

A Learning Trajectory for Ontology Building

David Ribes, University of California, San Diego

Geoffrey C. Bowker, Santa Clara University

Abstract:

Building ontologies involves taking a domain knowledge, formalizing this knowledge into a machine computable format, and encoding it in an ontology language. But what does this mean in practice? In the actual work of building ontologies a specialist will often find that domain knowledges are not readily prepared for ontology building. Written sources such as textbooks and technical treatises are often not precise enough for transformation into logical operators, there may be competing accounts of the same phenomena, overlapping taxonomies and standards, or outright contradictions. Similarly in consulting authorities in a domain, a programmer may find that these experts are not immediately able to state domain knowledge in the terms necessary for ontology building. In short, the experienced ontology specialist often finds that what participants in a domain consider their validated and structured knowledge is not readily compatible with knowledge representation. This paper draws on ethnographic observations of the broader process of building ontologies within the GEON project, a cyberinfrastructure for the geosciences.

An initial conceptual understanding by the domain specialist of what is an ontology is crucial, but never sufficient. We describe a *learning trajectory for ontology building*, in which both domain and IT practitioners come to learn, through practice, the specification and transformation of domain knowledge into discourse accessible to knowledge representation. Coupled with this technical activity is the work of informing, and encouraging the participation, of a larger domain community. The product of the learning trajectory is not an ontology, but rather the practical ability for a IT specialists and domain experts to co-produce a community resource.

Keywords: Interoperability, information infrastructure, database, ontology, learning trajectory, IT-domain interactions, community, organization, pedagogy

Introduction

Building ontologies involves taking a domain knowledge, formalizing this knowledge into a machine computable format, and encoding it into machine language. This is the stripped down technical understanding of knowledge acquisition. But in the work of building ontologies a specialist typically finds that domain practitioners are not readily prepared for ontology building. First there is the enrolment of the domain. To bring experts on board is to inform them of the technology of ontology, its strengths in the face of other interoperability strategies, and the particular work it will require. Enrolling practitioners is securing an investment in technological direction by a domain community. Second is the work of

knowledge acquisition. Written sources such as textbooks and technical treatises are often not precise enough for transformation into description logics: there may be competing accounts of the same phenomena, overlapping taxonomies and standards, or outright contradictions (Bowker 2000). Indeed, one key feature of scientific work is changing ontologies over time. Similarly in consulting authorities in a domain, a programmer may find that these experts are not immediately able to state domain knowledge in the terms necessary for ontology building. In short, the ontology specialist often finds that what participants in a domain consider their validated and structured knowledge is not readily compatible with ontology building. Finally the technical activity of ontology building is always coupled with the backgrounded work of identifying and informing a broader community of future ontology users.

In this paper we supplement a technical vision of ontology work with a broader frame which brings together the coordination of technological resources with the organization of production and the mobilization of domain communities. Drawing on ethnographic field research in the development of the GEON project, a cyberinfrastructure for the geosciences, this paper explores several dimensions of knowledge representation-in-action so as to develop stronger vision and tools for organizing ontology work.

In outlining a frame for ontology work we speak of a *learning trajectory for ontology building*. The learning trajectory includes technical action, such as knowledge encoding, but also a broader set of activities that stretches from introducing the technologies of ontology to the domain, through what is learned by a practice of building ontologies, and to the mobilization of a future domain user community. We can divide the learning trajectory into three components, although in practice these are often interwoven:

- i - the *enrolment of domain practitioners* in ontology building: this involves the initial phases of education by IT for the domain as to what an ontology is, what purposes it may serve and what some of the preconditions for ontology work are.

Conversely it is at this stage that IT will begin to understand the relationship of the domain to data integration projects;

ii- *learning acquisition* or the *practice of ontology building*: in our research we have found the abstract descriptions of ontology building are not sufficient to assist domain geoscientists in formal knowledge representation. Practical learning, learning-by-doing, is required in order to help domain scientists begin to translate their knowledge into inherited categories, logical operators, and predicates. It is at this head-to-head encounter between IT and domain that the configuration of knowledge within a domain will become apparent to the ontology expert. Practical learning and formalizing occur hand-in hand. Domain experts are often initially unaware of the particular configuration of consensuses, ambiguities, ambivalences, or disputes within their fields. IT experts must be prepared to offer one of multiple solutions available to assist in the formation of temporary agreements, represent multiplicities of knowledge, disagreement or uncertainty;

iii- *community enrolment*: an ontology is successful only if the technical work is coupled with identifying and collaborating with a broader community of future domain users. Identifying a future community and finding means to elicit participation is an emerging skill-set. A domain community must, explicitly or implicitly, consent to the 'solidified work' within ontologies.

In identifying a trajectory we are simultaneously typifying an observed phenomena from our research but also providing a conceptual and planning tool for future ontology builders. We take ontologies to be a form of infrastructure (Star and Ruhleder. 1994). By this we mean that it is not a tool for a single scientist or even a research team, rather it is an investment intended to serve as a long-term resource for a broader community. Within information infrastructure projects interoperability is often defined as the common goal of a collective, and ontologies are one means to achieve it. Thus, the definition of success in an ontology project stretches beyond technological deployment to its uptake and usage by a domain practitioners.

This paper is primarily directed at those who sit at the intersection of ontology and domain: knowledge representation (KR) specialists directing their efforts at building ground-up ontologies, and domain experts with the goal of building community resources for interoperability. While in this paper we draw substantially from the sociology of scientific knowledge and the interdisciplinary field of science and technology studies (STS), due to space constraints it has not been possible to thoroughly review the literature. However, we explicate relevant insights from this literature and hope that we point to useful affinities between ontology building endeavours and STS. Before we turn to the empirical analysis of the learning trajectory the next two sections introduce the case, as well as the methodology and research approach.

Case and Methods

For this study we have chosen GEON – the geosciences network www.geongrid.org – as a particularly rich site for understanding the work of building ontologies. In scientific domains, knowledge is highly nuanced and technical while simultaneously under continuous revision and debate. Thus science can serve as an excellent site for understanding the more general difficulties of ontology building. Moreover, the diversity of sub-fields and the emergence of new interdisciplinary efforts (such as bio-geology) have made the call for

interoperability particularly vocal in the sciences – *a fortiori* since the major questions facing us geopolitically (the continuation of fresh water supplies, dealing with climate change and so forth) demand the integrated work of scientists from multiple fields. In the near future ontologies will be deemed crucial for facilitating interdisciplinary work, making scientific research more accessible to policy and decision makers, and ensuring public accountability (Arzberger, Schroeder et al. 2004)

GEON is one such project for the sciences. As an umbrella cyberinfrastructure (Atkins 2003) for the geosciences, data within GEON comes from sources as diverse as geophysics, palaeobotany, metamorphic petrology and geochemistry – fields which otherwise have thus far had few venues for intercommunication. Mandated to serve such a heterogeneous constituency, GEON has the daunting task of providing data and resource interoperability, along with computation resources, data -storage -management and -operating tools (such as mapping and visualization). Organizationally centered at the San Diego Supercomputer Center (SDSC), GEON is a national project physically distributed across the US. The SDSC is the focal point for multiple cyberinfrastructures and a leading edge American site in the development of ontologies for scientific and engineering applications.

In the words of its developers, the GEON project:

represents a coalition of IT and Earth Science researchers that has been formed in response to the pressing need in the geosciences to interlink and share multidisciplinary data sets to understand the complex dynamics of Earth systems. ... The GEON (GEOscience Network) research project is being proposed in response to the pressing need in the geosciences to interlink and share multidisciplinary data sets to understand the complex *dynamics of Earth systems*. ... Creating the GEON cyberinfrastructure to integrate, analyze, and model 4D data poses fundamental IT research challenges due to the extreme heterogeneity of geoscience data formats, storage and computing systems and, most importantly, the ubiquity of *hidden semantics* and differing conventions, terminologies, and ontological frameworks across disciplines. GEON IT research focuses on modeling, indexing, semantic mediation, and visualization of multi-scale 4D data, and creation of a prototype *GEON Grid*, to provide the geoscience community a head start in facing the research challenges posed by understanding the complex *dynamics of Earth systems*...

The ultimate goal of the project is to provide for the development of a more holistic picture of earth processes than is possible with the current information infrastructure, which grows out

of and under-girds the splintering of the field into sub-disciplines. The GEON proposal constitutes a superb example of the kind of work that is going on in many fields of science and human endeavour to best use the multiple data sources and high data flows that characterize all of modern science and most of the important policy work that has a scientific basis (one need only think of the prospective role of the Global Biodiversity Information Facility (www.gbif.net) in determining international biodiversity policy and the role of heterogeneous data in world climate modeling (Edwards 1999)).

The primary data collection method for this research has been ethnography. Ethnography is the study of people and things in their natural settings, with the goal of producing a descriptive account of both meanings-for-informants and action-as-observed by trained researchers. Primarily based on qualitative methods, such as detailed observations, unstructured interviews, and the analysis of documents, in the ethnographic study of science and technology it is necessary to become deeply familiar with technical concepts and practices. Ethnographic observation in its various forms have been utilized for knowledge acquisition within knowledge representation for quite some time. In knowledge acquisition, information technologists turn their attention to understanding the knowledge of a given domain for formal representation. Ethnography, and its coupled qualitative methods, remain a staple of knowledge capture techniques, appearing in its various guises as participant observation (Meyer 1992), expert elicitation (Forsythe and Buchanan 1989), on-site observation (Waterman 1986), apprenticeship learning and teachback interviews (Boose 1989). In this paper, ethnography is used not to understand the domain, but to understand the process of knowledge capture, formalization and ontology building. This turn has allowed us to observe not only the knowledge configuration of a domain, but also the process of translation into the formalizations and language of ontologies.

