GROUNDED THEORY AS FOUNDATIONS FOR METHODS IN APPLIED ONTOLOGY

John W Lamp
School of Information Systems,
Deakin University, Geelong, Victoria,
John.Lamp@deakin.edu.au

Simon K Milton
Department of Information Systems,
University of Melbourne, Melbourne, Victoria,
smilton@unimelb.edu.au

Abstract

Research into domain specific ontologies is difficult to treat empirically. This is because it is difficult to ground domain ontology while simultaneously being true to its guiding philosophy or theory. Further, ontology generation is often introspective and reflective or relies on experts for ontology generation. Even those relying on expert generation lack rigour and tend to be more ad-hoc. We ask how Grounded Theory can be used to generate domain specific ontologies where appropriate high level theory and suitable textual data sources are available. We are undertaking generation of a domain ontology for the discipline of information systems by applying the Grounded Theory method. Specifically we are using Roman Ingarden's theory of scientific works to seed a coding family and adapting the method to ask relevant questions when analysing rich textual data. We have found that a guiding ontological theory, such as Ingarden's, can be used to seed a coding family giving rise to a viable method for generating ontologies for research. This is significant because Grounded Theory may be one of the key methods for generating ontologies where substantial uniform quality text is available to the ontologist. We also present our partial analysis of information systems research.

Keywords: Ontology, Grounded Theory

Introduction

Ontology generation is often haphazard relying on the skill of the person generating the ontology. It is also difficult to remain true to a guiding theory or philosophy while grounding the ontology in what is happening in the world. Other approaches based on automatic textual analysis fail to capitalise on expert knowledge or awareness of the domain under study. Empirical approaches can also use broad methods that result in summarising the collective view of a broad population but cannot be applied in specific highly technical settings.

The veracity of a domain ontology is judged using principles of clarity in concepts and relationships between concepts. Further, the quality of the top-level categories into which the lower categories fit can be judged for sensibility. Also, one can trace the relationship between categories and the world that gives rise to them. In any of these cases it is still an open area for research.

We are seeking methods that are simultaneously grounded in appropriate data while being guided by high-level ontological theory or philosophy. Grounded Theory has been successfully applied in Information Systems to elicit concepts from rich textual data. Thus, we propose to explore the usefulness of Grounded Theory as a method in applied ontology for generating domain specific ontologies that are tied back to high-level theory. This would most directly apply where a high quality source of relevant textual data about specific domains can be found thus allowing qualitative research. Indeed, the authors of this paper are presently involved in a project, a significant aspect of which involves the analysis of papers reporting information systems research in academic journals using the Grounded Theory method. A broader description of this project and a discussion of the rationale for using Ingarden's frameworks can be found in Lamp & Milton (2003, 2004, 2007).

In pursuing this research we have sourced a coding family from an ontological theory that explains written scientific works. Following this we applied the Grounded Theory method using this coding family.

This paper begins by exploring the use of grounded theory in information systems before examining the sources of coding families. The method based on Grounded Theory is then described. Following this we describe a coding family based on Roman Ingarden's philosophy that covers the literary work of art, of which academic journal papers are a boundary case. This theory is the relevant ontological theory for the work. We then describe an initial application of the method, and discuss the preliminary outcomes from this exercise.

Grounded Theory Method and its Use in Information Systems

Grounded Theory method provides a logically consistent set of data collection and analysis procedures which can be used in textual analysis. The original aim of Grounded Theory method was to develop a research methodology which would systematically derive theories of human behaviour from empirical data. Grounded Theory method seeks to discover what is going on. Typically it is applied to texts obtained by interview, observation or other data collection methods. Explicit in the use of Grounded Theory method is data collection from participants who may have different views of the phenomena being studied, and which must be accommodated in the development of theory.

Since its original announcement it has become an accepted qualitative research technique (Urquhart, 2001). Its adoption has probably been assisted by its originators' assertions that it is "not bound by either discipline or data collection" (Glaser, 1992: 18). Grounded Theory method is concerned with the identification of categories or properties of categories as a major issue (Glaser, 1992: 4). While Grounded Theory method was described to have emerged from symbolic interactionism, Annells (1996) classified it as being ontologically based in critical realism. Glaser & Strauss (1967: 6) state "our position is not logical; it is phenomenological."

Reviews and examples of the use of Grounded Theory method in information systems research can be found in Calloway & Ariav (1991), Pidgeon et al (1991), Hughes and Howcroft (2000), Urquhart (2001), Fernandez (2004) and Lings & Lundell (2005). Hughes and Howcroft (2000) review a number of uses and adaptations of Grounded Theory method by information systems researchers. After considerable discussion of the implications of the ways in which information systems researchers have applied Grounded Theory method, Hughes and Howcroft argue against the rigid application of Grounded Theory method in practice. They also note that "[i]f the research community is to mature then it would be of far greater benefit to tell the story as it were, and this should include the researcher's perspective, actual use of the method, and a reflective evaluation." Hughes and Howcroft further assert "that the adoption and diffusion of the method should be welcomed since it represents its usefulness as a pragmatic tool for research."

