Information and Organization xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Information and Organization

journal homepage: www.elsevier.com/locate/infoandorg

Between meaning and machine: Learning to represent the knowledge of communities

David Ribes [a,*], Geoffrey C. Bowker [b]

[a] Georgetown University, Communication, Culture and Technology, 3520 Prospect St. NW, Suite 311, Washington, DC 20057, United States
[b] Santa Clara University, Science Technology and Society, 500 El Camino Real, Santa Clara, CA 95053, United States

ARTICLE INFO

Article history:
Received 4 June 2008
Received in revised form 25 March 2009
Accepted 20 April 2009
Available online xxxx

Keywords:
Ontology development
Routine
Semantic interoperability
Sociology of knowledge representation
Ethnography

ABSTRACT

Representing knowledge in codified forms is transformative of ones orientation to that knowledge. We trace the emergence of a routine for knowledge acquisition and its consequences for participants. Over time, participants in the earth science project GEON, first learned about ontologies and then learned how to create them. We identify three steps in the routine: understanding the problematic of interoperability; learning the practice of knowledge acquisition; and engaging the broader community. As participants traversed the routine they came to articulate, and then represent, the knowledge of their communities. In a process we call reapprehension, traversing the routine also transformed participants' orientation towards their data, knowledge and community, making them more keenly aware of the informational aspects of their fields.

1. Introduction

Ontologies are an information technology for representing specialized knowledge in order to facilitate communication across disciplines, share data or enable collaboration. In a nutshell, they describe the sets of entities that make up the world-in-a-computer, and circumscribe the sets of relationships they can have with each other. They are a complex and ambitious technical approach to address the problem of diverse languages, heterogeneous categorizations and varied methods for organizing information. In the wake of ontologies the information of a domain is substantially reorganized, facilitating data exchange and reuse. These are the *goals* for ontologies. Their *development* is a practical and organizational achievement, and the topic of this paper. We focus on the practical processes surrounding

* Corresponding author. Tel.: +1 202 687 4831.
E-mail addresses: dr273@georgetown.edu (D. Ribes), gbowker@scu.edu (G.C. Bowker).

1471-7727/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.infoandorg.2009.04.001

the design and deployment of ontologies within the infrastructure project GEON (the geosciences network) and trace the emergence of an organizational *routine* for their production. This routinization mirrored the learning trajectory of participants as they came to understand what is *at stake* in ontologies, that is, that they were representing the knowledge of their communities. As they traversed the routine, participants' experience of expert knowledge and of their communities were transformed in a process we call *reapprehension:* an increased orientation to the informational organization of ones field e.g., data, databases, encoded knowledge and their capacity to "flow" (or interoperate) across technological, disciplinary and institutional divides.

Organizational routines are "repeated patterns of interdependent actions, performed by multiple actors," (Feldman, 2000; Feldman & Pentland, 2003). They serve as a resource, a malleable and locally adapted recipe or template for how to go about a task. While a routine must always be practically enacted, having no existence outside its performance, it also becomes embedded in the configuration of material resources that enable practical work (Jordan & Lynch, 1998). We will see that the activities we trace and call the *acquisition routine* rendered the complex and uncertain activities of knowledge representation into an outline of "steps," reducing the work of reinvention on each new occasion of ontology building. However, traversing the routine also changed the experience participants had of their data archive, knowledge and community. In particular we focus on the emergence of a structured concern for ensuring the representativeness of ontologies: the practical methods for creating representations which *stood in for* and *were used by* a larger knowledge community.

Ontologies are intended to serve a community, making accessible data and resources for its members; as such they are a form of *infrastructure* (Star & Ruhleder, 1994). Participants in GEON quickly realized that they comprised a small subsection of the geoscience community (i.e., scores of participants in a community often cited to be in the thousands). Without the work of making ontologies representative of their domain community – of generating venues for feedback and for participation – their ontologies would be open to contestation, or, more likely, be ignored and remain unused. However, modeling ontologies involved articulating knowledge in ways that appeared alien to that domain community. For ontologies to appear representative, the community itself would have to learn the goals and language of knowledge modeling.

The phenomenologically transformative consequences of learning and traversing routines are almost completely undiscussed in the literature. In our case, participants' orientation to information in their discipline was changed by traversing the routine. We name this reapprehension, and emphasize the practice and material tools that accompany the reworking of, for instance, knowledge in informational terms. Participants came to learn: (i) the purposes and goals of ontology, what we call the problematic of interoperability; (ii) how to articulate their knowledge in forms amenable to formal representation, and (iii) how a broader community's interests are at stake in this process, and what activities would be necessary to engage and enrol that community in the use of ontologies. In order to do so participants had to rearticulate their knowledge in forms amenable to formal modeling, and also encourage their colleagues in the ontologies' use, maintenance and upkeep. A keener awareness of the informational aspects of their fields changed the orientation of participants to their own data and knowledge; it also entailed redirecting more time and resources to their integration and maintenance.