Our research of ontologies within GEON is part of a larger organizational study of cyberinfrastructure development and of a comparative project on the contemporary strategies of interoperability (interoperability.ucsd.edu). Ethnographic observation began in November 2002. The primary research sites have included:

i- *The Organizational and Communications Structure of GEON*: the weekly workgroup meetings of top administrative managers and the IT team are an excellent vantage point from which to observe the general organizational emergence and functioning of GEON.

ii- *Concept-Space Workshops*: These workshops have been foci for the production of scientific workflows and ontologies, they are one of the points of greatest interaction between IT and domain sciences.

iii- *Geo-Scientist Sub-Groups*: Each geo-science PI works relatively autonomously on GEON projects with a local teams of academic geo-scientists and information technologists. GEON IT and geo-science participants meet collectively two or three times a year in PI and all-hands meetings.

These three sites have permitted observation of ontology building both in a narrower sense of knowledge acquisition as well as in the broader pedagogical and community building functions which are described in this paper. Additionally, the GEON team has granted access to internal discussions, such as email, discussion forums and the technical resources under development themselves. Data collection and analysis have been facilitated by the qualitative research software suite NVivo. The collection of data has followed the methodology of grounded theory (Glaser and Strauss 1973) – initial inductive research is complemented by iterations of deductive analysis which guide future investigation. For example, while our respondents initially described ontology work as a technical activity centered on knowledge acquisition workshops, reflection on our initial findings encouraged us to continue research by following our informants beyond this narrow definition to the preceding activities of enrolling participants and concurrent activities of enrolling the community (Latour 1987). These inductive-deductive iterations are described as theoretical sampling by sociologist Barney Glaser (Glaser 1978, see esp. ch.3). Our central analytic category ‘the learning trajectory of ontology work’ is adapted from Anselm Strauss’ research. A trajectory is a conceptual tool of the analyst for understanding i) the course of any

experienced phenomenon as it evolves over time and ii) the actions and interactions contributing to the evolution of a phenomena (Strauss 1993). The notion of a trajectory allows us to conceive ‘a routine of ontology work’ as emerging in relation to the development of the technology, the organization of GEON, and the knowledge of geo-science itself.

What does knowledge look like ‘in the wild’?

The artificial intelligence experiment is not just a problem of engineering or psychology but an empirical test of deep theses in the philosophy of the social sciences -- (Collins 1990, p.8)

Sociologist of scientific knowledge Harry Collins wrote the above quote during his studies of expert systems in mid-1980's. In some ways ontologies are the inheritors of a tradition of AI and expert systems research. The field of knowledge representation faces questions similar to its parent disciplines, although we believe this inheritance has been substantially reframed. Today's ontology work is informed by more complex understandings of knowledge, of the practice of expert work and of the design of information systems. Earlier efforts in artificial intelligence, expert systems, and automated reasoning were plagued with methodological underdevelopment: the knowledge bases generated were frequently both sloppily constructed and very expensive to build up; new information gleaned by crude knowledge acquisition techniques was highly problematic and poorly substantiated, and rigid logics and design demanded untenable practices by the domain (for a critique of early knowledge engineering methods see Stefik and Conway 1982). Recent work has aimed to ensure more robust methods for knowledge capture, acknowledging the importance of understanding the domain at a fine granularity and that the very question '*what is knowledge*'— the core of epistemology – is a contentious question. Supplementary methods have been imported from disciplines as diverse as anthropology, sociology, psychology, cognitive science and of course philosophy. These efforts have paralleled attempts in the larger computer science and IT community to include user studies and participation in the

design process (Schuler and A.Namioka 1993; Star and Ruhleder. 1994; Mackay, Carne et al. 2000; Oudshoorn and Pinch 2003).

It is this growing awareness and sophistication within the knowledge representation community itself that has prompted us to take a distanced stance on epistemology. In philosophy, 'ontology' is often coupled with *epistemology*, the theory of knowing, or *how* we know what the world is and what knowledge looks like: 'what is knowledge'. While in this paper we argue for the importance of 'how to know' by observing the work of domain scientists and IT specialists as they build ontologies, our argument stretches further than epistemology to the larger development arc of ontologies, what we call the learning trajectory. In this article we take as a methodological principle an agnostic position towards the question 'what is knowledge?' It is precisely this question which is at stake in the production of ontologies. In knowledge representation the object of activity is to root out the location of knowledge itself, to make it available for transformation into discourse and eventually formalization in machine language. In contrast, a *sociology of knowledge representation* takes as its object an entire repertoire of action surrounding knowledge work, what sociologist of science Knorr-Cetina has called an epistemic community (Knorr-Cetina 1999). Our own method is not the identification a site of knowledge for acquisition, but instead to follow our informants across the entire range of heterogeneous activities (Callon 1986) which constitute knowledge work.

What is an ontology?

Just as knowledge is a highly contentious issue, within computer science ontologies themselves remain a going concern. Are ontologies realist, utilitarian or pragmatic? Can automated deductive reasoning produce reliable new knowledge? In this paper we put aside these debates, and instead trace the discussions of our informants: 'what has ontology been for GEON?' Because many GEON participants are also participants in the broader KR community, the understandings of ontology within GEON have reflected many of the meanings which are currently at play within computer science. Below are excerpts from oral presentations or the accompanying slides used by IT practitioners presenting for the principal investigator (PI) team of GEON in its second year – at this point ontology had already become a relatively familiar term for the domain, and the discussion is sophisticated in relation to the earlier phases of introduction (in the learning trajectory 'enrolling the practitioners'):

“ontology specifies a theory (as set of models), defining and relating, concepts representing features of domain of interest”

Other terms which loosely overlap with ontologies: controlled vocabularies, database schemas, conceptual schema, thesauri, taxonomies, inform/semi-formal representations (concept spaces, concept maps, labelled graphs, semantic networks), formal ontologies (OWL, formalization of a specification, constraints of terms).

What are ontologies used for?

1-conceptual models of a domain or application

2- classification of concepts, data/object instances

3- analysis of ontologies, automated reasoning, deductively derived knowledge

4- targets for semantic data registration

5- conceptual indexes and view for searching, browsing, querying or, integration of registered data

When we think of knowledge we often think of formal encodings first: in science we think of textbooks, articles and other publications; in business we may look to statements of work and paper trails; in law to templates and precedent. Secondly, we may think of experts themselves as repositories of knowledge e.g. scientists or engineers. In researching knowledge work in action, sociologists of scientific knowledge have followed scientists and engineers across a surprising diversity of locations and kinds of activity (Star 1995). Sociologist of science Steve Shapin (Shapin 1989) opened the ground for looking at the technicians in the work of Robert Boyle and his air-pump. Shapin points to technician's significant expert work in knowledge production and simultaneously notes a moral order of the historical period which left them out of most accounts: 'invisible technicians'. Historian of science Peter Galison (Galison 1997) has extended this work in his studies of 20th century microphysics. Galison has remarked that modern physics has relied on complex collaborations between experimentalist and theoretical scientists along with another group of materially oriented technicians which support the work of both. Cambrosio and Keating (Cambrosio and Keating 1988) in their studies of molecular biologists have noted that much knowledge is distributed in local practices and methods of communication, accessible to negotiation and argument, but difficult to codify.

Thus social studies of science have 'found' knowledge in many sites, but more importantly the field has fostered a methodological tradition of openness to locations of

epistemic development. By shifting the unit of analysis from formal knowledge and experts to *relationships* between these, a more complex image of knowing has emerged, the object of which is not to pin down knowledge, but rather to track its production, shifts and movements. There is an affinity between ontology builders and sociologists of knowledge: both share an interest in knowing, and an necessity to maintain a broad field of inquiry into the possible repositories and vectors of knowledge.

Neither 'science' -- itself an enormously heterogeneous body -- but also not disciplines such as geology nor sub-disciplines such as geochemistry have come to solid practical or theoretical conclusions as to 'what is knowledge.' In the face of novel instrumentation, methodologies and concepts, this too is a question for scientists. In one GEON ontology workshop a geochemist remarked:

When we talk about knowledge recovery, we talk of this: the rock record. The rock record covers four billion years of planetary history. [...] If we can work on this planet and understand its ontology, we have a good chance of understanding the other ones as well [...] These are knowledge reservoirs, this part that is called the lithosphere contains knowledge of different kinds. How do we integrate the knowledge reservoirs?

This comment characterizing the Earth as a knowledge reservoir sparked a minor revolt and debates ensued over whether knowledge could be manifested materially in 'mere rocks' or whether it was a human capacity alone. Some suggested that rocks be thought of as the earth's memory, but this also led to accusations of anthropomorphism. Finally, frustrated information technologists turned the conversation away from "metaphysics" in favor of practical ontology building. The irony of the elision was noted by no one. The question of knowledge is at much *at play* in a sociology of knowledge as within knowledge representation and domain sciences.