Urquhart (2001) provides considerable detail of an instance of using Grounded Theory method to examine client – analyst interaction and behaviour, and also lists a number of IS researchers using Grounded Theory method. One instance of the use of Grounded Theory method (Orlikowski, 1993) received *MISQ*'s Best Paper Award in 1993. Fernandez (2004) also provides a background to Grounded Theory method, plus a report of the experience of using Grounded Theory method. He agrees with many of the points made by Urquhart (2001). Fernandez (2004) employed the software package ATLAS.ti to assist with his analysis and makes some cautionary points about the limitations and negative aspects of software assisted coding, in particular the potential for automatic coding (e.g. coding all occurrences of a word or phrase) to have a negative effect of obscuring the discovery from the researcher. This confirms the cautions given by Glaser about hiring coders and the use of automated systems which remove the analyst from close contact with their data (Glaser 1978: 58-59, 71).

At this point it should be noted that Grounded Theory method has come to include two divergent approaches. We now distinguish between those two approaches and the degree to which they suit the needs of our research project.

In 1992 Barney Glaser published *Emergence vs. Forcing: Basics of grounded theory analysis* (Glaser, 1992) in which he set out a comprehensive and vigorous attack on the contents of a 1990 book written by his original collaborator, Anselm Strauss, and Strauss' research partner, Juliet Corbin (Strauss & Corbin, 1990). Glaser considered their book "distorts and misconceives Grounded Theory method, while engaging in a gross neglect of 90% of its important ideas" (Glaser 1992: 2). In the years since this conflict, researchers have found both approaches useful as their different emphases make them more or less appropriate in specific research settings (Fernandez, 2004). The Glaserian approach is described as abstract conceptualisation, and the Straussian approach as full-description (Fernandez 2004).

This difference relates to a disagreement on the granularity of the unit of analysis. The Straussian approach emphasises word by word analysis (Strauss & Corbin, 1990: 81-84), while the Glaserian approach deals with

units of meaning at the line or sentence level (Glaser, 1978: 57). Strauss & Corbin give an example (1990: 82-83) of taking an hour to discuss what an individual meant by the word "once". Our analysis aims to discover what research is being reported on in journal papers. Our aim is to abstract, rather than commentate or interpret. We are assisted in this, as the purpose of journal papers is the transmission of cognitive knowledge, a concomitant of which is for the papers to be written using clear unambiguous language. In this case then, there should be little need for extensive analysis of individual words and their possible meanings. The Glaserian approach would seem to be more appropriate in this study.

Another significant difference in the two approaches for our purposes is the method of coding data. The Strauss & Corbin method requires that all data be coded against a single coding family – context, conditions, action/interactional strategies, intervening conditions and consequences. The Strauss & Corbin coding family is clearly inappropriate for ontological analysis. This coding family is a variation of only one of eighteen coding families proposed as a significant part of the book *Theoretical Sensitivity* by Glaser (1978: 72-82). Several of those coding families refer to ontological and mereological concepts (eg The Dimension Family, The Type Family, The Theoretical Family and The Conceptual Ordering Family – see Table 1). The list presented by Glaser is inclusive rather than exhaustive, and it is clearly intended that researchers using this method could derive their own coding families (Glaser, 1978: 73, 82) and this occurs in practice (Urquhart, 2001: 128). Glaser expanded on his original list of coding families in two later books (Glaser, 1998 & 2005).

Table 1. Selected examples of coding families (from Glaser, 1978: 75-79).

Coding Family	Categories
The Dimension Family	Dimensions, elements, division, piece of, properties of, facet, slice, sector, portion, segment, part, aspect, section
The Type Family	Type, form, kinds, styles, classes, genre
The Theoretical Family	Parsimony, scope, integration, density, conceptual level, relationship to data, relationship to other theory, clarity, fit, relevance, modifiability, utility, condensability, inductive-deductive balance and interfeeding, degree of, multivariate structure, use of theoretical codes, interpretive, explanatory and predictive power, and so forth
The Conceptual Ordering Family	Specification of concepts, and in developing properties of categories. An example is the ordering for motivation generally in the socialisation of people: Achievement orientation, institutional goal, organisational value, personal motivation.

We have seen that several of the coding families found by Glaser and others have heavy ontological undertones. Further, there are few rigorous methods for generating domain ontologies that tie back to high-level ontological theory. If ontological theory can be used to generate an appropriate coding family and then be used empirically in Grounded Theory then we have the makings of a rigorous qualitative method for ontology generation.

We now seek a method that is grounded in published research where a coding family is based on appropriate high-level philosophy. Thus, the following section describes a method, based on Grounded Theory, to analyse textual data sourced from published peer reviewed research. After that we turn to Roman Ingarden's ontology of the literary work of art, to source an appropriate coding family that we call 'The Reported Research Family'. We then present initial results of using Grounded Theory method with an ontologically grounded coding family as a tool for generating a domain specific ontology.

Using Grounded Theory method in domain ontology generation

There are two parts to using grounded theory in creating a domain ontology. Firstly we must decide on a certain stance, thus placing this approach into the broader context of Grounded Theory. Secondly, we must describe the coding process itself.

Stance

There are three areas needing attention when adapting Grounded Theory to new contexts. Firstly, one must decide what the researcher is looking for when analysing rich text. Secondly, a stance needs to be taken with respect to whether to enter the research with a coding family or seek emergence. Thirdly, an appropriate level of granularity needs to be selected (eg, word, phrase, sentence, or paragraph level). We address these below.

