

Ontology Building as Practical Work: Lessons from CSCW

Dave Randall¹, Wes Sharrock², Yuwei Lin³, Rob Procter³ and John Rooksby⁴

¹Department of Sociology, Manchester Metropolitan University

²School of Social Sciences, University of Manchester

³National Centre for e-Social Science, University of Manchester

⁴Department of Computing, Lancaster University

Email address of corresponding author: d.randall@mmu.ac.uk

Abstract. Ontology building can be thought of sociologically: the work undertaken and the problems and difficulties entailed can be understood in terms of the practices of knowledge workers and the practical nature of ‘sorting things out’. It appears that many problems in the work of ontology building resemble problems in software engineering, particularly engineering cooperative systems. We discuss research in the field of Computer Supported Cooperative Work that focuses on classification and which throws light on ontology building. We then introduce some data from our own early ethnographic studies of ontology building.

Introduction

The development of ontology-based computer systems for various kinds of knowledge work continues apace. Hitherto, it has focused in the main on the capture and re-use of scientific knowledge, though expanding into areas implicating procedural work, and the work of professionals (medicine and knowledge management being obvious examples). These developments come with a recognition that increased complexity – located in greater heterogeneity of professional knowledge and skill – creates a different order of problem. Dealing with procedural matters raises the spectre of ‘exceptions’; dealing with professional endeavour with that of an acceptable allocation of function, and dealing with heterogeneous environments with that of the ‘knowledge gap’ between one community and another. In the ontology-building community this is often expressed in terms of the size and scope of ontologies themselves, and the concomitant issue of modularization. The point we will make is that they can equally well be thought of *sociologically* – to do with the working assumptions knowledge workers make; the practical nature of their enquiries, and the assumptions they carry when they make them. This is sometimes glossed as the problem of ‘situatedness’ and stimulates recent attempts to incorporate this sociological notion into ontology-related work (see Pike and Gahegan, 2007). If this provides some purchase, then there is considerable mileage in considering what sociologists can contribute to support of ontology-building.

Research in the field of Computer Supported Cooperative Work (CSCW) has addressed problems that very much resemble the problems faced by those engaged in the building of ontologies. In particular, we note a resemblance between ontology building and other classification work. Rather than re-invent the wheel, it seems prudent to address this prior work. In this paper we discuss recent work by Martin, O’Neill, Randall and Rouncefield

(2007) and Bowker and Star (1999). We then introduce some early data from our ongoing studies of ontology building.

Ethnography and Classification

As social scientists working over a period of time in the field CSCW, we are interested in how the kind of ethnographic work we are typically engaged in can be related to ontology building. This is predicated on prior experience of the social sciences in the system design process, and notably the contribution that ethnography has made to it (Anderson, 1994; Hughes et al, 1993) The original complaint upon which CSCW was founded concerned the inadequacies of mechanistic, top-down models in the system design process revealed by a new order of complexity in patterns of use – specifically, when user communities worked on networked machinery, and where patterns of input and output were distributed across work environments in ways which were not well-known. Equally, it was argued that designers of computer systems, characteristically, were apt to be very optimistic about how simply and quickly formal schemes and computational devices could be deployed to service practical tasks. That is, the sociological ‘turn’ indicated that designers under-estimated the practical difficulties involved in fitting simple schematics as models for variable, flexible orders of procedures (because they make the design work manageable) and, of course, prioritise articulating the design over understanding the use to which the design will eventually be put.

It is our contention that, as ontology building moves out into the areas described above, so the problems that have to be dealt with look increasingly like those experienced in CSCW. One of the solutions proffered has been the use of ethnographic methods to uncover detail about the business of work in this context. If so, then we should consider how this might work. Two important contributions to overcoming some of these difficulties have been to:

1. make use of ethnographers for observing and understanding practical activities as a means of providing designers with a familiarity with areas of application; and
2. engage users in the design process, enabling them to contribute their hands-on understanding of how work is actually done and also to provide feedback on how well the developing design meets the requirements that will, in use, be made of it.