The word 'ontology' derives from philosophy. Philosophical ontology is a theory of being and existence, or to put it another way a theory of what there is in the world. It has accompanied the development of computing theory from Leibniz to the present day (Leibniz,

Cook et al. 1994; Slaughter 1982) While there are several large scale efforts to build universal upper level ontologies – a ‘unified field theory of ontology’ – the majority of work in the field is much less ambitious, primarily directed at ‘on the ground’ local applications. Our informants’ ontology work focuses on building these local ontologies in which goals of upper level integration still seem distant. In the close encounter with domain that comes with lower-level KR, and especially with our IT informants, ontology development has led to the common understanding that there are many ontologies, each demarcating a domain-specific reality. In a graduate seminar for ontology building an IT GEON participant remarked:

the standard textbook examples in databases are always 'employees' and 'projects', and 'departments', or you know 'courses' and 'professors', its very clear what these things mean. That's not how real data looks like. Well in the commercial world, like bank transactions, and so on, its probably fairly clear, but in the commercial world there are standards for how businesses present their data [...], and also the complexity of the data is not so immense. Its kind of well specified... But with the sciences, here we have chaos. Every scientist lives in their own world, and they do their own stuff, so that's why the scientists need to come together and integrate, work towards common standards and so on.

We do not need to agree with this informant's view of science to understand that ontologies are being applied differently in science and ‘bank’ – or business/commercial – spheres. Science does not serve as a particularly fruitful site for the sociological investigation of ontology building because its knowledge is a more significant, rather we take as a premise that when explicitly discussed *knowing* is always a charged issue. The significance seems instead to lie in the kinds applications of ontologies within business and science. Ontologies in science are seen as an aid to the research process itself – the sometimes ephemeral and always emergent practice of knowledge production. The objects of scientific research, as such, are themselves not fully specifiable. As historian of science Rheinberger has noted about scientific objects: “Such entities, then, have a peculiar, paradoxical time structure characterized by “recurrence” [...] These research entities, for the very same reason, do not belong to the realm of deliberate construction either. The mode of scientific existence peculiar to such entities derives precisely from their resistance, resilience, and recalcitrance

rather than from their malleability in the framework of our constructive and purpose ends.” (Rheinberger 2000, p.272). To put it another way, the science applications for ontology within GEON are focused on interoperability and knowledge discovery at precisely those points which are of interest to contemporary geo-scientists. This means that there can be a great deal of 'heat' at the site of ontology building for the sciences – with the reputations and careers of particular scientists or institutions at stake, not to mention 'truth.' As the site of explicit epistemic and ontologic development, empirical sociological investigation is facilitated because controversy and disagreement is drawn forth and made observable by informants themselves. To the contrary, within 'bank' or business based knowledge acquisition, conceptual complications are left buried in history where agreement has long been secured and where thus far the application of ontologies has focused on agreed-upon knowledge forms.

In this paper we identify a learning trajectory of ontology development. We take as part of our research that work which identifies and transforms 'knowledge in the wild' into computable knowledge representations, while simultaneously producing a wider frame for the observation of ontology work (Hutchins 1995). The learning trajectory stretches beyond a narrow definition of knowledge, to the work of *aligning* both *within* the domain, and *across* IT and domain boundaries. By alignment we mean simultaneously a practical know-how for building ontologies but also the substantive work of bringing together the organizations and institutions of a domain. In the first step, participant enrolment, domain practitioners and ontology experts must come to a collective understanding of the goals and methods of interoperability and knowledge representation. This pedagogical relationship is initiated in enrolment, but it is only in the learning-by-doing of building ontologies that domain practitioners come to be able to resolve their knowledge into discourse. It is through this process that domain practitioners come to see their work as a commitment, and feel they must

turn to their broader epistemic community. Enrolling the community involves developing techniques for communicating to a larger constituency, methods of 'community building', or 'community outreach'. What we observe is a two fold alignment: first, between *IT and domain participants* as they build ontologies, but we also see an alignment of the emerging ontology and the broader *community of the domain* participants.

The Learning Trajectory – Participant Enrolment

GEON's 'kick-off' meeting was held in November 2002. It is difficult to capture the vast scope and ambition of this initial meeting of geoscientists, information technologists, education experts, and a lone sociologist. Held at the San Diego Supercomputer Center (SDSC), the technical and administrative core of GEON (Scott 1992), many of the IT experts were already collaborating on other projects. Meanwhile many of the geoscientists, spanning broad disciplinary differences, were unfamiliar with each other, the IT experts, and the future technologies of GEON including ontologies. The kick-off meeting served as much as an introduction to the nascent GEON vision of a 'cyberinfrastructure for the geosciences,' as to the participants with each other, to the host organization SDSC, and to the information technologies. The following data is drawn both from the initial GEON kick-off meeting and from the all-hands meeting held six months later.

GEON was created with the mandate of producing data and digital resource interoperability, along with the provision of sophisticated search and query tools. The IT team had, from early on, already looked to ontologies as the means for this interoperability, but the task of enrolling the geoscience community remained. At GEON's kick-off meeting knowledge representation specialists unveiled a barrage of pedagogical presentations intended to educate geoscientists in the arcane realm of ontologies and concept-spaces. Many IT-savvy domain scientists were present, along with representatives of endeavours in digital libraries (Digital Libraries for Earth Science Education - DLESE) or national institutions

which often double as standard setting bodies such as the United States Geological Survey (USGS). With such an audience the enabling possibilities of ontologies – database compatibility, inter-disciplinary collaboration, controlled vocabulary queries and the potential for softening the politics of standardization – were not lost on the audience. Thus curiosity was piqued in the domain practitioners. This said, unfamiliarity with the technical terms of ontology building, diverging presentation styles between IT and geo-science, and even the kinds of visual representations familiar within computer science but foreign to geo-trained eyes (e.g. software architecture diagrams) resulted in several miscommunications, frustrations and confusions.

The first step in enrolling participants for knowledge representation work is the task of posing of the problem which ontologies seek to remedy. While stovepipes, incompatibility, and legacy databases are the everyday headache of computer science, and the bread and butter of ontology efforts, for geoscientists these problems have often been relegated to a little noticed substratum of technical infrastructure. Even the principal investigators of GEON, a highly technically informed cadre, have rarely had to encounter on such close terms, its database architectures. Knowledge representation requires a doubly-technical engagement between IT and domain: the specificities of knowledge must be brought forth and aligned with the logical requirements of ontology languages. In the first four days, with many hours of PowerPoint aided descriptions and followed by open discussion, it was only possible for the geoscientists to gain an impressionistic understanding of the detailed work of ontology building. A nuanced understanding would only be learned later, in the practice of knowledge acquisition; enrolment served to bring domain scientists on board, to begin a pedagogical trajectory of understanding the problematic of interoperability, and the particular solution ontologies promise.

Enrolment is not unidirectional – while domain begins an understanding of the problematic of ontology, the IT practitioners begin a familiarity with the 'data-politics' of the

domain. A predecessor of GEON is BIRN – the bio-informatics network – also centered at the SDSC. Many of the IT experts within GEON are also participants in BIRN – and thus it served as an early model for the organization and technological planning of GEON. One initial vision was to begin to define a ‘unified geosciences language system’ (UGSL) in a similar fashion to the Unified Medical Language System (UMLS) already established in medicine (i.e. metathesaurus, semantic network, lexicon). However, unlike medicine, geology has not had a one-hundred year history of standardizing its specialist language across its various sub-domains (Bowker and Star 1999). While as far back as the 1830’s there are tracts calling for the stabilization of a geological language – particularly fossils – it is an odd feature of geology that there has been greater disagreement in the field over time.

Consequently a controlled vocabulary would be a substantial undertaking: going out in the field may involve idiosyncratic data collection methods; there are many sub-disciplines in the geosciences which do not have traditions of data sharing; data is frequently considered proprietary. This is true even in organisations such as USGS where data is mandated as public, but in practice many practical barriers stand in the way of dissemination¹. It is only with a progressive familiarity with the domain that knowledge representation experts can begin to gain a realistic understanding of what challenges are faced, and what solutions are plausible.

Aside from data-politics IT experts began to understand the place of information technologies themselves, including diverging understandings of advancement and success. Within physics Peter Galison has described the rise of computer/model based approaches to research. Physics has a long tradition of experimentation, with established methodological criteria and means of communicating results. With the introduction of model or computer based simulation, a methodological conundrum arose in physics: was this a form of experimentation? Was this an empirical research, or some form of thought-experiment?

¹Years later, the UGLS was abandoned as a GEON project – partially due to funding constraints, but also in the face of realizing the enormity of the task.

Debates as to the status of knowledge produced by 'in-silico' means (Atkins 2003) ensued, and the future careers of hybrid graduate students was held in question. It was decades later that a place for models and simulations came be established within physics with its corresponding criteria for communication, journals for publication and undergraduate curricula. As computing has entered the earth sciences similar sorts of debates have played out, with various specificities. For example, geo-physics, with its large quantitative datasets and ties to other branches of physics has had the benefit of learning from the experience of the broader field of physics, while paleobotany and its stronger historical and qualitative traditions is only beginning to face these questions.

In GEON's kick-off and all-hands meetings, information scientists began to understand the general orientation of the geosciences towards IT. While in computer science, an ontology has intrinsic value – its very development is a success – in the geosciences IT projects would not be considered ‘science contributions’ to geology. Put another way, a built-ontology would not be considered a success, rather, only geologic knowledge derived from an ontologies' existence would constitute the success of GEON². GEON's ties to its funding body, the NSF, is both through the computer science (CISE) and the geo-science (EAR) directorates, and thus it must negotiate these multiple commitments in the distribution of its development efforts; GEON must produce both IT advances and knowledge about the earth. In this first step in the learning trajectory geoscientists began to *know of* the techniques and technologies which would be employed for producing interoperability, while IT experts began to understand the data politics of the domain to be represented.