Firstly, traditionally in Grounded Theory, the researcher is looking for *Basic Social Processes* (BSPs) in rich qualitative data. Initially the researcher considers three questions (Glaser, 1978: 57):

- What is this data a study of?
- What category or property of a category, of what part of the emerging theory, does this incident relate?
- What is actually happening in the data?

In considering ontological analysis of research reported in journals, we are looking to find reports of research activity. Thus, we define the equivalent concept for BSP to be a *Reported Research Activity* (RRA). An RRA is central to the understanding of how a paper contributes to knowledge, as a BSP is to understanding a human activity. To paraphrase Glaser's three points above:

- What research activity is this scientific work reporting on?
- What category or property of a category, of what part of the emerging theory, does this research activity relate?
- What is actually being undertaken in the research activity?

These questions would be appropriate for any domain ontology grounded on published research papers. It remains to be seen if adaptation of questions would be required for other ontological studies.

Secondly, we intend using ontological theory to underpin a coding family with which we analyse rich text (the papers). In the example in this paper, we use an ontological framework from Roman Ingarden and we explain the choice of Ingarden's theory. We introduce Ingarden's theory and we describe the parts of the theory that form the basis of a coding family for the Grounded Theory method. Further categories dealing with matters relating to publication, other than what is contained within the papers themselves (time to publication, reviewing status, intellectual property status), which were suggested in Lamp (2002), may be added. The Grounded Theory coding technique united with Ingarden's ontological analysis would appear to meet our needs, but we do not claim that what we are undertaking is Grounded Theory method as it is not in accord with either the Glaserian or Straussian approaches. We appropriate the coding technique from Grounded Theory method and adapt it to ontological analysis of text.

Thirdly, as mentioned earlier, we intend coding at the sentence level or coarser granularity. We are analysing research reported in journal papers through a process that results in an abstract understanding of the domain of information systems. We are examining the top journals in the field and mostly expect the articles to be clearly written and to use sentence and paragraph structures in ways that clearly convey knowledge. We thus expect comparatively little recourse to analysing specific words for multiplicity of meaning.

It should be noted that while this is not fully in accord with either the Glaserian and Straussian approaches, it draws on aspects of both. The unit of analysis and aim of conceptualisation tend towards the Glaserian approach. The use of a largely predetermined, rather than fully emergent, coding family is a Straussian feature, but the coding family is not that prescribed by Strauss. For the purposes of this research, we are seeking a technique which can be applied to discovering concepts within scientific works, and which is philosophically compatible with Ingarden's framework. At this point it should be noted that while we are seeking to develop a methodology where all components share a consistent philosophical perspective, the content of the papers being analysed would not be limited to this perspective. The nature of the philosophical perspective used and reported on by researchers in the papers studied would be a value assigned to a category in our analysis.

The Coding Process: Foundations of our Technique

Grounded Theory method can be considered to consist of two phases; *substantive coding* and *theoretical coding*. Substantive coding itself consists of two sub-phases: *open coding* and *selective coding*.

In open coding the analyst aims to "generate an emergent set of categories and their properties which fit, work and are relevant for integrating into a theory" (Glaser, 1978: 56). Units of meaning are examined and coded against as many categories as may fit. New categories emerge, and new units of meaning fit existing categories.

Selective coding occurs when the analyst identifies core categories and limits his coding to "those variables that relate to the core variable in sufficiently significant ways to be used in a parsimonious theory" (Glaser, 1978: 61). In the context of this research program this corresponds to creating a parsimonious categorial tree where more specific terms are subsumed in more general categories. The researcher moves from open coding to selective coding when *theoretical saturation* is achieved. Theoretical saturation is said to be achieved when consideration of further data is not contributing further categories. Selective coding would be inappropriate for a study of only one year of one journal, where the aim is to generate a categorisation for the IS domain, as it is highly unlikely that theoretical saturation would be achieved.

Summarising, we have adapted Grounded Theory to the context of research. We have also flagged the use of a pre-determined coding family based on appropriate reference theory. Finally, we have described the coding process we intend following.

In the following section we describe the basis of the coding family used in this research: Roman Ingarden's examination of literary works of which scholarly articles are a special case making this theory appropriate for the study.

Towards a research publication coding family

Roman Ingarden developed and applied phenomenology to the examination of literary works. He developed a number of conceptual and methodological frameworks for ontological analysis of texts, which are documented in his books *The literary work of art* (1965) and *The cognition of the literary work of art* (1968). While Ingarden's primary focus was on mainstream literature, he also considered scientific works along with a number of other literary forms as borderline cases of the literary work of art.

The issue of applying Ingarden's framework to scientific works is significant because, while his work has been extensively applied to mainstream literature (e.g. Thomasson, 1996), there are no reports in the archival literature on developing his ontological analysis of scientific works into a technique which can then be applied.

The research results in the later sections report on an analysis of one volume of *Information Systems Research*. To illustrate Ingarden's framework based on scientific works in this theoretical section of the paper, we use Broadbent, Weill *et al* (1999) as our exemplar. This paper is from *MIS Quarterly*. While *MIS Quarterly* is often rated as the most significant information systems journal (eg Peffers & Ya, 2003; Katerattanakul *et al*, 2003a, 2003b; Bharati & Tarasewich, 2002; Mylonopoulos & Theoharakis, 2001; Walstrom & Hargrave, 2001), the choice of the particular paper is not significant, but was a paper conveniently at hand.