Ethnographies of work for CSCW subsequently have been conducted in a wide range of domains, including air traffic control, emergency services work, transport coordination, knowledge management regimes (see, e.g., Groth and Bowers, 2001; Petersson et al, 2002). However, ethnographies tended for a time to be local, and small-scale. In recent years, they have turned to more complex domains, and to the knowledge work associated with them (e.g., Harper et al, 2000). Here the emphasis has been on an alternative conception of knowledge, one which is quite distinct from the information-theoretic one embodied in much ontology-based work. The dominant metaphor in the former has become that of ‘expertise sharing’ (see Ackermann et al, 2003) and research has emphasised the ordinary practical ways in which knowledge or expertise is (or is not) shared across knowledge communities.

We do not set out to critique assumptions about knowledge here, nor to privilege one set of beliefs about the nature of knowledge work above another. Our purpose is to identify how this problem of knowledge work and its behavioural dimensions resonates with some of the issues we have mentioned above. On the one hand, ontologies are a formal method of encapsulating knowledge and are intended to provide for the flexible re-use of data, and for rapid inferencing concerning data held. On the other, it is clearly the case that some ordinary, practical considerations go into the use and exchange of knowledge, and that some of that knowledge is not easily captured. In turn, it is a reasonable bet that an understanding of the

latter will have some impact on the procedures for capturing the former, and the design of tools to support it.

Of course, and in principle, ontologies can come in any size, from the one-size-fits-all approach which we would associate with a realist paradigm, to the modularized, limited scope approach we might associate with a more pragmatist paradigm. It is the latter we can be more readily associated with here, but we have nothing to say in this paper about the philosophical underpinnings of disputes between realists and pragmatists in the ontology community.

New lamps for Old: The Allocation of Function Problem

To illustrate the problems that may well turn out to be relevant for the deployment of ontologies in complex organizational environments, we explore some recent work on classification schema as they are used in call-centre work (see Martin et al., 2007). Although nothing that looks like a fully-fledged ontology is in place in this work, there are embedded classification schema. In this paper, the authors, borrowing from Bowker and Star (1999), suggest that three dimensions are particularly important when looking at the degree to which the deployed classification systems turn out to be useful or not. As Martin et al. put it:

“[we are] ... looking at the design relevance of classification through examining its actual everyday, operational nature as it features in call centre work. We do so because it became increasingly evident to us, when looking over the material we were examining, that the standardization processes we were looking at were almost always classification processes – operators in the customer-facing work we were interested in were expected to see their encounters with customers not only in terms of a standardized procedure but also as ‘types’ of encounter, generated by schema embedded in machinery.” (ibid)

In effect, a set of categories are fixed in the system that do not comprehensively fit the range of materials or circumstances that will have to be fed into the system, meaning that the operators must make available classifications do whatever work it is that needs to be done – the system does not fit the work, the work is made to fit the system. Hence, Martin et al, use the 'wide-ranging archaeology' of classification which Bowker and Star discuss, and issues that they raise in association with it. Hence:

“Imposed standards will produce work-arounds. Because imposed standards cannot account for every local contingency, users will tailor standardized forms, information systems, schedules and so forth to fit their needs.... When designing tools for distributed, organizational decision and policymaking, a detailed catalogue and analysis of such responses could become part of the designers’ tool kit; incorporated in the system, it could point out styles of work arounds at the level of coding.” (159).

Bowker and Star, then, are concerned with a ‘gap’ they perceive between the ‘formal’ and the ‘informal’, and the ways in which we might better understand the relationship between the two. They raise ‘challenges’ for classification schema in terms of comparability, visibility and control. Decisions are made in the design of any system about how the work will be distributed between the system and its users, but these are seldom made as deliberated, thought through decisions, but are more usually default consequences of quite other kinds of design decisions, and it is only when the system is brought into serious use that designers and users find out how that divisions of labour works out. One of the big tasks in ontology development is, as so often, that of inputting the materials that the classifications of the ontology will work on, where the annotation of research reports, data bases and the like will be the means of coordinating the materials with the search device, and where one possible solution to the division of work is to organise it so the scientists producing the materials can and will do the appropriate annotations. One of the root considerations of usability is not merely having something that can be used, but finding ways of ensuring that it will be used.