² What does and does not count as a scientific contribution is often invoked by scientists themselves, sometimes within a general consensus, but often as a debate over the boundaries of the category itself: ‘what counts as new knowledge in geology.’ This is particularly the case with IT-science collaborations where IT developments are not considered advances – ‘databases are not science’. From a historical perspective, however, it is clear that what counts as a science result is constantly shifting within any given science. For example Galison (ibid) in his studies of computer in experimental physics; Winkler has shown the rise and fall of images as knowledge within early-modern astronomical communities (Winkler and Van Helden 1992); and Epstein (Epstein 1996) has shown the reconfiguration of standards for medical clinical trials in the face of AIDS activism.

In enrolment there is also the creation of the particular division of labor which characterizes relations between domain and IT in knowledge representation: a relationship in which KR experts attempt to establish distance from the *content* of scientific knowledge. We will see that re-creating this division is a continuous effort, which was initially enacted primarily by IT practitioners but as ontology work continued in GEON was progressively taken-up by domain geo-scientists. Below is an excerpt from a class taught on ontology development, intended primarily for computer science students, the example is drawn directly from applications in GEON:

B: Here we have the scientist's question [reading from the slide]:

“What is the distribution of the ... I don't know ... uranium lead [hesitating] surplus of A-type plutons in Virginia?”

I don't what A-type pluton's are. I barely know what plutons are, Ok? You see that these guys use a language that we as non-geologists have trouble understanding, and then they use databases and they want us to help them integrate their data. What can we say? We can say put all the relevant information in the database, but still you have all these different databases. What we have to do is get them to tell us how to connect ... the A-type pluton column in this database to the uranium lead in this other database.

In this extract B is first expressing a relative ignorance regarding the details of geochemistry or geophysics, while also marking-out some familiarity with the domain: B knows, to some extent, the nature of a pluton. Knowledge acquisition requires a comfort with the language of the domain, but specific details are left to experts. Secondly, he is instructing his students to leave aside the particular content of the scientific knowledge and to focus on the relations between concepts, and specific connections to database schemas – plutons and uranium lead become predicates connected by a query, which in turn must access particular columns in at least two databases. Here, an ontology must lay on the topology of scientific knowledge like a wet blanket, capturing a surface to support what is pragmatically necessary for interoperability or information navigation. When domain specific IT experts create knowledge representations, e.g. a geoinformaticist, this division of labor is not as clearly visible. However, the ontology experts within GEON are not tied to any geo-science, and so there is an effort to capture while remaining distant as to the content of the domain. To put it

more bluntly, ontology specialists often are indifferent to the configuration and controversies of domain expert knowledge per se, rather the concern is to create a functional knowledge representation which faithfully reproduces internal conceptual relations with a sufficient granularity to achieve automated computer action. In practice a clean division between specifying content and creating representation becomes difficult to maintain but what we wish to draw attention to is the effort to inculcate in geo-scientists the sole responsibility of specifying domain knowledge. Since the act of naming objects in the world is invariably political in science (Bensaude-Vincent 1989), particular attention must be paid to the actors' own categories. In the words of a GEON ontology designer:

One of the most important principles is to utilize terms and methods derived from the way experts communicate in their local, day to day work. In the context of the challenge problems [an ontology being built], if the experts think and refer to the first input parameter as A, then we use the term *A* when eliciting its estimates. Likewise, if the initial elicitation demonstrates that experts think of uncertainty as a range or interval of values, then it makes sense to elicit in those terms.

The scientist can only be enrolled in the ontology if she can see her own worldview in it. This is not immediately apparent at this stage of the trajectory, in fact, IT participants may even downplay the knowledge involvement required. In the initial pedagogic stages of introduction to ontology we have often observed the use of 'pop examples' such as the as wine classifications, or 'smart' searches for commercial goods. For example, an ontology enabled search for 'beauty' may automatically include 'skin, hair and nails' but exclude 'plastic surgery.' For IT educators these kinds of pop examples have several advantages: they are light and entertaining and maintain the attention of the audience; they require little expert knowledge and thus can travel across domains with the IT specialists; and because they are not closely tied to the domain are unlikely to foster internal debate amongst those experts present. However, it is precisely these kinds of examples that precludes an early sophisticated understanding of the future difficulties in building detailed knowledge representations. The triviality of the example for domain participants obscures the knowledge stakes involved . This is not to say that there is no *epistemic stake* involved in wine ontologies, rather that the

use of a domain distanced knowledge (e.g. geology from wine) underplays an epistemic significance in pedagogic explanations of ontology. One can be assured that building ontologies for wine involved detailed and specific knowledge capture work, which was taken very seriously by wine aficionados. Proximity to knowledge endeavours charges epistemic issues – just as wine classification is distanced to the average geologist, we can also look within geology to see that pluton classification is epistemically-distanced for the paleobotanist. The relatively distanced importance of ‘pop examples’ has both pragmatic uses, and unintended consequences. If in describing ontologies detailed domain examples are used, it is quite possible that domain scientists would focus on the technical knowledge, perhaps even leaving aside the ontology education at hand in favor of deconstructing the represented scientific knowledge. Thus by using pop examples such as wine ontologies, the technology becomes foregrounded rather than the knowledge representation. On the other hand, using wine ontologies can portray ontologies as 'depoliticized' hiding the kinds of detailed knowledge articulation that will be necessary to complete the learning cycle of ontology.

In summary, the initial phases of the learning trajectory bring forth the problematic of interoperability for the domain and the particular solution offered by ontologies, they also inaugurate a division of labor which keeps domain content separate from encoding practices, and they introduce the IT members to a configuration of data politics within the domain. These initial encounters are presentations *about* ontology, and can only begin to familiarize the domain with the kind of work for building: in these initial discussion we observed very little consideration of the difficulties of actually creating formal knowledge representations, and few discussions of the commitment ontologies imply for a larger domain community. In fact, in examples such as wine ontologies or smart shopping searches, we observed an displacement of the complex issues for a later time. It is in the practice of ontology building

that acquisition is learned as a skill, and the commitment of the individual domain participants to a larger community becomes clearer.

The Learning Trajectory – Learning Acquisition and the Practice of Ontology Building

The second step in the learning trajectory is a practical pedagogy, learning by doing. Domains carry with them epistemological traditions: ways of knowing, and criteria for what is considered knowledge. The methods of knowledge acquisition must be tailored to the configuration of knowledge in a domain – and in turn the domain must learn to 'speak' their knowledge in a language accessible to machine encoding. While in the received understanding of science the myth of a single scientific method abounded, more recent research in the history, philosophy and sociology of science have uncovered a plethora of domain specific methodologies, trials and vehicles for the establishment of an accepted knowledge, and great shifts over time in these methods and criteria. For example, historian of geology Martin Rudwick has traced the evolution of visual languages of evidentiary production and dissemination within geology. He argues that this language of visual coding – a means for representing topographical, distributional, structural or even causal features – emerged in tandem with an increasing knowledge base of the field itself:

During the period in which 'geology' emerged as a self-conscious new discipline with clearly defined intellectual goals and well established institutional forms, there was thus a comparable emergence of what I shall call a visual language for the science, which is reflected not only in a broadening range of kinds of illustration but also in a great increase in their sheer quantity (Rudwick 1976).

The visual tradition has remained strong in geology to this day, and in GEON ontologies and interoperability are closely linked to 'needs' for visualization and GIS mapping. In developing ontologies an abstracted notion of need is easy to identify – “ontologies for interoperability, for knowledge discovery or to facilitate visualization” – but it is in the practice of ontology building that domain scientists come to understand the involved complexities of specifying

their knowledge with specific tasks in mind, identifying specific technology requirements, and means to achieve their goals.

What is knowledge?

In this outtake we set aside our epistemic 'agnosticism' in order to review some of the ways in which social studies of science and technology has characterized its descriptions of knowledge – these findings are drawn from the sociology, history and philosophy of science. It is in learning by doing that domain scientists begin encountering surprising and unexpected difficulties discussing knowledge in the format that it is being elicited. We hope this section will help facilitate understanding why formalizing semantic relations for computable encoding is an exceeding arduous enterprise:

Knowledge:

- < **has a *tacit dimension*** not immediately available to discursive expression (Polanyi 1964; Ravetz 1971)
- < **is *distributed*** between people, things and systems of action, including formal and informal procedure (Douglas 1986; Latour 1988; Hutchins 1995)
- < **is *immanent or local***– it is developed in response to certain problem sets (Polanyi 1964; Lakatos 1976) and requires a skill for movement between the specific and the general (Kuhn 1977; Forrester 1996)
- < **is *embodied*** – it is known as an action-in-practice not immediately accessible for transformation into a formal set of propositions (Lave 1988; Shapin 1989)
- < **is *heuristic*** – knowledge claims may serve to help understand a phenomena, even when practitioners question claims of faithfully representing a reality (Sismondo 1999; Bowker 1988)
- < **is *multiple*** – each sub-domain has its own set of methods, tools, heuristics and criteria for truth – styles of reasoning and roles for logic may vary (Livingston 1986; Hacking 1990; Messer-Davidow, Shumway et al. 1993)
- < **is *contentious*** – knowledge is often the focus of debate, struggle and contestation – science is both a repository for knowledge but also a site for its development and review (Collins 1985; Burchfield 1990)

In the first step of the learning trajectory, enrolment, the curiosity of scientists in GEON was piqued. Over two meetings geoscientists had established a more detailed understanding of the problem that ontologies offer to solve and a nominal understanding of a division of labor between geoscientist knowledge specification and IT encoding. Next, scientists began turning to their IT collaborators and asking to build an ontology. But at this point a significant gap was revealed between abstract understandings of ontology and a know-how for proceeding. In the following quotation by an ontology expert, many of the initial encounters with practice are summarized:

different scientists come to me asking "we want to have a workshop to define ontologies." That's very good, we're very happy to host that, to do that, and help them with that. But the issue is to do what? What kind of ontology do you want? What do you need it for? But sometimes it's actually useful to conserve that [ambiguity], to get people together from the domain. We've had people here from geo-chemistry, people here from seismology, so within that group, lets say seismologists, scientific representative persons from a domain, they start all of a sudden arguing heavily about the things they do, the way they view the world. But if you put them into this exercise of trying to find ontologies, of what are the things they care about, what is important for them, what are your analytic methods, how do all these things work together, and how can you create more... uh.. how can you share knowledge, how can you work together in some sense. Ontologies can be that catalyst, or they can create a lot of tension, you know...