Ingarden asserts that works of fiction contain no genuine judgements. Instead they contain quasi-judgements. Works of fiction concern a portrayed world, in which assertions, or statements by portrayed persons, can only be considered within the context of the portrayed world (1968: 147-8). Contrastingly, scientific works are asserted by Ingarden to consist almost exclusively of genuine judgements (1968: 147), the most significant ontic items of which are:

- the states of affairs described;
- schematized aspects; and
- the represented objectivities.

In the following sections we examine these concepts in Ingarden's framework in some detail.

States of affairs

Ingarden asserts that the role of the scientific work in the transmission of cognitive knowledge requires that its context is that of states of affairs in the real world. Consequently, because there is a real world to which judgements refer, he calls the judgements in scientific works genuine judgements.

Ingarden states that genuine judgements are assertions that may be true or false, but they lay claim to truthfulness; eg a paper may report "The management style of company A was undemocratic" which is a result perceived as true by that researcher, and yet a second researcher may report a different result. Despite their essential contradiction, both statements are genuine judgements. Ideally, genuine judgements allude to means of confirmation which may be found in experience, or are contained in literary proofs based on reasoning and written in conceptual language. Failure to provide means of confirmation weakens the paper and reduces its functional value (Ingarden, 1968). Broadbent, Weill et al (1999) shows many examples of the first form of confirmation.

> "From March to September 1992, a small team further examined process approaches concurrently with a detailed study of the capabilities of the firm's current systems and infrastructure."

The above quote is a genuine judgement regarding the efforts of the business process reengineering team in the subject organisation, called "CostCo" in the paper, as collected by the authors of the paper; that is, reporting an event actually experienced by the person reporting to the authors of the paper, and therefore claiming authority based on reporting an actual contemporary experience.

> "Business process redesign (BPR) is a pervasive tool for transforming organizations (Grover et al. 1993) and ranked as one of the most important issues for information systems (IS) executives since the early 1990s (Brancheau *et al*. 1996; Index Group 1994; Watson *et al*. 1996)."

The above quote is a genuine judgement regarding views of IS researchers on BPR; that is, reporting a state of affairs reported in the IS literature, and therefore claiming authority based on previous accepted research.

Genuine judgements whose authority is based on literary proofs are relatively unusual in information systems research and are not found in Broadbent, Weill *et al* (1999). An example of this form of genuine judgement is the following definition of one sided separability from Smith (1998).

$$OS(x,y) \coloneqq x \leq y \wedge x \neq y \wedge \exists w(w \leq y \wedge \neg O(w,x) \wedge SD(w,x)) \wedge \neg \exists w(x \leq y \wedge O(w,x) \wedge SD(x,w))$$

Broadbent, Weill et al (1999) also contains questions such as in the following extract:

> "Important questions to consider include:
>
> - To what extent does the firm have at least the 10 core infrastructure services (see Table 1) together with the seven boundary-crossing services in place?
> - What is the reach in terms of who can be seamlessly connected?
> - What range of services are available: only the ability to access information or the capacity to perform complex business transactions across multiple systems?"

Such questions could be reworded as assertive statements, reporting evidence or argument supporting the affirmative or negative, and hence should also be considered genuine judgements (Ingarden, 1965: 329).

Schematized aspects

Objects represented in a literary work are derived purely intentional objects projected by units of meaning (Ingarden, 1965: 218). They are intentional because an author has written them with a purpose. For literary works of art it is to tell a story or generate a particular aesthetic effect. For scientific works it is the transmission of cognitive results (Ingarden, 1965: 330). In both cases they are derived, because we cannot enter the mind of the author. Finally, they are projected, because it is only through language (in this case written language) can we understand what is intended.

Consider the following extract also from Broadbent, Weill et al (1999):

> "CostCo has a robust network with numerous LANs in place at its head office, in large cities, and other major sites. 'About 2,000 PC users have whatever multihost connectivity is required for their business needs. We have moved from computer-centric to network-centric computing,' noted the CIO. Business units utilized these networks as a basis for a new distributorship, retail and electronic funds transfer, and point of sale (EFTPOS) systems."

This extract concerns the existence and nature of the network in place at CostCo – it describes one aspect of CostCo. The physical nature of the network is described along with the scope and purpose of the network and the attitude of management towards the use of the network. An outline of the functionality offered by the network in place is also described.

However, what is represented by this extract does not stop at the network, but extends further to linkages to provide EFTPOS, the retail activities of CostCo and other related activities and the general management policy framework within which the specific network policy lies, even though none of this is directly given to us. This is also typical of scientific and other literary works because there is seldom enough room to completely describe a state of affairs. Equally, this description does not delve into the details of the precise networking protocols, hardware, operational requirements and other minutiae of CostCo's network. If this depth of analysis was provided, then the extract would no longer be talking about the network, but of the components of the network and how it was managed.