Comparability

Comparability refers to a ‘regularity in semantics and objects’ (1999: 231) and thus pertains almost by definition to ontologies. What is important in this context, however, is the degree to which this stability is, in practice, obtainable. Most ontology-design hitherto has been aimed at relatively homogeneous communities, where underlying concepts (if not terminology) stand a good chance of being commonly held. Ontologies that have to serve more heterogeneous situations and purposes may turn out to be serving one user group more successfully than another – a problem that has been well-attested in the field of medical informatics. Problems of this kind will be compounded as and when ontologies are deployed across organizational boundaries. These might be used for managerial as well as professional purposes, or might be deployed in contexts where parties to the work have no knowledge of the categories that underpin an ontology (as, for instance, with customer or client-facing work). The potential mismatch between the ‘requisite variety’ of terms in an ontology, and the actual use of terms is shown by Martin et al in the context of a help desk. They argue:

“Organizationally, the system was intended as a repository of problem types and solutions to improve efficiency; to aid continuity of service through customer record, and to measure work (by recording call times). This data could then be used to potentially re-arrange work by distributing personnel or assigning personnel to dealing with subsets of problems. However, some mismatches appeared between the imagined purposes of the system and the actual mechanics of classifying... [for instance] .. a large amount of calls were classified as ‘other’ or ‘miscellaneous’”.

And go on to show how:

“... many of the 100 or so categories of problem were rarely used. The types of problem listed under a category could vary quite greatly. One ‘password problem’ could easily be different from another, while ‘general problems’ showed massive variance. Within the flow of work the operators viewed problem classification as an inexact activity that was carried out under time pressure rather than an accurate portrayal of the work. The written commentaries in associated fields in fact often provided a clearer memory of the work undertaken than the classification scheme itself, offering the possibility to facilitate continuity of service through an audit trail of customer-organization interaction. They provided a way into a problem and were commonly backed up with the recollections of staff.”

Desk operators here utilized the system as a resource for standardizing practice, but actually managed the business of consistency through observing and talking with one another, supplementing ‘the records’ with occasioned collaboration. The records did not speak for themselves but were made sensible through cooperative techniques. Similar conclusions have been reached in other work which shows how distinct problems of classification occur in customer-facing work (see Harper et al., 2000) It is one thing for a professional knowledge worker to understand the uses to which an ontology can be put; it is quite another for that person to interact with ‘outsiders’ as they are doing so, and this relates to the notion of ‘visibility’.

Visibility

The second of the challenges to classification recognised by Bowker and Star is visibility. ‘Invisible’ areas of work are ‘by definition unclassifiable except as the residual category: “other”’ (1999: 231) ‘Invisible’ work here refers to those informal practices which do not, in themselves, constitute part of the ontology, but which may nonetheless be critical to how an ontology is constructed. For instance, it would be central to questions as, firstly, how explicit and complete should the categories in any classification scheme be, and hence to decisions about appropriate modularity for ontologies. For Bowker and Star there is an inevitable trade-off between comparability and visibility, insofar as comparability allows for use across a variety of settings, but risks an increasing degree of inappropriateness for each local setting

where it is used. They make the point that the more comparable (aimed at use across a number of settings) embedded classification schemes are, the less visible the work that goes into maintaining them will be and the less they will fit 'local' arrangements (see, e.g., Anderson et al, 2007). This, we think, is critical to decisions about the scope of ontologies, and, more importantly, raises the question of where our knowledge about these different settings – knowledge that would enable us to make appropriate decisions – is going to come from. Equally, and as pointed out by Martin et al., 'translating and mediating' work is frequently invisible, but necessary. In other words, in complex organizational environments, and dare we suggest even in professional knowledge work, we cannot presume that all users are equally adept at using ontologies. Put simply, what is being done with a classification system sometimes has to be explained to other people.