This outlines many of the principles that guide action within GEON. Ontologies, within computer science, are of course itself an active field of research with internal divergences in theoretical and methodological approach. For example, whether ontologies should be application specific or independent remains a general issue of contestation(Gruber 1993; Guarino and Giarretta 1995; Guarino and Welty 2000; Smith and Welty 2001; Smith 2003). Without necessarily subscribing to one view or another, knowledge representation practitioners within GEON suggest that ontology building should be *driven* by specific scientific application. "Application drivers" are believed ensure that the ontology will be useful in specific scientific inquiries (a mandate of GEON, *ibid.*) but also that the scientists themselves will continue to invest their efforts into the long and often laborious task. Application specificity ensures that that scientific interests in knowledge production will become tied to the functional completion of the ontology. Returning to the informants quote: the task of building ontologies also begins to inform scientists of their internally diverging knowledge commitments. What appeared to be a shared epistemic umbrella -- "geoscience" or even "geochemistry" -- begins to break down into a finely grained mapping of differences. It is this identification of difference within the knowledge domain, and its transformation into explicit discourse accessible to machine encoding that primarily characterizes the learning by practice of ontology building. Lastly, from this quote we can see that ontology work in practice encourages the domain to begin ordering their priorities, and specifying their 'needs' beyond abstract notions. At a workshop directed at GEON and geo-ontology building,

knowledge representation scholar Deborah McGuinness expressed the specification of needs as a emergent process over time:

There's no quick answers to it. There's answers, but it's a dialogue. Because we have to figure out what your requirements are [...] even the most articulate person that you can encounter today, and that we'd hope to encounter tomorrow, after even an hour of conversation, or even days of conversation, we actually identify over time that there's all these other requirements. It's only then we are really able to get going [...] There's plenty of starting points for this dialogue, but there's no one best starting point. And there's no way in world you're going to pick up one shrink-wrapped [software] and have it represent your needs. The field is young. Even though there are actually all these starting places, there's tremendous variation in needs.

What tasks an ontology should facilitate, what the current requirements of a community may be, are not immediately accessible for expression, but require time and discussion. Interactive dialogue between IT and domain is necessary for elaboration on available solutions but also to curb excessive technological faith. Misconceptions of the current state of technology can lean both towards under- and over- estimating performance, and requirements are shaped in relation to shared models of technical capacities. Thus we should understand the development of *specific* needs as emergent, dialogic and concurrent to, rather than preceding, ontology practice.

The GEON IT team has called their particular method for knowledge acquisition 'concept-space workshops' and more recently have begun calling it 'community-based ontology development'. The distributed nature of GEON requires bringing together experts from across the continent for punctuated bursts of capture, followed by extensive revision via e-mail and video-conferencing. During the concept-space workshops geo-scientists and IT experts co-located for two to three days and sit in conference rooms, 'hammering out' formalized representations of geological knowledge:

Here is the recipe that we've currently applied with some success:

One, you lock up scientists for 2-plus days; add some CS or knowledge representation types to hang around there; then you create concept maps; you refine those, following the meeting, or we turn to local geoscientists [at the SDSC]; then, have other scientists to visit us [at the SDSC], so we can

work on these, so we iterate. In this way we go from napkin drawings, which is a very useful start, to concept maps, to sometimes really formal approaches. We need not always go to the formal, we can sometime stop at the concept maps. But in order to go to correct knowledge representations we need this situation between scientists and domain scientists.

This method for knowledge acquisition is in the lineage of expert elicitation (Meyer and Booker 2001) and more generally is based on a social model of expertise (see Gaines 1989 for an elaboration from within the KR community) in which knowledge is considered the shared jurisdiction of a domain community. Knowledge is seen as a *distributed* phenomenon rather than localized in particular individuals or texts. This method of a collective practice of knowledge acquisition -- rather than an individual interaction between formal knowledge (e.g. textbook) and captor -- is particularly apt for science, where the content of knowledge itself is often the site of debate, controversy and tension. Concept-space workshops in GEON have become a location for both debate and consensus building discussion. The IT teams hoped that through these discussions geoscientists would either come to an accord on knowledge (and thus be able to stand-in for the community) or IT would be able to represent the uncertainty in their ontologies. Concept space workshops were thus an attempt to ensure that the responsibility of knowledge determination did not force a *de facto* shift onto the shoulders of the IT team.

IT ontologies have their own epistemology: what and how the computer can 'know' is very particular, limited by the availability or development of description logics and the level of formalization in an ontology. Ontology development can range from a tool facilitating the work of a user, to system-accessible formalisms. Formal ontology requires a substantial investment of time, effort, and as we shall see, community enrolment. Furthermore, it may not always be possible to produce knowledge representations at the granularity necessary for system automation, especially within the nebulous edges of research knowledge. For this an alignment between the domain and the language of ontologies is necessary. The second step of the learning trajectory includes what we call *epistemological alignment*: a double movement of the domain specialist reflecting on her own knowledge

base and beginning to reconceive it terms of a language of logical operators; in turn IT experts must establish a sufficient personal familiarity with the domain knowledge and with participants such that they may offer themselves as a resource for this translation, all this while skirting the difficult line of intervening in the domain knowledge.

In the initial phases of face-to-face ontology building at GEON, the concept-space workshops, we found that a great deal of time was spent educating the domain scientists on criteria for successful ontology building. For example, KR experts have asserted that no one should assume that scientific data are necessarily better defined than scientific theory:

When it comes to scientific data collection – I mean when they are collecting their data of the world out there – sometimes even they don't know what they're looking at. If it's a new instrument, or if they're measuring something they've never measured before, they might only have a vague notion of what they're up to. And somehow they're supposed to be telling us how to relate this data to other data, or to general categories of knowledge? But they haven't even come up with a category for themselves!

This is not a criticism of science, rather a necessary understanding of what kinds of data to expect: research methodologies will produce *raw* data. This form of indeterminacy is a characteristic of science rather than an exception. If we conceptualize science as a *process for knowledge development* rather than a source of truth, it becomes easier to understand that contemporary developments are under continuous scrutiny, debate and revision; in the words of one ontology expert:

It's both frustrating and exciting that we have to think about these things [ontologies] changing. After all science is about the movement of ideas, not just unchanging fact. Our tools are going to help science, and so they have to somehow match that mobility, rather than somehow holding it back by being too solid.

Given that the IT expert does not wish to intervene in scientific debates, but also requires some consensus in the domain community to do his work, this kind of controversy, or *contentiousness*, can represent a significant derailment in building ontologies. Once identified and encountered the ontology expert must avail himself to the domain and offer strategies for going on in the work of ontology development. In one example that might seem trivial to those unfamiliar with geology, an extended discussion emerged regarding the definition of

the color red. Over several minutes the debate become somewhat heated. Knowing the data politics of domain can mean following such fine grained distinctions as color schemes in map-representation. There have been various standardized legend color schemes previous to digitalization, and many since – some of these are tied to state entities such as USGS, or to similar bodies outside the US. Thus, in determining the spectral band of a particular shade of red, there is also a running backgrounded discussion of alignment with national state bodies or larger world-wide trends in map representation. Picking one standard for ‘red’ may exclude another. Exclusion, or mismatched categories can lead to complications in data integration, such as leaving out an important dataset during an ontology enabled search. While in even the most technically minded geological discussions such details can be set aside when encoding ontologies detailed granularity becomes crucial and momentary agreements can become programmed commitments. Conversely, having never explicitly discussed such details, domain experts can find it challenging to produce consensus at such scales of granularity.

In encountering controversies, the inability of domain scientists to agree on a particular semantic definition or relations between definitions, we have observed several solutions employed during the workshops or suggested by knowledge representation experts:

- i- *decrease granularity*: deal with the issue at a higher conceptual level where the domain has established a stronger consensus;
 - ii- *pragmatics*: encourage the domain experts to form a working consensus in order to continue the process;
 - iii- *rain check*: leave the problem aside for a later time, experts may be able to resolve the issue with a review of evidence, referring to the literature, consultation with experts or, in the long term, production of new evidence;
- or

- iv- *represent the uncertainty*: technical solutions within ontologies work permit encoding multiple knowledges, disagreement, uncertainties, ambiguities or ambivalences³

The techniques for resolving difference are a skill-set. Both the identification of differences, understanding their implications, and knowing what kinds of resolutions are possible is learned by practice. Sub-communities in a discipline may hold diverging beliefs about a particular phenomenon, or commitments to particular domain institutions, and previous to exposure through the explicit formalization of ontologies these divergent beliefs or commitments may be held unproblematically. An initial conceptual understanding by the domain specialist of what is an ontology, is indeed crucial, but it is never sufficient. Domain specialists have learned a specific way of knowing and have particular definitions of what constitutes valid knowledge in a field; ontologies require particular configurations of knowledge representation which are rarely readily compatible with the current structure of domain knowledges. But our real concern in this paper is not to show that we all know differently, more practically to ask how different ways of knowing are getting represented on the ground. In epistemological alignment the IT specialists gather a feel for the domain and its native language, while the domain scientist begins to transform into discourse her knowledge into a machine computable format. This learning trajectory is achieved through the practice of ontology building – the communicative work between IT and domain – rather than through reading ontology guides or other formalizations which a domain specialists often find alien, impenetrable and irrelevant (see Orr 1996 for a discussion of the ecology between personal communication and formal flow charts in technical repair). Within scientific communities it is often the concepts themselves that are at stake in debates, and we

³ In practice we have never seen the application of this technological solution, rather we have only heard it referred to.

have observed patient IT specialist wading through these internal controversies while suggesting resolutions for domain scientists. It is important for the IT specialist to understand her involvement in this social dynamic, and to learn to facilitate the usual difficulties which arise -- domains experts are, by their very position, less experienced or interested in the general problems of ontology building, while for the IT specialist it is potentially a life-long career. This fine grained activity of knowledge acquisition is coupled with a larger procedural development for how to go about ontology building.