For these reasons, literary works necessarily consist of incomplete descriptions, termed schematized aspects, which contain fulfilled (explicitly described) components and unfulfilled components, which while not explicitly described, may not be indeterminate. The reader may fill these out from aspects held in readiness from previous experiences. Prompting the most appropriate aspect is influenced by the word choice and represented objectivities selected by the author. For example, by using two words identical in meaning, but different in word sound, the reader may be influenced towards different aspects (Ingarden, 1965: 277). Substituting "many" for "numerous" in the preceding quote does not alter its meaning. Both convey a sense of a large quantity of items, but while many is completely unqualified, numerous contains an implication of a collection that can be numbered, an act which then intrinsically imposes order. This difference may influence the reader's choice of aspects between perceiving CostCo's LANs as an unordered collection or as an ordered and therefore managed collection.

Because these aspects are based in perception, and aspects of the same object which are experienced by different individuals must differ in various respects, it is not possible for the reader to actualise with complete accuracy the same aspects intended by the author (Ingarden, 1965: 265). The degree of this type of perceptual error in a scientific work is reduced, as schematized aspects are intended only for assistance in the transmission of cognitive knowledge. The use of decorative or evocative aspects is unnecessary, and may hinder the essential aim of a scientific work – accurately transferring knowledge as intended by the author (Ingarden, 1965: 330).

Represented objectivities

A literary work of art describes people, animals, lands, houses and other items. This represented world is not the real world – the represented objects within it may not exist in the real world or may behave differently to such objects within the real world. As a reader reads a passage of words and phrases (meaning units) containing a represented objectivity, he or she relates directly to the state of affairs that the represented objectivity is helping to clarify. Consequently, a particular represented objectivity within a scientific literary work causes us to direct ourselves to corresponding states of affairs. Because we are dealing with a scientific work rather than a literary work of art, this directional ray passes through the content of these represented objectivities so that they refer to objectively existing states of affairs, or to objects contained within them rather than to some fictional creation (Ingarden, 1965: 329; See Figure 1). It is through this directional meaning ray that the represented objectivities claim to determine objects in the real world as they are in themselves and thereby claim to be genuine judgements (Ingarden, 1968: 148).

Figure 1. The directional meaning ray

In a scientific work, clarity in writing directly affects the transparency of represented objectivities. Where readers have difficulty in relating to the state of affairs beyond the text then the represented objectivities are not 'clear'. Ideally the represented objectivities are transparent, that is the way that they inform the reader regarding the particular state of affairs under discussion requires little conscious interpretation. In the extract above, the IT infrastructure (the state of affairs being discussed) is correlated with the represented objectivities "computer-centric" and "network-centric". For readers of *MIS Quarterly* these would be readily understood concepts and hence transparent. It should be noted that if that was not the case, and the authors went on to define the concepts, then the concepts would become states of affairs for which other, hopefully more transparent, represented objectivities could be found to describe them. In the absence of such explanation from the authors, the reader re-reads the sentences, concentrating on the word meanings and syntactic interconnections until the represented objectivities become clear and unequivocal (Ingarden, 1968: 149-50).

Summarising, any scientific work will contain genuine judgements, states of affairs, schematized aspects, and represented objectivities. Effective subject indexing in any academic field requires the terms in these ontic categories to be identified and the relations between them defined. The terms in these categories are meaningful to some group(s) of people and form the basis of the subject indexing. Clearly, there are potentially scores of specific terms that fit within each of the categories. Further, the relative importance of each term is not evident *a priori*.

The Research Publication Family

Thus, a coding family based on Roman Ingarden's philosophy should consist of the three categories described above: states of affairs, schematized aspects, and represented objectivities. Additional to these are dimensions of the publication which interest us: time of publication, review process, etc … Using this coding family, and the method of coding published research using a Grounded Theory approach, we seek terms that (a) are meaningful to a group of people, (b) exhibit cognitive economy (Rosch, 1978), and (c) are discovered through a process that is repeatable and rigorous. The principle of cognitive economy is that categorisation should provide a great deal of information about the item categorised with minimal resources expended.

In the following section we outline a textual analysis of published information systems research in order to generate a categorisation of keywords applicable to the domain of Information Systems.

Preliminary results: partial analysis of *Information Systems Research*

In this section we cover some preliminary results of analysing published research in information systems. It is important to note we have not completed our analysis and that this paper is largely about the process of adapting the Grounded Theory approach to generating domain ontologies.

An initial application of the method has been made to Volume 16 of *Information Systems Research*. An outline of these results is given here using passages from an article in that volume, Majchrzak *et al* (2005) just for illustrative purposes. Nvivo was used as a tool to track the coding and documents.

When applying our method, specific attention was paid to passages containing phrases such as "we derive," "we specifically focus on," "we include" and "we do not try to" to identify RRAs. Statements with phrases like these often characterise the represented objectivities or states of affairs contained in the RRAs being studied. Consider the following paragraph:

> "We derive a theoretical model from Te'eni's (2001) cognitive-affective model of communication to elaborate how information technology (IT) can support an individual's communication of context to develop collaboration know-how."

The paragraph was coded as using the represented objectivity "model building" concerning the states of affairs "collaboration," "communication," "know how" and "IT support". The represented objectivity, the process of model building understood by researchers, is providing information on how the authors approached their investigation of the states of affairs, collaboration, communication, use of know how, and levels of IT support found in the organisations under study.

