Research on Conceptual Modelling: Less Known Knowns and More Unknown Unknowns, Please

Jan Recker
Information Systems School
Queensland University of Technology
PO Box 2434, Brisbane 4001, Queensland

j.recker@qut.edu.au

Abstract

In this invited paper I describe some personal views on the research field of conceptual modelling. I argue that the field has become entrenched in some "bad habits" that usually emerge in evolved paradigms and that we need to proactively pursue a dual research strategy incorporating new and different avenues that lead us to novel and impactful research contexts of conceptual modelling. I provide a framework that can guide this exploration and finish with some recommendations about how conceptual modelling research programs could proceed.

Keywords: Conceptual modelling, empirical research, research agenda, paradigm.

1 Introduction

> *There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know.*
>
> Donald Rumsfeld, U. S. Secretary of Defense, Statement to the Press on February 12, 2002.

At a press briefing in 2002, Donald Rumsfeld addressed the absence of evidence linking the government of Iraq with the supply of weapons of mass destruction to terrorist groups using this statement. The statement has become an internet sensation and not only sparked political but also linguistic quarrels. On the one hand, Rumsfeld was criticized for his application of the English language, but on the other hand some linguists lauded his statement as "impeccable, syntactically, semantically, logically, and rhetorically" (Pullum 2003).

I have argued before (Recker 2012) that Rumsfeld actually provides a quite useful frame to conceptualize research fields in terms of both what a field has been doing in the past and what it should focus on in the future.

In this invited commentary, I will use Rumsfeld's views and apply them to conceptual modelling as a research field. I will briefly state my understanding of this research field, its role and its major research streams and then outline some arguably personal views on what we should be researching and how. I do this not with the intention of discrediting the research efforts that have brought conceptual modelling to where it is today – in fact, this research has been so successful that conceptual modelling is now clearly considered a core area within Information Systems research. Instead, I wish to provide some stimulus for future research, especially for research on conceptual modelling that is notably different compared to the traditional trajectories of the field. This, I hope, will allow the community to further grow and expand by complementing new fields and approaches to inquiry with what we are researching anyways.

In what follows, I will first briefly recap my views on the conceptual modelling field. Then I will draw on discussions of scientific paradigms (Kuhn 1996; Lakatos 1970) to offer some arguments why I believe the research field of conceptual modelling is stuck in some habits that may not be entirely beneficial to its further development. Finally, I will use Rumsfeld's statement to develop a conceptual frame to guide research in conceptual modelling that will hopefully become even more impactful and relevant.

2 Conceptual Modelling as a Research Field

What is conceptual modelling? In writing this commentary I initially wondered not only what my own definition would be but also how the APCCM community would respond to this question. As one potential answer to this question, consider Figure 1. I extracted all paper titles from the conference proceedings of all APCCM events to date and generated a Wordle image from these titles. The resulting image visualizes the most prominent terms used in the APCCM papers published in the proceedings to date. Unsurprisingly, most papers deal with models of processes, data, business or other requirements, often in associated with the design of systems.

Figure 1: Wordle image generated from titles of papers from APCCM2004-2014

Looking beyond APCCM into the wider field of conceptual modelling literature, my view is that in most cases, we characterize conceptual modelling as a diagramming technique used in the requirements analysis phase of information systems analysis and design to capture, specify, communicate and document both static and dynamic phenomena of a real-world domain intended to be supported by an information system (Wand and Weber 2002). This view is probably most prominent in the field, and has given rise to research on questions about the grammars that we use for conceptual modelling (e.g., Shanks et al. 2010; Recker et al. 2009), which elements should be used in a conceptual model to aid analysts (e.g., Parsons and Cole 2004; Masri,Parker and Gemino 2008; Figl,Mendling and Strembeck 2013) or what makes an expert conceptual modeller (e.g., Shanks 1997; Batra and Davis 1992). It also sparked research in computer science and software engineering about better conceptual modelling grammars (e.g., Kwon 2011), better methods of modelling (e.g., Siau 1999), better modelling tools (e.g., Jarvenpaa and Machesky 1989) or better model management systems (e.g., Reijers,Mans and van der Toorn 2009).