Even for IT specialists the initial framing of the ontology workshop method was supplemented over time with an evolution of the various organizational commitments which have to be secured from the domain scientists. Because transforming knowledge into accessible discourse is a learned skill, and requires extended discussion amongst the domain scientists, single ontology workshops are only the first step in building these. Initial meetings within GEON resulted in “napkin drawings” which required substantial discussion, revising, and formalization before they could be 'handed-off' to knowledge representation experts:

D: So you guys did some recordings and some notes that you are going to pass to us?

J: Ah yes, so once this meeting is done, we are going to set up some action items, and people will attempt to send the information and send it around to make sure that the intent is right, and then get it back to the GEON folks, so that we can put on the web, and then ultimately we need to build a formal ontology. But I think that we have some clear hierarchical relationships defined.

D: one thing, J, works the best if you could hand-deliver that report.

A: Hand deliver?

D: i.e. spend a couple of days with us, and not just [give us] an ontology report, that would be the ultimate.

A: oh yes, what we were just thinking, is that the next step is to start some formal ontologies. And we've already talked to participants -- that actually to formalize it we need to be there, so that we can start talking with the computer people, to make sure that what we're proposing is actually going to work.

D: and that's what I'm saying, its not going to work by email. When you build these things you u have to bite the bullet at some point and come down [to the SDSC]

Having already educated various groups of domain scientists at the San Diego Supercomputer Center the GEON IT team was familiar and had already developed a procedure for the initial introduction of ontologies: presentations, PowerPoint slides, open discussion and demos. Similarly the framework for ontology workshops preceded GEON formal inception at the 2002 kick-off. But it was only later that the procedural organization for sketching ‘napkin drawing’s and then revising, was collaboratively determined between the geo and IT participants: can initial ontology meetings be held between geoscientists without IT experts? Can revision meetings be held over the phone or video-conference? Is it necessary to physically co-locate at the SDSC to capture more formal ontologies? Over time the method of building ontologies in GEON has grown from ‘ontology workshops’ to an entire repertoire of resources and procedures for organizing work over time and national distances. The particular procedures developed are specific to the kinds of resource available for conducting co-located and distanced work: email, and videoconferencing and groupware facilitates distanced work, in GEON they found that only co-located offered the ‘bandwidth’ to substantively create ontologies.

In summary, the second step of the learning trajectory – learning acquisition and the practice of ontology building – includes the practical skill of bringing knowledge into discourse, of learning techniques to establish the agreement necessary for representation, and the development of procedures for following-through on the development of ontologies. The practice of ontologies itself involves many of the modalities of knowledge that we have described in this section. We have characterized the ability to transform technical domain knowledges into a discursive form accessible to knowledge acquisition as a skill which can only be learned in doing. Even though many of the geo-science participants in GEON have dedicated their lives to education, university pedagogy has a distinctly different character than the requirements of a knowledge to be formalized for system accessibility. Similarly

while geo-scientists are adept at discussing the state of the field, to disagree on findings amongst themselves, and to bring new evidence to bear on current understandings, KR requires qualitatively different expression of knowledge. This too is a learned skill. Our future research will focus with finer granularity on the activity of negotiating knowledge for representation. Finally, the involvement and length of time necessary for building ontologies requires procedures for initiation, follow-up, and revision; in GEON's case organization for distributed work was also necessary.

The Learning Trajectory – Community Enrolment

The final step in the learning trajectory involves enrolling a broader community in the use of ontology. In a project such as GEON, ontologies are constructed not only to serve the research of participating scientists but as a community resource, an infrastructure. There is always the possibility that within a community of science – where knowledge is seminal and research may be hotly contested – that an ontology may receive dissent from the group as a whole or from sub-disciplinary groups. Because expert communities are considered the arbiters of domain truths, the front stage work of formalizing knowledge, forming temporary pragmatic consensuses, or representing uncertainties must be coupled with the backstage work of securing consent, building alliances, and holding standards in place within a community. As development continues the practice of building ontologies and securing community consent progressively begin to overlap. In forming a working consensus on debated geo-scientific knowledge GEON participants had to think not only of the scientists present, but also of a larger community of future geo-scientific users of the ontology. The properties of knowledge we listed in the excerpt above are precisely what makes the movement of ontologies back to a community such a potentially explosive undertaking. The possibility of a contentious ontology should perhaps itself be sufficient motivation to take the development process extremely seriously; of course, the most likely outcome of a contentious

ontology is its being thoroughly ignored by a community and thus becoming high-investment 'vapourware'. In this sense the domain community becomes seminal in defining the success of an ontology endeavour.

In an analysis of knowledge representation we must be cautious not to hypostatize *a community*, for in many senses this entity is both the object and outcome of the emerging methods for enrolling community. In building information systems for GEON the term 'community' has had at least two general uses: i) community has referred to an already existing but amorphous body of the domain, but; ii) community has also acted as the identification of a future body of users who will be linked by interdisciplinary ties and the computing resources of GEON. The two meanings of community are used somewhat interchangeably, and reflect the mandates and purposes of the GEON project. As 'cyberinfrastructure for the geosciences' GEON is simultaneously serving a constituency of U.S. earth researchers, but is also a community building endeavour that seeks to bring these scientists together. The category 'geo-scientists' includes thousands of academic and industrial researchers distributed in universities and research institutes across the US, in disciplines as varied as seismology to archaeology. It is difficult to reconcile this use of 'community' with traditional understandings where community is characterized "as a network of social relations marked by mutuality and emotional bonds" in which a "sense of self and community may be hard to distinguish" and where "a community is an end in itself," (Bender 1978 p.8). Geo-scientists only rarely know each other directly, most often do not interact at all, and in fact are usually only tied by loosely framed interest or by common organizational or funding umbrellas (*e.g. USGS or NSF*). In this first usage of the term community, GEON is referring to its mandated constituency – the broader geo-science collective as a beneficiary of its computing resources. Meanwhile there is another usage of the term which comes closer to what we understand as community building or outreach. We can think of that small selection of geoscientists which come together to build an ontology as

a *sample* of the larger geo-science community, the work of community enrolment is partially that of making this sample representative of the larger body. In the practice of ontology building, domain practitioners come to see their work as not only an investment of their own time, but also as a knowledge commitment. In building a computable semantic map of geoscience knowledge, individual domain scientists are standing-in as representatives of a larger epistemic community, for this 'representativeness' to hold the individual geoscientists must ensure that a larger collective will stand behind their knowledge representations.

As a by-product of the 'learning by doing' of ontology work, domain scientists begin to consider methods for building community consent, inviting participation and gathering a user-base. Because ontologies remain a relatively novel technology, the act of informing a community brings GEON participants back almost full circle to the first step in the learning trajectory; the broader geo-science community must be educated about a novel technology and of the epistemic commitments made in their name:

This is the opportunity to reach out to the geological community as a whole, whether or not they may have a view, even if they somehow, even randomly, bump into this, and say 'oh this is interesting, I didn't know this is what was going on' -- and this all comes back to accessing What's going on in as easy a way as possible, and not, say, the only way to get to these [ontologies] is go to the GEON portal, and then sign on, to do this and then press this button. And that's the challenge, just how to lay this stuff on, this is what is going on, that this is under development.

This selection by a geoscientist speaks to the dual difficulty of simultaneously communicating the existence of ontology technologies and making them accessible for community participation. Ontology development is a going concern in the worlds of business, science, government (Fountain 2001) and in the public sphere (Berners-Lee, Handler et al. 2001). Over time the stakes will become clearer as the technology itself becomes more familiar, but in these early stages of science-application GEON is encountering a generally uninformed constituency and few methods for making their work accessible.

A period of 'domestication' is not uncommon following the introduction of new technology. For example, common sense would tell us that following the introduction of the

printing press in the 15th Century that we would see an decrease in the number of copy errors over the previous tradition of hand copied manuscripts. After all, the printing press reproduces mechanically, while hand written texts are open to human error. However, a close inspection of early printed texts shows a distinct increase in copy-error. Historian of technology Elizabeth Eisenstein (1983) argues that over centuries of copying manuscripts by hand, a tradition and repertoire of procedure had been derived for ensuring accuracy. The introduction of the printing press, with its shift in locations of production and individuals participating, dislocated many of these correction practices, and it was only over time that new methods could emerge for the new technological order.

Today ontology visualization remains in the early stages, and conventions of representation have not yet become established. As the informant above notes, logging-in to the GEON portal, navigating to the appropriate ontology representation and taking the time to understand it, are significant barriers to participation. Enrolling of participants and enrolling of community can be distinguished by the level of possible engagement – participants will experience learning by doing, the practice of ontology work, thus coming to understand first hand the kinds of difficulties of specifying knowledge as code – meanwhile enrolling the community, for the moment, remains action at a distance. While a plethora of software solutions are under development for communicating ontologies, none have completed a full development arc and become accessible and familiar to non-expert users.