Similarly, the passage:

> "In sum, we argue that the opportunities for misunderstanding when performing nonroutine tasks are so great that collaboration know-how development will benefit from any IT support for contextualization even if the support is partial. When individuals perform routine tasks, however, partial IT support will lead to reduced collaboration know-how development because individuals not only must perform the task, but must also expend cognitive resources at the same time to resolve and reconcile the implications of the missing context."

was coded as a represented objectivity "a priori hypotheses," as it was asserting a view held by the authors before their investigation took place, therefore describing the way they viewed their investigation.

The method described in this paper identified forty-five represented objectivities and forty-three states of affairs, a total of eighty-eight terms. Eleven represented objectivities and thirteen states of affairs were used in more than one article. It is also worth noting that, examining the terms we have identified, the ones recurring the most, concern the research process or artefacts (eg. quantitative results and quantitative analysis). This is not surprising but worth noting as a dimension often overlooked – that of the research method, tools, techniques etc.

Table 2. Represented Objectivities in *ISR* volume 16

Term	Documents
a priori hypotheses	16
analytical framework	1
attitude theories	1
conceptual framework	1
configuration theory	1
consumer-piracy behaviour model	1
contextualization	1
contract-theoretic framework	1
cost	4
design science	1
developing country	1
diversity of participants	1
econometrics	1
economic analysis	3
elaboration likelihood model	1
ethnography	1
experimentation	6
field study	5
game theory	3
grounded theory	1
habit-automaticity perspective	1
induction	1
instant activation perspective	1

Term	Documents
interviewing	2
methodology development	1
model building	12
multi-country	2
observation	1
ontology	1
optimal control theory	1
perceptions	1
practice-theoretical approach	1
psychological contract violation	1
qualitative data collection	1
quantitative analysis	20
Reflection-in-action	1
resource based theory	1
set-theoretical notation	1
social impact theory	1
structural equation modeling	1
survey	7
task-technology theory	1
technology acceptance	1
technology-organisation-environment	1
trust	1

Table 3. States of Affairs in *ISR* volume 16

Term	Documents
accounting firms	1
accounting information systems	1
algorithms	2
automatic use	1
b2b ecommerce	1
buyer-seller relationships	1
collaboration	3
communication	1
consumers	1
control	1
data mining	1
data structures	1
descriptive results	9
e-business	5
expert systems	1
groups and teams	1
human resource practices	1
implications for practice	11
information processing	1
information quality	3
information sharing	1
IT support	1

Term	Documents
know how	1
market segmentation	1
online auctions	3
online brokerage	1
power	1
pricing strategies	2
privacy	1
project management	1
quantitative results	15
reputation mechanisms	1
security	3
social loafing	1
software engineering	1
system development	1
system quality	1
theory building	4
user satisfaction	3
users	5
virtual teams	1
web personalisation	1
website promotion	1

Relations between terms

A key aspect of developing an ontology is identifying the relationships between the terms. The 25 articles examined to date constitutes 12% of the total data collection. The question of the amount of work needed to approach theoretical saturation is presently indeterminate. Nevertheless, it is likely that much of theoretical coding will involve grounding relational terms. An important part of this will concern the types of relations that are evident when describing research. These are not restricted to general ontological relations such as part-whole and subsumption (is-a) relations. Indeed a well grounded ontology will comprise well defined primitive or fundamental relations applicable in the domain under study as well as the terms identified.

Tables 2 and 3 show the terms coded as states of affairs or represented objectivities identified in Volume 16 of *ISR*. As has been mentioned before this analysis constitutes only a preliminary study, however there are some interesting observations which can be made at this stage. There is a rich selection of theoretical approaches in Table 2, none of which appear in Table 3. At this point it would appear that *ISR* focuses on applying theory to analysis of information systems, rather than review articles where the theoretical approaches are themselves the primary subject of study, and would thus be coded as states of affairs.

Within each table, it is possible to intuit further categories into which the terms in the tables could be grouped, to extend categories or document relations. In Table 2, grouping based on theoretical approaches, as mentioned previously, is one potential grouping. The facilities provided by Nvivo allow for more than intuition. Reports listing all text coded to particular terms can be quickly generated by Nvivo. These reports can be used to compare the actual passages associated with terms to provide a basis for grouping or identifying relations that is more substantial than intuition. Memos can be raised to document the comparison and provide a means for evaluating the basis for that analysis. This facility will be useful when updating the analysis, either with more journals or material covering other publication years, to guard against inaccurate inclusion of new material in existing categories.

This approach should provide a better documented and more transparent approach to developing domain ontologies that can be defended on the basis of a clear and documented process, rather than by an appeal to the mystique of expert opinion or introspection.

Some observations on author chosen keywords

Although the main focus of this research is the generation of a domain ontology using rich grounded methods, a by-product is the generation of keywords that are comparable with those researchers routinely assign to articles. Further, the first journal we chose to analyse was *ISR* – a journal that has a canonical list of keywords recommended for authors to use. Thus, we took the opportunity to compare the keywords emerging from applying this method and those suggested by the author, and by the journal's editorial committee. The authors of Majchrzak et al (2005) chose the keywords "knowledge management, collaboration, virtual teams, distributed teams, knowledge sharing, group support systems" to describe the article. Our analysis of that article identified a larger set of terms than the keywords chosen by the authors.