Following a discussion of case, method and an outline of knowledge capture we trace each of the three steps of the routine. In Section 6 we return to how, by traversing the routine, "knowledge" and "community" took on new meaning, as they were rearticulated in the language of logic and information as predicates and users,[1] respectively.

1.1. Case and method

GEON, the GEOscience Network, is a cyberinfrastructure project (Atkins, 2003) which sought to produce a repertoire of high-end information technologies for the broader earth sciences:

[1] There is a close relationship between acquisition and user studies or requirements analysis. In both cases people are recast as users of future systems, the object of studies that make them known, so as to inform a process of technology design (Mackay, Carne, Benyon-Davies, & Tudhope, 2000; Woolgar, 1991). This topic is analyzed more extensively by Ribes and Finholt (2008) in which the authors explore the simultaneous constitution and knowing of a user community.

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

> The ultimate goal of GEON is to establish a new informatics-based paradigm in the geosciences, to provide a holistic understanding of the Earth's dynamic systems, thereby transforming the science (GEON Proposal: 3).

The project was funded by the National Science Foundation (NSF) with the goal of producing an "umbrella infrastructure" bringing together heterogeneous earth science disciplines. It drew together a wide range of earth and computer science experts representing multiple institutions across the US. To name only a few, from the earth sciences GEON includes paleobotanists, metamorphic petrologists and geophysicists; from computer science: database specialists, grid developers and knowledge representation experts (Ribes & Bowker, 2008). The hope was to build tools *specific* to earth science research but *general* enough to support work in the various specialties of that community. Ontologies were part of the GEON's "knowledge mediation solution," facilitating communication and collaboration across disciplinary difference.

Between 2002 and 2005 we conducted ethnographic research, attending the meetings, workshops and conferences organized by GEON members. The routine we describe emerged as a driving concept of our research as we iteratively returned to the field (Star, 1999). We identified the steps of the routine through theoretical sampling (Glaser, 1978), as we focused on the iterations of ontology development. In total we observed the (partial) creation of four ontologies over three years. We conducted observation in the various sites of development work: face-to-face meetings, phone calls, video conferences, emails, and through Listservs – GEON participants provided generous access to all venues and the resulting data stream was quite rich. We supplemented observation with 12 interviews with computer and earth scientists. We have chosen to focus on a single project (GEON) to follow the *development arc* of a routine over a period of years. Even as we completed our research it was clear that the routine continued to evolve within the project. However, over that period it stabilized to a significant degree as participants became familiar with the method and as material arrangements came to support that work. On various occasions we have presented our preliminary research back to our respondents in GEON, and in turn they assisted us in refining our formulation of the routine.

As a methodological principle, we adopt an agnostic position towards the question "what is knowledge?" In this paper we take the language of knowledge representation at face value. That is, we will not question whether knowledge or meanings are *really* captured in ontologies (c.f., Forsythe, 1993). Rather, our focus is on the practical activities of participants as they went about the task of developing ontologies and shaping a routine to facilitate future development. 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 modeling in machine language. In contrast, our "sociology of knowledge representation" takes as its object the entire repertoire of action surrounding ontology development. Our method is not to identify a site of knowledge for acquisition, but instead to follow our informants across the entire range of heterogeneous activities which constitute knowledge work (Latour, 1987) – the kinds of work are reflected in the routine itself: learning a problematic, practical knowledge capture and community outreach.

2. Why capture knowledge?

A useful metaphor for understanding the *goals* of knowledge mediation and ontology is the concept of "gateway technologies." In their economic analysis of the American electrical network, David and Bunn (1988) describe how heterogeneous technologies were rendered into functional units through mediating linkages such as transformers and switches. Similarly, ontologies can be considered a kind of semantic gateway technology: they enable communication and coordination across heterogeneous disciplines, languages, categories or database schemas. This is a central goal of cyberinfrastructure ventures: enabling data and resources to move freely across the institutions of science.

Gateway technologies are "some means (a device, or a convention) for effectuating whatever technical connections between distinct production subsystems are required in order for them to be utilized in conjunction, within a larger integrated production system," (ibid.: 170). A gateway links between existing standards, schemas or formats, permitting these to work as though a single functional unit. A common example is the electric adaptor which permits you to plug in your laptop in the US, England or Continental Europe. Focusing on information technologies, Egyedi (2001) argues that gateway technologies add flexibility (Hanseth, Monteiro, & Hatling, 1996) permitting local adaptation and tailoring while maintaining compatibility with the larger system. They increase the ability of information to move across heterogeneous platforms, environments and institutions. Nominally, because standards can be maintained while still linking the entire system, gateway technologies are often touted as softening the politics of standardization (Bowker & Star, 1999). However, we need not take this claim at face value. While Egyedi has argued that gateway technologies increase flexibility, David and Bunn have shown that they can be substantive economic actors, reshaping market relations. Coalitions of standardized technologies can leave excluded participants in marginalized positions. Similarly, while ontologists in GEON have claimed that "we don't do standards" (a commonly repeated phrase), the question remains: whose knowledge and what databases are being interlinked through ontologies and which remain "silos" and "stovepipes"? This question is a topic of future research.