This, of course, implicates the interface – certainly at the GUI (graphical user interface) level and possibly also at the API (application program interface) level as well. It is one thing for the knowledge worker to understand what is in front of him/her in OWL or in an editor such as Protégé-OWL or Swoop, but quite another for he or she to use it to explain to others what work needs to be done. Whilst it is becoming common to visualise ontologies in various ways, particularly as a graph, we would suggest this problem can be addressed far more extensively. The work of Buckingham-Shum and his collaborators on rationale and argumentation has already served to point e-Science in this direction. In their work, attention has been paid to integrated and interoperable technologies for presenting and manipulating data but also to "the art and craft ... to know how to use the tools well enough that they are constructively disruptive, delivering immediate value to those using them, as well as supporting longer term memory" (Buckingham Shum et al., 2006, p129). If ontologies are to be deployed in heterogeneous domains where interfaces need to be shared amongst people with varying degrees of expertise then a great deal of thought needs to go into how classifications are represented.

Control

Regarding control, Bowker and Star argue that,

"... any prescription contains some amount of control to be exercised by the user, be it as small as in the most Taylorist factory or prison or as large as the most privileged artists' retreat the managerial trick is to measure the degree of control required to get the job done well, for most people, most of the time."

Again, and especially where ontologies are to be extended into procedural requirements, this would seem to be critically important, because users of any given ontology will themselves make decisions about its use in the light of their need to get the job done well. As Martin et al. suggest, based on their data on call-centre work- problems often occur when organizations attempt to switch the expertise from the operator to the system. They say:

"Control is central to the design of classification-based systems, in our view, because it is central to the problem of the residual. Skilful decision-making work might turn out to be a means by which the residual can be avoided or alternatively residual categories might create problems for that work. In other words, work can be, and may have to be, done in order to translate what would remain residual into one category or another ... Rehearsals of rules and their application, or as they are more commonly termed in CSCW research, plans and situated actions, are precisely examinations of this control problem. As already noted, however, Bowker and Star (op cit) ... point to the practicality of issues surrounding degree of control. As they say, "This balance can never be fully resolved (as novices and strangers are always entering the field of work) ..."' (op cit: 232)

If Bowker and Star are right in that these issues are central to any classification system, and if we are right that they pertain as much to ontology-building as to any other kind of

classification scheme, then it would seem that there are good reasons for examining the work that goes into the construction of ontologies – classification practices, if you will – because that work will ramify in the development of ontologies and of systems to support their development. Their success or failure when deployed will not only be a matter of their internal consistency but also the degree to which they meet organizational requirements. There may be a number of dimensions to this, and below we sketch out some of them, based on our observations of ontology-building work conducted over a six month period.

An ethnography of ontology building

Our task here is not to critique the very idea of ontology-building, but it should be said that the existence of various controversies and disputes within this community does indicate that philosophies, purposes and methods are not universally valued (cf. Shirky) It is worth trying to understand what factors might be involved. We will suggest that they can be understood largely in terms of the issues of comparability, visibility and control referred to above. Certainly, those ontology builders with whom we have been involved seem to be developing ontologies in response to problems that they have, such as achieving standardisation in certain areas of web-based practices, making museum-like collections of materials available on line, being able to manage the huge quantities of research data generated by contemporary genetics, make these comprehensively accessible online and enable their use for research purposes. Much of the development is that of codifying the knowledge available to experts in an ontology that will make their expert knowledge shareable, and where one of the acknowledged research problems for the developers is to find ways of capturing expertise in the domain, sufficient to achieve adequate categories for the ontology and that of sharing knowledge about ontology building. The work we are familiar with is not a matter of experienced ontology builders applying their specialised skills to the design of specific ontologies, but, much more, those with particular domain expertise learning how to build an ontology whilst attempting to do so, and looking at the same time to understand what people in other domains are doing to see if they can identify generic problems, can find aids and solutions to their own domain specific problems in other fields etc.

In the bioinformatics area much of this work is intellectually demanding on both sides of the equation and we must not be understood as suggesting that problems in developing an ontology for front-line research in an area of genetics are due to neglect of the organisational problems of project work, since the task itself is a difficult one, involving considerable complexity of biological understanding and a very demanding level of comprehensiveness and integrity in the formation of the ontology. There remains, though, the organisational problem familiar from software more generally, of the distributed nature of experience and understanding of ontology building, such that what has been learned about common problems is not accessibly codified and shared, meaning that developers have to find out about and resolve for themselves many of the general risks and problems of the process. We have already seen aspects of this in our observations of ontology-building work. What we do below is discuss aspects of ontology building work-in-practice that seem to have ramifications for supporting community based ontology building. This discussion covers issues of identifying purpose, rationale recording and timeliness.