The following author assigned keywords were not identified by our method: knowledge management, distributed teams, knowledge sharing, group support systems. "Knowledge management" is an *ISR* prescribed keyword but only appears once in the body of the article as "knowledge management systems," as a type of system which the conclusions in the paper could be extended to cover. "Distributed teams" is used as a synonym for "virtual teams." It is not an *ISR* prescribed keyword, while "virtual teams" is a prescribed keyword. "Knowledge sharing" appears when discussing a form of survey instrument and when mentioning types of IT support applications. "Group support systems" does not appear in the body of the article. Neither "knowledge sharing" nor "group support systems" are *ISR* prescribed keywords.

One relevant incidental observation which emerges starkly from this initial analysis is the tiny degree of adherence to the keyword scheme prescribed by *ISR*. Nine out of 109 keywords actually used in volume 16 were in the *ISR* list of keywords. If this degree of adherence is a norm within the journal, it means a researcher attempting to search an archive of articles using prescribed keywords, would have an extremely low recall score and one would have to question the usefulness of such an archive. Of the 109 author chosen keywords, only two keywords were used more than once. Again, the recall score of a search under such a regime would be very low. Neither of the two available approaches is effective. This observation reinforces the need for a more rigorous technique in this area.

As has been noted previously, only two author chosen keywords were used in more than one article. The method described in this paper identified forty-five represented objectivities and forty-three states of affairs, a total of

eighty-eight terms. Eleven represented objectivities and thirteen states of affairs were used in more than one article.

Fourteen author chosen keywords were also coded as states of affairs. Eight author chosen keywords were also coded as represented objectivities. In some cases the keywords were *equivalent* rather than *identical*, because there were minor variations. For example "web personalisation" was our term, but "web personalization" was used by the authors. Failing to cater for regional spelling variations is another well known issue with author assigned keywords. We assigned "technology acceptance" as a term, but the authors assigned "technology acceptance model." For the most part, terms emerging from the analysis method tended to be more general than those chosen by authors.

It is also worth noting that, examining the terms we have identified, the ones recurring the most, concern the research process or artefacts (eg. quantitative results and quantitative analysis). This is not surprising but worth noting as a dimension most often overlooked – that of the research method, tools, techniques etc.

It could be argued that these results are already more useful as categories as they appear in multiple articles, have a degree of overlap with author assigned keywords, but extend beyond them and therefore will be more useful as search terms. However, it should not be assumed that repeated use is always a virtue. For example, one term, "quantitative analysis," is coded against all but one article. As a search term it would not be useful.

Conclusions

In this paper we asked how Grounded Theory Method could be used in ontology generation for a domain where high-quality rich textual data is available and where domain expertise is critical. One important step in the process involved deriving a coding family from a guiding philosophy for the domain that shares a heritage with the Grounded Theory approach. We have described the method and related it to recognised schools of Grounded Theory and presented preliminary analysis. That analysis is not the key focus of this paper. Nevertheless, the product of this analysis would constitute the starting point for a categorial scheme for that domain. The data collection and analysis techniques provided by Grounded Theory method give an approach to the analysis of textual material in which a guiding philosophy can direct, but not constrain, the analyst. Grounded Theory method and Roman Ingarden's literary ontology have similar philosophical heritages. They are both from the realist tradition and share assumptions about the world and the ways in which it may be understood. Importantly for this research, they both explicitly provide for the accommodation of differences in perception of states of affairs by individuals. The proposed tool therefore conforms to the definition of a method as a "coherent and systematic approach, based on a particular philosophy" (Fitzgerald et al, 2002: 5, *cit* Lings & Lundell, 2005). As such, it has the potential to move the development of domain ontologies into the realm of defensible analysis, rather than oracular pronouncements by experts, however well meaning.

We have found that Roman Ingarden's ontological analysis of the literary work of art, the theory upon which the research is based, readily formed the basis for a coding family. The family, with a set of guiding questions for the researcher applying the method has yielded terms about the published works. It will be fascinating to see how theoretical coding proceeds and helps us to move beyond the terms found thus far to more abstract terms and to primitive types of relations between terms.

The application of Grounded Theory method to ontology generation may be novel, but Grounded Theory is approaching its 40th anniversary and as it appears in information systems literature for the past 15 years, it should not be unfamiliar to information systems researchers. This last point is seen as significant as we see this tool as having applications beyond the immediate project for which we are developing it.

The use of this tool by a number of researchers for the purpose of this project would provide a consistent analysis which would reflect the intentionality of the researchers using it. This could begin to capture the diversity of perspectives which exist in information systems.

The tool may also be applicable for eResearch in other interdisciplinary subject domains, such as health informatics, and to other forms of literature intended for transmission of cognitive knowledge in the same way as scientific works – systems and user manuals, requirements specifications etc. Journal editorial boards could use it from time to time as a check on whether what was being published actually reflected their stated aims and scope.

What remains is to be seen, is if the approach is practically feasible. Initial results, mentioned in this paper, appear promising. Hence, the next step is to apply the approach to a significant sample of papers from the top five information systems journals. We then will be able to comment on the practicality of the approach.

References

Annells, M. (1996) "Grounded theory method: Philosophical perspectives, paradigm of inquiry, and postmodernism," *Qualitative Health Research* 6(3)

Betteridge, H. (1978) *Cassell's German-English English-German Dictionary* London, Cassell & Co Ltd

Bharati, P., & Tarasewich, P. (2002). "Global Perceptions of Journals Publishing E-Commerce Research." *Communications of the ACM* 45(5): 21-26.