Over time, domain scientists and the IT team in GEON have begun to put together alternate methods of outreach, based on previous experience or by analogy:

sometimes, I mean, I don't have that experience, some people who are in the standardization of programming languages, proposed changes for Fortran. You have some idea, you propose it in the [list serve] forum, everyone jumps on it, chews it up, you respond. Then eventually it comes up, sometimes exactly as it was proposed, but now everyone is saying 'oh you should have done that at the very beginning now its right, now its good'. This [ontology] can be something that geologists debate, contribute, and then they feel better, and they may feel more comfortable accepting those ontologies, if its not something that we, usurpers [...] are dumping onto them. But if its something that they are discussing themselves, it can only help.

While ontologies are often presented as an alternative to standardization, experienced practitioners quickly began to see procedural similarities. In standard building, top-down approaches can be softened by establishing multiple mechanisms of community involvement, participation, and feedback. In the selection above, changing code for Fortran is compared to building ontologies: while a joke is being made that even following a period of initial dissent by the community the final code may resemble the initial code, the informant is expressing a need to produce forums for discussion rather than having an ontology appear *deus ex machina*. This is hardly a sophisticated strategy for intervention, however it is only over time that detailed methods can emerge. In the same day as the previous quote another geoscientist proposed to widen the scope of concept-space workshop for ontology development:

so there is this specific action item, which is, once you prepare the report, when you want to come down and do the formal representation? Or rather I should say the specific follow up. But [KR specialist] is going to be here tonight... and I was talking to him and one thing we want to do is, also for all the ontology work going on so far, I think we need a system for follow up. Because ontologies are not just one time things, obviously, so as a group how do we follow up? Is it subsequent meetings, *clearly these can be exposed on the website, but then we need to get this real engagement going, we might even bring in some representatives from the community. Like I said, put it up there, let the community look at it. We need some process.* [emphasis added]

It was following this point that the 'concept-space workshops' became the 'community-based ontology development.' This is not a purely a rhetorical move, rather we point to a processual development over time in which representing the community becomes significant to the ontology building activity. Later in the same year the first geo-ontology workshop was held as the SDSC, backed by GEON, but directed at a much broader collection of computer and geo -scientists than previous such efforts.

Apart from community outreach another method of GEON's ontology community building efforts could be characterized as an institutionalization of the procedures of ontology, and the formalization of relationships. In the larger GEON project an alignment with already existing institutions is an ongoing effort: maintaining ties with the primary earth science research institutions, leading publications , 'sister' cyberinfrastructure projects and

other geoinformatic endeavours. Similarly within ontology work a parallel alignment must be maintained: following the leading trends in the stabilization of ontology languages, technique and applications; but also with ongoing KR projects in the earth sciences. For example GEON's ontology efforts have always drawn on already established work within the Canadian Geological Survey or the British Rock Classification, making these available as options within the configuration of its search or data-registration engines. While the geoscientists of GEON have already led efforts to publish special editions of journals and books on issues such as geo-informatics, more recently they have begun looking at a geo-ontology edition or book publication. Finally, geoscientists have considered the creation of geo-ontology 'facilities,' which would mirror the already existing data repositories and standard bodies in the field, and thus provide points of contact for the community, means for arbitrating the construction of new ontologies, the registration of data and perhaps the possibility of upper-level ontologies. This latest trajectory of action – formalizing relations through institutionalization or publication – remains novel within GEON. Few practical strategies have yet to emerge; however it is clear that these formalizations will involve building institutions beyond the initial scope of the GEON project, and will require stronger alliances with the established regulatory bodies of geo-science.

This section cannot serve as a 'how-to guide' for community enrolment, or make any claims to providing a comprehensive list of possible methods. We are in the midst of a domestication of the technologies of interoperability, ontology one amongst those. The activities described in this section will be a continuing site for sociological investigation. Today the methods and techniques for identifying and acting on a domain community are emerging, mirroring the technical evolution of ontology itself. Rather, this section points to the kind of learning which has characterized ontology development. Beyond constructing an ontology, but also beyond the first two facets of the learning trajectory -- understanding 'what

is an ontology' and 'how to ontology' -- both IT and domain participants must enact (Fountain 2001) the implementation of the technology.

We can characterize the entire arc of ontology development as an hourglass shaped process (see fig.1): beginning as a broad amorphous community – e.g. 'geoscientists' – domain experts are selected to represent this community as a whole, or task-specific portions thereof; at the tight neck of the hourglass this small selection of individuals must come to states of agreement and gain the discursive capacity necessary for knowledge representation, and; then, in turn, the domain scientists must broaden the links built by ontologies back to the community. Enacting the specific methods for intervening in community – e.g. open workshops, forums,

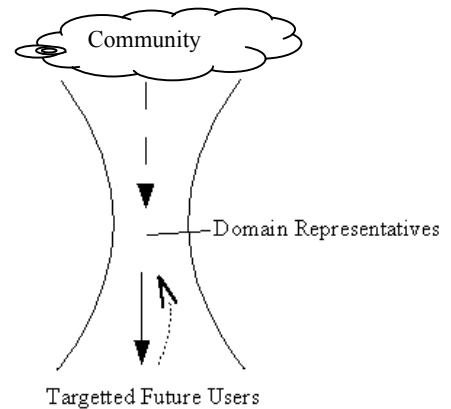


Fig.1

publications, standard bodies &c. -- leads to a specification of that community as particular groups are targeted. Enrolling the community is a movement from a broadly defined domain to an identified body of future ontology users (Woolgar 1991; Mackay, Carne et al. 2000). In identification the meaning of community changes, or doubles, referring both to the amorphous entity and to particular relations with groups. These specific targeted members of the domain may act on the ontology development itself; depending on the particular mechanism of community enrolment they may become participants themselves.

In summary, the third step of the learning trajectory – enrolling community – includes a growing awareness of the commitment to a larger knowledge community, of developing particular techniques for the identification of a relevant group of future users, and the development of particular techniques for engaging with that user community. In the acts of engagement, such as workshops, an amorphous domain community is transformed as specific relationships are formed which link people and organizations through ontology building or use. Technologies, and in particular novel technologies, do not simply diffuse into usage, nor

does the illocutionary force of a 'best ' or 'most efficient' method lead from invention to innovation. Innovation – which we define as the progressive domestication of a technology and the work of its uptake in specific implementations – of ontology has required the simultaneous invention of technique for education, dissemination and application.

Conclusion

Many 'social' studies of knowledge representation suggest that the direct participation of sociologists in acquisition could facilitate the process (Bowker, Star et al. 1997). This is not our conclusion, we think that knowledge representation endeavours are always already a kind of *sociological work* (Latour 1996). In researching ontology building in GEON we have observed an increasing sophistication of methods for pedagogy, organizing acquisition and invoking the participation a broader community. Computer science has always had as part of its output the goal of producing tools – e.g computing power, visualization, data management and storage. In this sense ontologies are no different: they are a tool for the user to execute sophisticated searches and queries. But building ontologies requires something that few applications have required before: they require the programmer to have a detailed understanding of the knowledge to be represented and they require the domain experts to communicate their knowledge in a computable format. Ontology builders cannot sit on their neat and tidy side of the 'design wall' and throw over a finished product to the user - an intense pedagogic relationship is required between IT and domain. The epistemic commitments of ontologies, and their purpose as community resource, means not only a close knowledge-based collaboration between domain and IT, but also negotiating a relationship with a broader domain community.

Interdisciplinary work always involves investments in translating between domain languages, practices and organization, but the shared space built is often ephemeral and rarely officially encoded. With ontologies semantic relations and equivalencies become formal and

executable, and as infrastructure are intended to serve as long term investment.

Anthropologist Mary Douglas notes: “Nothing else but institutions can define sameness. Similarity is an institution,” (Douglas 1986, p.55). The front stage work of local agreements must be coupled with backstage work of securing a larger consent from a community of practice and its host institutions. Sociologist Bruno Latour reminds us that there are two simultaneous meanings for translation, and that we should maintain both in mind in any such act (Latour 1987). Translation, in its more linguistic sense can refer to the production of an equivalence, 'sameness' – equal meanings across languages for example. In its physical sense translation refers to a to a movement, to a displacement. These understandings of translation are particularly apt for building ontologies, as a translation of a form of knowledge from *wild* to formal. Without sufficient caution, through the double action of translation the various modalities of knowledge (tacit, embodied, and so forth) can be buried in the *bloodless* languages of description logic.

At the writing of this paper GEON has been an active project for almost three years. The various phases of the learning trajectory have been slow to develop; for example, a sophisticated dialogue of investment and the methods for engaging the community have only emerged in the last year. It has taken substantial time to come to collective understanding within the project of procedures and practicalities for problem definition, knowledge acquisition, and community building. Because GEON is a large and ambitious project with various goals outside knowledge representation, the efforts of participants are not exclusively directed at ontology building. Furthermore the distributed nature of the organization also presents its own complications (Olson and Olson 2000). A co-located team of dedicated domain and IT experts could perhaps expedite the process of ontology building, but we believe that the conceptual and methodological maturity described in the learning trajectory must be traversed even in more concentrated efforts and furthermore that there is a moderation of pace required for pedagogy and community acceptance. It is our argument in

this paper that the learning trajectory – enrolment of participants, learning by practice, and enrolment of the community – requires a certain practical breadth of scope which has usually not been considered by narrower technically oriented descriptions of ontology work.

