenarios, similar to the example above when more data records are appended.

We used semi-structured interviews with managers in various industrial organisations to determine relevant scenarios/decisions. We also used software simulations, originally developed for the research in (Woodall et al., 2014), to assess the effects of the following criteria on these scenarios/decisions:

- Distribution of errors,
- Frequency of errors,
- Types of errors (the different types of inconsistencies and completeness issues)
- The complexity of the data (number of distinct data values etc.)
- The nature of the decision (based on thresholds, types of statistics used etc.)

This is an ongoing study where we expect to report the findings within the next two months in the full paper for the conference.

References

Mayer-Schonberger, V. and Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*, John Murray.

Wang, R.Y. and Strong, D.M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, 12 (4), p.pp.5–34.

Woodall, P. et al. (2014). An investigation of how data quality is affected by dataset size in the context of Big Data analytics. In *Proceedings of the International Conference on Information Quality*.