I want to relax the assumptions that all conceptual models need to be explicated in diagrams and that these diagrams must be more or less formalized, or indeed that conceptual models always serve purposes of eventually leading to an IT-based implementation of some system diagrams. Instead, let us assume that, in more general terms, conceptual modelling can be seen as representing knowledge about a real-world domain, often deliberately constructed such that implementation concerns or details are omitted (Lukyanenko,Parsons and Wiersma In Press; Mylopoulos 1998). This view acknowledges that conceptual modelling is a useful technique to analyse and design information systems; but it also extends to further application areas. For example, we use conceptual modelling when we try to make sense of a real-world domain, we use it to analyse or redesign business processes possibly independent from any involvement of IT-based artefacts, and we use it simply to communicate our view of a real-world domain or phenomena to other people around us. In fact, we all engage in conceptual modelling all the time in our mind – but very rarely do we document these mental conceptual models explicitly in the form of diagrams: Expert data modellers for example are known to conceive real-world domains in terms of entities and relationships between them; but they may or may not feel the need to document these conceptual models in an entity-relationship diagram. Those trained in process modelling often conceive domains in terms of events that occur and sequences of tasks that are performed in reaction to those events. Even in emerging areas such as user-generated content it was shown that participating content-providers conceive of their content in different models – some of which are better or more effective than others (Lukyanenko,Parsons and Wiersma In Press).

3 Some Perceived Problems in the Conceptual Modelling Field

Why do I describe such a view of conceptual modelling? I do so because I believe that most existing research is stuck in a narrow conception of what the field is about, and in turn identifies and addresses only a limited set of research questions that are worth asking and exploring. In other words, I believe there are more research questions in and around conceptual modelling that remain hidden because of the way we have been thinking about conceptual modelling so far.

A stable and limited view of what core phenomena and core research questions are is of course not only a problem for the conceptual modelling field but in general a problem for all research fields that approach the status of paradigm development (Kuhn 1996).

Of course, paradigms per se are not a bad thing. In fact, most science endeavours wish to achieve a status where core ideas, core methodologies and core questions are well-known and accepted. This is also true for information systems where the yearning for a paradigm status is well and truly alive and active (Chen and Hirschheim 2004; Hassan In Press; Weber 1987).

Yet, there are also problems with paradigmatic research (Kilduff,Tsai and Hanke 2006). And it is my view that some or even most of paradigm problems are also evident in the conceptual modelling field:

1. Paradigms are *stable*: the phenomena of interest remain largely unchanged. The set of constructs to explain those phenomena, too, remain largely unchanged, and so do methodological and analytical procedures.
2. Paradigms are *predetermined*: Most research work within paradigms is on delineated smaller and smaller "puzzles" whose outcomes are largely conceivable ex ante because mostly, problems are chosen that can be solved in accordance with the paradigm.
3. Paradigms are *divided*: typically, researchers fall into two camps: those in support of the paradigm tend to gravitate toward the hard core of the paradigm and work on those problems without considering phenomena or challenges on the boundaries, and those in scepticism of the paradigm tend to remain on the fringe to debate or critique the limits of the paradigm.

I believe all three problems that appear often in most research paradigms (e.g., Kilduff,Tsai and Hanke 2006; Hassan In Press) also manifest in the conceptual modelling field. It is not my ambition to prove this conclusively in this commentary but I still want to present some arguments in support of this view.

For instance, one way to examine whether or not conceptual modelling has paradigm problems would be to examine the influential and prominent papers in this field. I ran a citation analysis using the Harzing Publish or Perish tool (Harzing 2010) to identify some of the most impactful papers on conceptual modelling. The search returned, amongst others, the following types of papers:

- Commentaries that introduce constructs to describe conceptual modelling (Wand and Weber 2002)
- Frameworks that describe quality in conceptual modelling (Lindland,Sindre and Solvberg 1994)
- Conceptual and ontological foundations of conceptual modelling (Hirschheim,Klein and Lyytinen 1995; Guizzardi 2005).

Note that these selected papers in essence provide definitions of the core constructs in the conceptual modelling field (such as the status of a grammar or the notion of semantic quality). Unsurprisingly, these papers have defined the universe of discourse for the field. They also set forth a very stable research agenda that builds upon quite a limited set of research questions such as:

- How do we evaluate the quality of conceptual models (e.g., Moody 2005; Mendling,Reijers and Cardoso 2007; Recker 2007)?
- How can we measure their correspondence to ontologies (e.g., Fettke and Loos 2007; Gehlert and Esswein 2007; Guarino and Guizzardi 2006; Hadar and Soffer 2006)?
- How do we devise methods and grammars such that models are better ontologically or qualitatively (Brinkkemper 1996; Schütz and Schrefl 2014)?