Broadbent, M., Weill, P., & Clair, D. S. (1999). "The implications of information technology infrastructure for business process redesign." *MIS Quarterly* 23(2): 159-182.

Calloway, L. J., & Ariav, G. (1991). "Developing and using a qualitative methodology to study relationships among designers and tools". In H.-E. Nissen, H. K. Klein & R. Hirschheim (Eds.), *Information Systems Research: Contemporary Approaches and Emergent Traditions* (pp. 175-193). Amsterdam: North Holland.

Fernandez, W. D. (2004) "The Grounded Theory Method and Case Study Data in IS Research: Issues and Design" *Proceedings of Information Systems Foundations Workshop: Constructing and Criticising* Australian National University.

Fitzgerald, B., Russo, N.L. & Stolterman, E. (2002), *Information Systems Development: Methods in Action*, McGraw-Hill, London.

Glaser, B. (1978) *Theoretical Sensitivity* Mill Valley CA, Sociology Press.

Glaser, B. (1992) *Emergence vs Forcing: Basics of Grounded Theory Analysis* Mill Valley CA, Sociology Press.

Glaser, B. (1998) *Doing Grounded Theory: issues and discussions* Mill Valley CA, Sociology Press.

Glaser, B. (2005) *The Grounded Theory Perspective III: Theoretical Coding* Mill Valley CA, Sociology Press.

Glaser, B., & Strauss A. (1967) *The Discovery of Grounded Theory* New York, Aldine de Gruyter

Hughes, J., & Howcroft D. (2000). "Grounded Theory: never knowingly understood." *New Review of Information Systems Research* 1(1): 181-199.

Ingarden, R. (1965) *Das literarische Kunstwerk* (translated GG Grabowicz *The Literary Work of Art: an investigation on the borderlines of ontology, logic and theory of literature* Evanston, Northwestern University Press, 1973) Max Niemeyer Verlag, Tübingen

Ingarden, R. (1968) *Vom Erkennen des literarischen Kunstwerks* (translated RA Crowley & KR Olsen *The Cognition of the Literary Work of Art* Evanston, Northwestern University Press, 1973) Max Niemeyer Verlag, Tübingen

Katerattanakul, P., Han, B., & Hong, S. (2003a). "Objective quality ranking of computing journals." *Communications of the ACM* 46(10): 111-114.

Katerattanakul, P., Razi, M., & Kam, H. (2003b). "IS journal rankings versus citation analysis: consistency and concerns". *Ninth Americas Conference on Information Systems*, Tampa, Florida.

Lamp, J.W., & Milton S.K. (2003) "An Exploratory Study of Information Systems Subject Indexing" *Fourteenth Australasian Conference on Information Systems*, Perth Australia, Edith Cowan University.

Lamp, J.W., & Milton S.K. (2004) "The Reality of Information Systems Research" *2nd Biennial ANU Workshop on Information Systems Foundations*, Canberra Australia, Australian National University.

Lamp, J. W. & Milton, S. K. (2007) "Indexing Research: An Approach to Grounding Ingarden's Ontological Framework", in Hart, D. & Gregor, S. (eds) *Information Systems Foundations: Theory, Representation and Reality*, ANU E-Press

Lamp, J.W., (2002) *IS Journal Categorisation* [Online] Available: http://lamp.infosys.deakin.edu.au/journ_cat/ Accessed: 18/05/2004.

Lings, B., & Lundell, B. (2005). "On the adaptation of Grounded Theory procedures: insights from the evolution of the 2G method". *Information Technology & People*, 18(3), 196-211.

Mylonopoulos, N. A., & Theoharakis V. (2001). "Global Perceptions of IS Journals: Where is the Best IS Research Published?" *Communications of the ACM* 44(9): 29-33.

Orlikowski, W.J. (1993). "CASE tools are organisational change: Investigating incremental and radical changes in systems development" *MIS Quarterly*, 17, 309-340.

Peffers, K., & Ya T. (2003). "Identifying and evaluating the universe of outlets for information systems research: Ranking the journals." *J Information Technology Theory and Application* 5(1): 63-84.

Pidgeon, N. F., Turner, B. A., & Blockley, D. I. (1991). "The use of Grounded Theory for conceptual analysis in knowledge elicitation". *Int. J. Man-Machine Studies*, 35, 151-173.

Rosch, E. (1978) "Principles of Categorization" in Rosh E and Lloyd B [eds] *Cognition and Categorization* Lawrence Erlbaum Associates, Hillsdale p27-48

Smith, B. (1998). "The Basic Tools of Formal Ontology". *Formal Ontology in Information Systems*. N. Guarino. Amsterdam, IOS Press: 19-28.

Strauss A., & Corbin J. (1990) *Basics of qualitative research*, Sage, California

Thomasson, A. L. (1996). "Fiction and Intentionality." *Philosophy and Phenomenological Research* 56(2): 277-298.

Urquhart, C. (2001). “An encounter with grounded theory: tackling the practical and philosophical issues”. *Qualitative research in IS: issues and trends*. E. M. Trauth. Hershey, Idea Group: 104-140.

Walstrom, K.A., & Hargrave B.C. (2001) “Forums For Information Systems Scholars: III” *Information and Management* 39 117-124