Our first data is taken from an ontology building course on a Masters' degree. The following is an example of the kind of instruction students are given concerning how they might proceed with the business of building an ontology:

“The first thing you have to do is establish the purpose. Without a clear purpose, there can be no scope, no requirements and no evaluation. It’s hard to constrain the problem without something in mind, even though you might want to re-use it”

A tension between re-use and specific purpose is evident. In reference to term-collection, a necessary initial step in ontology building work, we hear the following:

“Organize them informally, paraphrase and clarify them to produce informal concept definitions ... paraphrasing is really important ... how will we know what you’re trying to do if you don’t make notes ... photographs are good too ... I mean it, if you have a digital camera or a mobile phone, take photos of the way you organise the cards ... felt boards are useful tools for organizing things ...”

What is interesting about this is the way in which the business of ontology building becomes the business of understanding and noting one’s own rationale for making the decisions one makes, and the use of some very prosaic techniques – photographs, card-sorting, felt boards, and so on – to do so. This is linked to an explanation of how one comes to decide that one’s hierarchy of concepts might be done this way instead of that:

“card sorting is as good a way as any ... the metaphor is highlighting ... here’s a list and you want to group things together ... you have to make decisions about how you’re going to do that ... this is sometimes called, ‘laddering’ ... group things and then ask why ... what do they have in common ... what are their parents and siblings? What other siblings might there be ... for example, if you’re doing an ontology of children’s animals, there are bacteria and fungi, but do you want to include them now? You’re going to get lost in the trees, and it’s very easy to lose sight of the woods ...”

Reference is made to disagreement, because one of the evident features of ontology building is lengthy disagreement, and the need to resolve it at some point, “without a paraphrase you can’t disagree on why we did something ... what can we say about all members of a class? ... all of this does some of that or all of this ... these are the only two constraints we’ve got.”

Equally, the practical issue of recording decision-making, remembering rationale, and so on, will become a great deal more important as ontologies are designed for more heterogeneous communities, in part because the actual process of building will become a more distributed affair. If we look at the kinds of technique deployed above for the recording of rationales, disagreements, etc., then it is quite evident that they will have to be replicated in some other way if distributed ontology building is to be possible. These issues tie in closely with the kinds of debate that inform the ontology-building community today – debates about problems of scale, scope, detail and usefulness. Whether or not, as the realists would have it, an ontology of everything is possible, as a matter of practical contingency important decisions have to be made about the number of classes to be identified, their properties and relations. That is, the ontology needs to classify persons and events in such a way that it covers all relevant possibilities (including the rare exceptions), and discriminates them sufficiently such that the consequences of different events and behaviours can be identified and dealt with. How we begin to identify what “relevant” might mean here is entirely non-trivial.

Closely related to this is an issue which we can describe for convenience as temporal. That is, understanding the processes of development across time may be revealing. The kinds of problem this may entail can be seen in the following:

‘Of course, a feature of microarray technology is that massive amounts of data are produced ...petabytes ... there are innumerable difficulties attached to this, not least that across the range of proteomics; transcriptomics; metabolomics, etc., there will be a range of different experimental methods with different approaches to normalization; different ranges tested, etc. ... even so, the experimental metadata I get is all the same the problem is that biologists don’t see sufficient gain in inputting this data themselves- it often remains invisible. An ontology which provides some of this data would minimize ‘cost’

and maximize 'benefit'. The trouble is, it involves an awful lot of drudge work, especially in respect of coming up with definitions that everyone agrees with. It would have a particular benefit, in that it might provide for 'environmental' information that some biologists would not otherwise think to provide because they don't need it.'