The way in which many researchers in the field attempt to address these questions, too, is very stable – and therefore somewhat limited. Ontological evaluations (Green and Rosemann 2004), for instance, are very prominent in analytical work. Empirical research largely builds on experiments with students (e.g., Bodart et al. 2001; Mendling,Reijers and Cardoso 2007; Recker 2013); and design research often relies on set-theoretics, calculus or similar mathematical formalisms (e.g., van der Aalst 2005) as the way to specify new constructs, grammars or methods. Different approaches to evaluation, empirical or design research on conceptual modelling do exist but they are few and far between.

Finally, it would appear that the conceptual modelling field is full of *proponents* of conceptual modelling, i.e., those of us that believe that conceptual modelling is important, relevant and worthy of attention. The natural scientific attitude that we bring to our research and studies is thus positive: we typically try to demonstrate the relevance, utility, benefits and other positive impacts and applications of conceptual modelling. While this a healthy and fruitful attitude, scientific progress also benefits from a critical attitude that pursues and explores potential failure as a means to better our understanding (Gray and Cooper 2010). By trying to understand where the limits of application and utility of conceptual modelling lies, and by studying the areas where conceptual modelling potentially yields negative rather than positive impacts, we in turn develop a much better understanding of how conceptual modelling actually operates and how we might use it most effectively. In fact, through trials of failure we may find out more about the success of conceptual modelling as well as its boundary conditions than by pursuing success directly.

4 Knowns and Unknowns

Where and how should conceptual modelling research be conducted? As visualized in Figure 2 and in analogy to the statement by Rumsfeld, we should move towards exploring unknowns – both known unknowns and unknown unknowns. The best pathway toward an exploration of the unknown is by moving from the known knowns to the known unknowns – the areas where we already know that we don't know enough (or nothing at all). From there, it is hopefully only a small step to explore the areas that we can't even anticipate yet.

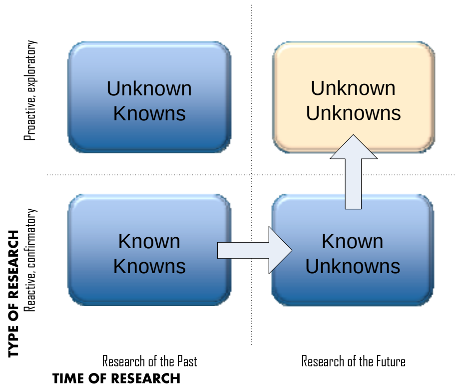

Figure 2: From Knowns to Unknowns. Adapted from (Recker 2012)

Note that I included two axes in the taxonomy in Figure 2. On the x-axis I differentiate *knowns* from *unknowns*, which helps separating research of the past from possible themes for the research of the future. The y-axis separates two different types of research – more *reactive* research to *confirm* known knowns and known unknowns – things that we know we know or things that we don't know, but at least we are aware that we don't know them. This type of research – and likely the approach to research – is different from the more *proactive* research required to *explore* the unknowns – things we didn't even know we know, and, perhaps more interestingly, the things we don't know that we don't know. Of course, this last category is the most challenging state to predict – but perhaps also the one that promises the most interesting findings?

In guiding the development of an agenda that moves conceptual modelling research towards the unknown, I have used the taxonomy in Figure 1 to develop several avenues for future conceptual modelling research. The outcomes of this effort are described in Table 1.

Note that the questions listed in Table 1 are subjective because they describe my own attempt at classifying both what we know and (which is much harder) what we do not know. I do not claim exhaustiveness or comprehensiveness of this list. In fact, some doubts remain that the suggested entries are entirely valid. The purpose of Table 1 is thus more to create a dialogue about what potentially the knowns and unknowns in conceptual modelling could be – and in turn where we should potentially direct our future research efforts.

Known Knowns	Known Unknowns
How can we create conceptual models that are understandable?	What is effective conceptual modelling?
How can we express data in structures that can be implemented in databases?	How do expert modellers model?
How can we generate software code from (some) conceptual models?	How do we manage version control in large model repositories?
Unknown Knowns	**Unknown Unknowns**
Conceptual modelling leads to shared understanding – but why and how?	What is conceptual modelling for big data?
Conceptual modelling reduces system design errors – but how much and why exactly?	Do all people model? Do they need to?
Novices create different models from experts – but why?	If we don't engage in conceptual modelling – how else do we make sense of the world(s) that concern us?
Which is the most effective modelling grammar?	Can conceptual modelling be automated?