<i>Enrolling Participants</i>	<i>The Practice of Ontology</i>	<i>Enrolling Community</i>
<ul style="list-style-type: none"> • the problematic of interoperability and the solution of ontologies • introduction to the data politics of the domain • division of labor – IT encoding and domain content 	<ul style="list-style-type: none"> • how to bring forth knowledge into discourse • resolving difference (granularity, pragmatics, rain-check, represent uncertainty) • procedural development 	<ul style="list-style-type: none"> • the commitment of ontology ('hourglass structure') • identifying community • develop methods of outreach or community building

As we have demonstrated through our development of the concept of the learning trajectory, this fitting process is integrally organizational, practical and conceptual. Formal computer-based ontologies are only as good as the knowledge that they encapsulate. The attempt to produce a standardized ontology – whether this be in the overweening encyclopaedism of Lenat’s cyc project, which attempts to completely categorize all the objects in the world (<http://www.cyc.com/>) or through a more grounded project like GEON – will always falter in the long term unless they recognize the nature of this fitting. The central practical point here is that scientific work is in many ways *about* ontology change; and building sharable ontologies between disciplines is *about* adjudicating the relative knowledge claims of each discipline.

Works Cited

Arzberger, P., P. Schroeder, et al. (2004). "An International Framework to Promote Access to Data." *Science* **303**: 1777-8.

Atkins, D. E. C. (2003). Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, National Science Foundation.

Bender, T. (1978). *Community and Social Change in America*, Rutgers University Press.

Berners-Lee, T., J. Handler, et al. (2001). "The Semantic Web." *Scientific American* **May**.

Boose, J. H. (1989). "A survey of knowledge acquisition techniques and tools." *Knowledge Acquisition* **1**(1): 3-37.

- Bowker, G. C. (1988). Pictures from the subsoil, 1939. Picturing Power: Visual Depiction and Social Relations. G. Fyfe and J. Law. London, Routledge.
- Bowker, G. C. (2000). "Biodiversity Datadiversity." Social Studies of Science **30**(5): 643-683.
- Bowker, G. C. and S. L. Star (1999). Sorting things out: classification and its consequences. Cambridge, Mass., MIT Press.
- Bowker, G. C., S. L. Star, et al., Eds. (1997). Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide. Mahwah, NJ, Lawrence Erlbaum Associates.
- Burchfield, J. D. (1990). Lord Kelvin and the age of the Earth. Chicago, University of Chicago Press.
- Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fisherman. Power, Action and Belief. J. Law and P. Kegan, Routledge.
- Cambrosio, A. and P. Keating (1988). "'Going Monoclonal': Art, Science, and Magic in the Day-to-Day use of Hybridoma Technology." Social Problems **35**(3): 244-260.
- Collins, H. M. (1985). Changing Order: replications and induction in scientific practice. London, Sage Publications.
- Collins, H. M. (1990). Artificial Experts: Social Knowledge and Intelligent Machines. Cambridge, Massachusetts, The MIT Press.
- Douglas, M. (1986). How Institutions Think. Syracuse, Syracuse University Press.
- Edwards, P. (1999). "Global Climate Science, Uncertainty and Politics: Data-Laden Models, Model-Filtered Data." Science as Culture **8**(4): 437-72.
- Eisenstein, E. L. (1983). The Printing Revolution in Early Modern Europe. Cambridge, England, Cambridge University Press.
- Epstein, S. (1996). Impure Science: AIDS, Activism and Politics of Knowledge, University of California Press.
- Forrester, J. (1996). "If p, then what? Thinking in cases." History of the Human Sciences **9**(3): 1-25.
- Forsythe, D. E. and B. G. Buchanan (1989). "Knowledge Acquisition for Expert Systems: Some Pitfalls and Suggestions." IEEE Transactions on Systems, Man, and Cybernetics **19**: 435-442.
- Fountain, J. E. (2001). Building the Virtual State: Information Technology and Institutional Change. Washington, D.C., Brookings Institution Press.
- Gaines, B. R. (1989). "Social and cognitive processes in knowledge acquisition." Knowledge Acquisition **1**(1): 39-58.
- Galison, P. L. (1997). Image and logic: a material culture of microphysics. Chicago, University of Chicago Press.
- Glaser, B. G. (1978). Theoretical sensitivity: advances in the methodology of grounded theory. Mill Valley, CA, Sociology Press.
- Glaser, B. G. and A. Strauss (1973). The discovery of grounded theory: strategies for qualitative research. Chicago, Aldine Pub. Co.
- Gruber, T. R. (1993). "A Translation Approach to Portale Ontology Specification." Knowledge Acquisition **5**: 199-220.
- Guarino, N. and P. Giarretta (1995). Ontologies and Knowledge Bases: towards a terminological clarificatin. Towards Very Large Knowledge Bases; Knowledge Building and Knowledge Sharing. R. Dieng and O. Corby. Amsterdam, IOS Press: 25-32.
- Guarino, N. and C. Welty (2000). A Formal Ontology of Properties. Knowledge Engineering and Knowledge Managemtn: Method, Models and Tools. 12th International Conference (EKAW 2000). R. Dieng and O. Corby. Berlin/New York, Springer: 97-113.

- Hacking, I. (1990). The Taming of Chance. Cambridge, UK, Cambridge University Press.
- Hutchins, E. (1995). Cognition in the Wild. Massachusetts, MIT Press.
- Knorr-Cetina, K. (1999). Epistemic Cultures: How the sciences make knowledge. Cambridge, MA, Harvard University Press.
- Kuhn, T. S. (1977). Second Thoughts on Paradigms. The Essential Tension. Selected Studies in Scientific Tradition and Change. T. S. Kuhn. Chicago, IL/London, Chicago University Press.
- Lakatos, I. (1976). Proofs and refutations: the logic of mathematical discovery. Cambridge, MA, Cambridge University Press.
- Latour, B. (1987). Science in action: how to follow scientists and engineers through society. Cambridge, Mass., Harvard University Press.
- Latour, B. (1988). The pasteurization of France. Cambridge, Mass., Harvard University Press.
- Latour, B. (1996). Aramis, or, The love of technology. Cambridge, Mass., Harvard University Press.
- Lave, J. (1988). Cognition in Practice: Mind, mathematics and culture in everyday life. Cambridge, Cambridge University Press.
- Livingston, E. (1986). The ethnomethodological foundations of mathematics. London, Routledge & Kegan Paul.
- Mackay, H., C. Carne, et al. (2000). "Reconfiguring the User: Using Rapid Application Development." Social Studies of Science **30**(5): 737-57.
- Meyer, A. M. (1992). "How to Apply the Anthropological Technique of Participant Observation to Knowledge Acquisition for Expert Systems." IEEE Transactions on Systems, Man, and Cybernetics **22**(5): 983-990.
- Meyer, M. A. and J. M. Booker (2001). Eliciting and Analyzing Expert Judgement: A Practical Guide. Philadelphia, PA, Society of Industrial and Applied Mathematics.
- Olson, G. M. and J. S. Olson (2000). "Distance Matters." Human-Computer Interaction **15**(2-3): 139-179.
- Orr, J. E. (1996). Talking About Machines: An Ethnography of a Modern Job. Ithaca, Cornell University Press.
- Oudshoorn, N. and T. J. Pinch (2003). Introduction: How Users and Non-Users Matter. How Users Matter. N. Oudshoorn and T. J. Pinch. New Baskerville, The MIT Press.
- Polanyi, M. (1964). Personal Knowledge: Towards a Post-Critical Philosophy. New York, Harper Press.
- Rheinberger, H.-J. (2000). Cystoplasmic Particles: The Trajectory of a Scientific Object. Biographies of Scientific Objects. L. Daston. Chicago, University of Chicago Press.
- Rudwick, M. (1976). "The Emergence of a Visual Language for Geological Science, 1760-1840." History of Science **14**: 149-95.
- Schuler, D. and A. Namioka (1993). Participatory Design: Principles and Practices. New Jersey, Lawrence Erlbaum Associates.
- Scott, W. R. (1992). Organizations: Rational, Natural and Open. Englewood Cliffs, NJ, Prentice-Hall.
- Shapin, S. (1989). "The Invisible Technician." American Scientist **77**(Nov/Dec): 554-563.
- Slaughter, M. M. (1982). Universal Languages and Scientific Taxonomy in the Seventeenth Century. Cambridge, UK, Cambridge University Press.
- Smith, B. (2003). Ontology. Blackwell Guide to the Philosophy of Computing and Information. L. Floridi. Oxford, Blackwell: 155-166.
- Smith, B. and C. Welty (2001). "Ontology: Towards a New Synthesis." Published in the Proceedings of FOIS '01.
- Star, S. L. (1995). Ecologies of Knowledge: Work and Politics in Science and Technology. Albany, New York, SUNY Press.

- Star, S. L. and K. Ruhleder. (1994). Steps Towards and Ecology of Infrastructure: Complex Problems in Design and Access for Large-Scale Collaborative Systems. NC, USA, Chapel Hill.
- Stefik, M. and L. Conway (1982). "Towards a Principled Engineering of Knowledge." AI Magazine **3**(3): 4-16.
- Strauss, A. (1993). Continual Permutations of Action. New York, Aldine de Gruyter.
- Waterman, D. A. (1986). A Guide to Expert Systems. Reading, MA, Addison-Wesley.
- Winkler, M. G. and A. Van Helden (1992). "Representing the Heavens: Galileo and Visual Astronomy." Isis **83**: 195-217.
- Woolgar, S. (1991). Configuring the user: The case of usability trials. A Sociology of Monsters. J. Law, Routledge.