It is not an especially new observation that it is difficult to get people interested in development when they do not see any immediate practical benefit for themselves. From the developer's point of view this can seem like a conflict between public and private interest, between the attempt to develop an ontology that will have utility for all sorts once it is developed, and the demands of people who want to see near term applications. There is, too, a felt lack of understanding between what the developer knows is going into the painstaking, time consuming difficulties of working out even a small part of the ontology and what the users are expecting in terms of speed with which improvements can be delivered. One aspect of the heterogeneous community we have been speaking about, then, is a disparate set of interests. The point is that these varying interests have all sorts of consequences for the way in which any proposed ontology might be structured. Hence:

'Some of what we want to do is just controlled vocabulary, but some of it is to do with other issues, like granularity ... problems we have to deal with include the fact that habitats are not discreet, they blend into each other. They can be described in many ways.'

Related to this is the way in which these different interests coalesce around the practical business of 'getting involved'. This turns out to be hugely problematic. Much of our data concerns the way in which it is difficult to get involvement at the point where the builder needs it, but easy when others see it in terms of their purposes. We would suggest, however, that where ontologies are to be deployed in complex environments, more exploration of what those purposes might be is needed. The reasons for this quickly become obvious:

'We're going to have to consult a lot over terms, but maybe will have to legislate to some degree. We're going to have to allow people to tag stuff up. To some extent, at least to begin with, there are some agreed boundaries ... tropical rainforests have agreed attributes ... and these would be largely unchallenged. There are, however, about twenty different types of grassland within that habitat, defined not only by the type of grass but also maybe by the fauna as well. And then there's all this to do with temperature, rainfall, etc, etc. We wouldn't want to include geospatial information. Habitats grow and shrink but we don't need that ...'

The issue of 'consultation over terms' again has practical consequences for ontology building, for there is controversy over when and how that should take place. On the one hand, as one respondent observed, 'GO [Genome Ontology] started from a use case ... that's why it was so successful ...' In other words, early engagement with users can be seen as desirable. On the other, 'It'll just be a controlled vocabulary to begin with ... with most successful ontologies, the complexity came later. Sometimes you feel like a lawyer, finding descriptions that no-ones going to disagree with ...'

It seems to be a common experience among ontology builders that getting any kind of community agreement over terms in the earliest stages is extraordinarily difficult. It can result in the kind of one-size-fits-all ontology we refer to. Thus:

M: OBI is such a big beast because people felt it had to be all-encompassing and centrally managed ...'

B: 'A big semantic cricket bat'

M: 'there was no modularity ...'

In contrast, it seems that success sometimes comes from a very slow and careful form of user engagement: 'we're trying to keep it quiet ... if we try to develop by committee we're not going to get anywhere for a long time ... but sooner or later we need to put something out there into the community ...'

Conclusions

We have argued that various issues that exercise ontology builders, notably the size and scope of their ontologies (including for our purposes arguments about modularisation; the contingent relevance of controlled vocabularies, folksonomies and fully-fledged ontologies, and the problematic relationship between terms and concepts) are equivalent to arguments about users and what we know about them. It is not surprising that these problems get glossed as problems of ‘community’ and our point is that this is best understood as a practical matter.

This is no simple thing. In the context of ontology building, the familiar problem of the immediate day-to-day purposes of the user measured against the longer term, and meta-level, purposes of the ontologist are particularly vexatious, and magnified by the highly technical knowledges being deployed by all parties. Much of what we have seen demonstrates how the user remains a largely imagined vector of requirements, understandings and practices, with occasional contacts with and feedback from actual persons who might count as users. Those ‘users’ who are involved seem typically to be project collaborators who are not *dedicated* to the use of the ontology, so ‘trying out’ is something they do in addition to other more workaday things. Thus, the interface between ontologists and ‘users’ is a problematic one, with respect to the balance between what ‘users’ know about the practices of ontologists and vice versa. Many ‘user’ reactions do not provide data on users as such, but instead become a problem for the ontologist, a missing feature of the *research/design* point of the immediate exercise, exhibiting a failure to understand the design rationale behind what is presented to them, thus becoming an added charge on the ontologist’s work, explaining to the collaborator/users how what the ontologist is offering is to be understood and what kind of response is wanted from them. Handing stuff over to ‘users’ can become a risky matter, involving the ontologist in attempts to control the amount of work that any trial release might generate from the user response – as so often in design, the inclination to delay release so as to get work completed, to fix problems and to provide a more acceptable release has to be weighed against the fact that one has to go public at some point. Our data indicates how such problems are typically managed, including attempts to manage the how and when of community involvement, and indeed of available ‘versions’ of the ontology.