Table 1: Selected Knowns and Unknowns in Conceptual Modelling

5 Conclusions

Conceptual modelling as a research field has matured into an established research area of information systems. Perhaps it is not as regarded in the same manner as research on technology adoption and business value of technology, but conceptual modelling stands as a cornerstone of the research discipline.

Yet, the standing and reputation of conceptual modelling within the discipline is not stable. As any other field, conceptual modelling research is rightfully under constant scrutiny in terms of its validity, applicability, relevance and utility in our ever-changing world.

To cement the place as a research field within information systems and surrounding disciplines it will be important to constantly review and revise our own research efforts on conceptual modelling.

In this paper I have argued that new pathways to research on conceptual modelling commences by (a) relaxing and challenging our own assumptions about what conceptual modelling is, and (b) moving our research efforts towards the fringes of the conceptual modelling paradigm, to areas where we are required to explore the unknown rather than confirm the knowns. In doing so, we may find out that conceptual modelling has its limits. But we will for certain increase our confidence in where, how and why conceptual modelling is effective and useful – and who knows, we may discover that conceptual modelling has premises and promises that we never foresaw.

6 Acknowledgments

The preparation of this paper was supported, in part, by a grant from the Australian Research Council (DE120100776). I am thankful to comments from Eike Bernhard on an earlier version of this manuscript, which helped improving the paper.

7 References

Batra, D. and J. G. Davis. 1992. Conceptual Data Modelling in Database Design: Similarities and Differences between Expert and Novice Designers. *International Journal of Man-Machine Studies* 37 (1):83-101.

Bodart, F., A. Patel, M. Sim and R. Weber. 2001. Should Optional Properties Be Used in Conceptual Modelling? A Theory and Three Empirical Tests. *Information Systems Research* 12 (4):384-405.

Brinkkemper, S. 1996. Method Engineering: Engineering of Information Systems Development Methods and Tools. *Information and Software Technology* 38 (4):275-280.

Chen, W. S. and R. Hirschheim. 2004. A Paradigmatic and Methodological Examination of Information Systems Research from 1991 to 2001. *Information Systems Journal* 14 (3):197-235.

Fettke, P. and P. Loos. 2007. Ontological Evaluation of Scheer's Reference Model for Production Planning and Control Systems. *Journal of Interoperability in Business Information Systems* 2 (1):9-28.

Figl, K., J. Mendling and M. Strembeck. 2013. The Influence of Notational Deficiencies on Process Model Comprehension. *Journal of the Association for Information Systems* 14 (6):312-338.

Gehlert, A. and W. Esswein. 2007. Toward a Formal Research Framework for Ontological Analyses *Advanced Engineering Informatics* 21 (2):119-131.

Gray, P. H. and W. H. Cooper. 2010. Pursuing Failure. *Organizational Research Methods* 13 (4):620-643.

Green, P. and M. Rosemann. 2004. Applying Ontologies to Business and Systems Modeling Techniques and Perspectives: Lessons Learned. *Journal of Database Management* 15 (2):105-117.

Guarino, N. and G. Guizzardi. 2006. In the Defense of Ontological Foundations for Conceptual Modeling. *Scandinavian Journal of Information Systems* 18 (1):115-126.

Guizzardi, G. 2005. *Ontological Foundations for Structural Conceptual Models*. Vol. 015, *Telematica Instituut Fundamental Research Series*. Enschede, The Netherlands: Telematica Instituut.

Hadar, I. and P. Soffer. 2006. Variations in Conceptual Modeling: Classification and Ontological Analysis. *Journal of the Association for Information Systems* 7 (8):568-592.

Harzing, A.-W. 2010. *The Publish Or Perish Book: Your Guide to Effective and Responsible Citation Analysis*. Melbourne, Australia: Tarma Software Research Pty Limited.

Hassan, N. R. In Press. Paradigm Lost … Paradigm Gained: a Hermeneutical Rejoinder to Banville and Landry's 'Can the Field of MIS be Disciplined?'. *European Journal of Information Systems* 24:forthcoming.

Hirschheim, R., H. K. Klein and K. Lyytinen. 1995. *Information Systems Development and Data Modeling: Conceptual and Philosophical Foundations*.

Cambridge, Massachusetts: Cambridge University Press.

Jarvenpaa, S. L. and J. J. Machesky. 1989. Data Analysis and Learning: An Experimental Study of Data Modeling Tools. *International Journal of Man-Machine Studies* 31 (4):367-391.

Kilduff, M., W. Tsai and R. Hanke. 2006. A Paradigm Too Far? A Dynamic Stability Reconsideration of the Social Network Research Program. *Academy of Management Review* 31 (4):1031-1048.