Associated with this is the ‘never finished’ quality of ontology work. New tools, new domains, and the representation of evolving knowledge are a continual problem. This means that the apparent tension between ‘ontologist’ and ‘user’ is magnified by the fact that ontology builders are not only having to contend with the regular transformation of domain knowledge, but they are also having to learn and re-learn the very business of ontology building as ‘another first time’ whilst they are working on their projects. They discover what they need to know by trying, as best they can, to work out their ontology whilst simultaneously working out how to build a better ontology. Not least, this has implications in respect of trust and authority. There is no doubt, as our colleagues put it, that ‘there are a lot of bad ontologies out there’ and much needs to be learned about the social distribution of expertise in each instance. ‘Expertise’ here is a complex problem since it refers both to evolving competence in the use of ontology tools and to knowledge about the various domains and their aspects that can be modelled (and equally importantly, cannot). Again, it has been no surprise to us that the community we have been studying has an abiding interest in ‘use cases’, where one can look for general, even generic features of ontology building in the solution of design problems in different domains. It is equally unsurprising to us that they find this extraordinarily difficult to do.

In this paper we have argued that the first step to designing solutions is to identify problems. The problem-set we have recognised has to do with a range of issues such as rationale-

recording; identifying purpose; the timeliness of community involvement, and so on. All seem to us to be deeply relevant to the success of ontology building processes, and depend on accurate and adequate knowledge of how those processes are currently managed.

References

- Ackermann, M., Pipek, V. and Wulf, V. (2003). *Sharing Expertise: Beyond Knowledge Management*, Cambridge, Mass., MIT Press.
- Anderson, R. J. (1994). Representations and Requirements: The Value of Ethnography in System Design, *Human-Computer Interaction*, 9, pp151-182.
- Anderson, S., Hardstone, G., Procter, R. and Williams, R. (2007). Supporting the Evolution of Organisational Information Systems. In Ackerman, M., Erickson, T. and Halverson, C. and Kellog, W. (Eds.) *Evolving Information Artefacts*, Kluwer.
- Bowker, J. and Star, S. (1999). *Sorting things Out: Classification and Its Consequences*. Cambridge MA, MIT Press.
- Buckingham Shum, S.J., Selvin, A.M., Sierhuis, M., Conklin, J., Haley, C.B., Nuseibeh, B. (2006). Hypermedia Support for Argumentation-Based Rationale: 15 Years on from gIBIS and QOC. In Dutoit, A.H., McCall, R., Mistrik, I. and Paech, B. (eds.) *Rationale Management in Software Engineering*. Berlin, Springer.
- Groth, K. and J. Bowers (2001). On finding things out: situating organisational knowledge in CSCW. *Proceedings of the European Conference on Computer Supported Cooperative Work*.
- Gruber, T. (1995). Toward Principles for the Design of Ontologies Used for Knowledge Sharing, *International Journal Human-Computer Studies* Vol. 43, Issues 5-6, November 1995, p.907-928.
- Harper, R., Randall, D. and Rouncefield, M. (2000). *Organisational Change and Retail Finance: An Ethnographic Perspective*, London, Routledge.
- Hughes, J.A. Randall, D. and Shapiro, D. (1993). Designing with Ethnography: Making work visible, *Interacting with Computers*, Vol5 No 2, pp239-253.
- Martin D, O'Neill J, Randall D, Rouncefield M (2007). How Can I Help You? Call Centres, Classification Work and Coordination, *Journal of Computer Supported Cooperative Work*.
- Petersson, M., Randall, D. and Helgeson, B. (2002). Ambiguities, Awareness and Economy: A Study of Emergency Services Work, in *Proceedings of CSCW '02*, New Orleans, Louisiana, ACM Press, pp 286-295.
- Pike, W. and Gahegan, M. (2007). Beyond ontologies: Toward situated representations of scientific knowledge, *International Journal of Human-Computer Studies*.
- Shirky, C. Ontology is Overrated: Categories, Links, and Tags, http://www.shirky.com/writings/ontology_overrated.html