Kuhn, T. S. 1996. *The Structure of Scientific Revolutions*. 3rd ed. Chicago, Illinois: University of Chicago Press.

Kwon, S. J. 2011. Conceptual Modeling of Causal Map: Object oriented Causal Map. *Expert Systems with Applications* 38 (1):360-370.

Lakatos, I. 1970. Falsification and the Methodology of Scientific Research Programs. In *Criticism and the Growth of Knowledge*, eds. I. Lakatos and A. Musgrave, 91-132. New York, New York: Cambridge University Press.

Lindland, O. I., G. Sindre and A. Solvberg. 1994. Understanding Quality in Conceptual Modeling. *IEEE Software* 11 (2):42-49.

Lukyanenko, R., J. Parsons and Y. F. Wiersma. In Press. The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-Generated Content. *Information Systems Research* (forthcoming).

Masri, K., D. C. Parker and A. Gemino. 2008. Using Iconic Graphics in Entity-Relationship Diagrams: The Impact on Understanding. *Journal of Database Management* 19 (3):22-41.

Mendling, J., H. Reijers and J. Cardoso. 2007. What Makes Process Models Understandable? In *Business Process Management - BPM 2007*, eds. G. Alonso, P. Dadam and M. Rosemann, 48-63. Brisbane, Australia: Springer.

Moody, D. L. 2005. Theoretical and Practical Issues in Evaluating the Quality of Conceptual Models: Current State and Future Directions. *Data & Knowledge Engineering* 15 (3):243-276.

Mylopoulos, J. 1998. Information Modeling in the Time of the Revolution. *Information Systems* 23 (3-4):127-155.

Parsons, J. and L. Cole. eds. 2004. An Experimental Examination of Property Precedence in Conceptual Modelling. *1st Asian-Pacific Conference on Conceptual Modelling, Dunedin, New Zealand*. Dunedin, New Zealand: Australian Computer Society.

Pullum, G. K. 2003. Language Log: No Foot in Mouth. http://itre.cis.upenn.edu/~myl/languagelog/archives/000182.html (accessed June 21).

Recker, J. 2007. A Socio-Pragmatic Constructionist Framework for Understanding Quality in Process Modelling. *Australasian Journal of Information Systems* 14 (2):43-63.

Recker, J. 2012. BPMN Research: What We Know and What We Don't Know. In *Business Process Model and Notation - BPMN2012*, eds. J. Mendling and M. Weidlich, 1-7. Vienna, Austria: Springer.

Recker, J. 2013. Empirical Investigation of the Usefulness of Gateway Constructs in Process Models. *European Journal of Information Systems* 22 (6):673-689.

Recker, J., M. Rosemann, M. Indulska and P. Green. 2009. Business Process Modeling: A Comparative Analysis. *Journal of the Association for Information Systems* 10 (4):333-363.

Reijers, H. A., R. S. Mans and R. A. van der Toorn. 2009. Improved Model Management with Aggregated Business Process Models. *Data & Knowledge Engineering* 68 (2):221-243.

Schütz, C. and M. Schrefl. 2014. Variability in Artifact-Centric Process Modeling: The Hetero-Homogeneous Approach. In *10th Asia-Pacific Conference on Conceptual Modelling*, eds. G. Grossmann and M. Saeki, 29-38. Auckland, New Zealand: Australian Computer Society.

Shanks, G. 1997. Conceptual Data Modelling: An Empirical Study of Expert and Novice Data Modellers. *Australasian Journal of Information Systems* 4 (2):63-73.

Shanks, G., D. L. Moody, J. Nuredini, D. Tobin and R. Weber. 2010. Representing Classes of Things and Properties in General in Conceptual Modelling: An Empirical Evaluation. *Journal of Database Management* 21 (2):1-25.

Siau, K. 1999. Information Modeling and Method Engineering: A Psychological Perspective. *Journal of Database Management* 10 (4):44-50.

van der Aalst, W. M. P. 2005. Pi Calculus versus Petri Nets: Let Us Eat Humble Pie Rather Than Further Inflate the Pi Hype. *BPTrends* 3 (5):1-11.

Wand, Y. and R. Weber. 2002. Research Commentary: Information Systems and Conceptual Modeling - A Research Agenda. *Information Systems Research* 13 (4):363-376.

Weber, R. 1987. Toward a Theory of Artifacts: A Paradigmatic Basis for Information Systems Research. *Journal of Information Systems* 1 (2):3-19.