These are the goals of ontology. The goals are important in understanding what comes to be the orientation of participants: why do they invest their time and energy in capturing knowledge? This said, in this article we focus on the production of ontologies and how that process itself casts the free movement of data and knowledge as inherently valuable, thus transforming the orientation of participants and, potentially, the domain.

2.1. The language of ontologies

Ontology development is a relatively new technique within the field of computer science research (Gruber, 1993; Guarino & Welty, 2000), its application in the sciences and for data integration is even more novel. In GEON it was a central strategy for the production of data and resource interoperability across geoscience disciplines (Berkley et al., 2005). Ontologies are often described as producing *semantic interoperability*[2] because they represent knowledge in the form of meaning based relations, and link heterogeneous entities e.g., databases, resources, concepts and terminologies (Sheth, 1999). In an ontology, knowledge is represented or modeled in one of multiple computer readable languages, for example, the ontology web language (OWL) or resource description framework schema (RDFS) (Horrocks, Patel-Schneider, & van Harmelen, 2003).

In GEON, the experts of ontology were computer scientists. This meant that in addition to building useful tools for geoscientists, they had an interest in contributing to a body of computing knowledge[3]; these practitioners are called ontologists, knowledge engineers or knowledge modelers. Meaning, knowledge or conceptual relations represented in ontologies are said to be captured, acquired or elicited from domain sources such as human experts, scholarly publications or even textbooks (Forsythe & Buchanan, 1989). These terms carry slightly different connotations (Button & Sharrock, 1994), but are used relatively interchangeably by participants themselves.

Knowledge capture may be conducted through interactional means such as interviews, surveys or oral questionnaires (Boose, 1989; Meyer, 1992; Waterman, 1986); or by using semi-automated methods such as machine learning or semantic content analysis of journal articles (Maedche & Staab, 2001). The final products of the acquisition and development process are ontologies themselves, sometimes also called concept maps or knowledge bases.

Ontologies are conceptual definitions which are coded and interrelated using logical operators (description logic); however, participants often refer to ontologies as a technologies, as tools, or as software. In this article we will not problematize the language of ontologists, but rather use that language to follow and elaborate on the practices which they reference (see methods, ibid.).

Fig. 1 is an image used by GEON knowledge representation experts to convey the larger vision for data integration. On the bottom are the raw databases to be integrated, at the top is the user running a

[2] The other usually cited forms of interoperability are syntactic, schematic, or systemic, each of which addresses different aspects of data and resource heterogeneity, respectively: the structure of data entry, the structure of a database, or the underlying operating system.

[3] For studies on the tensions between science and development in technology projects see (Lawrence, 2006; Ribes & Bowker, 2008; Ribes & Finholt, 2007).

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

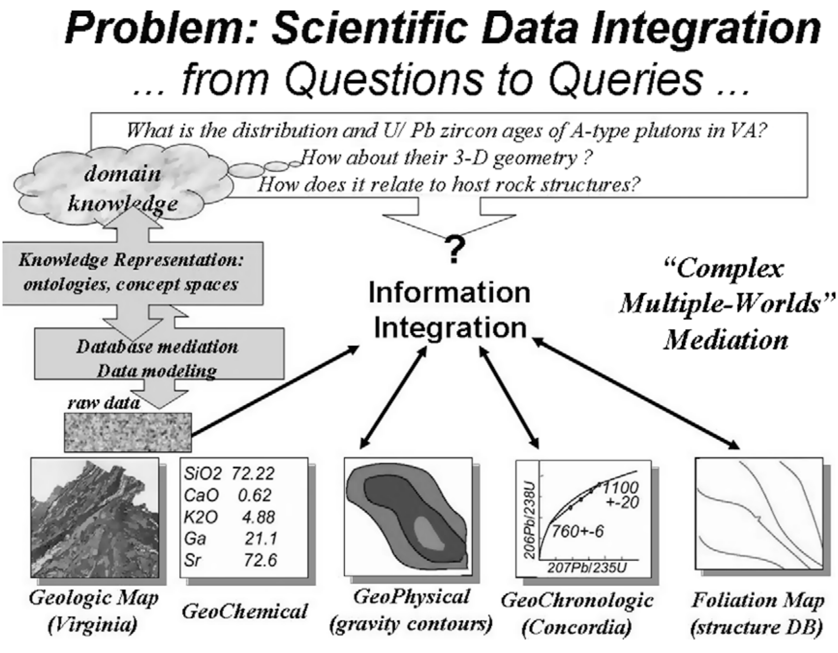

Fig. 1. GEON participants compose elaborate slide shows to share technical knowledge with the heterogeneous experts of GEON. This slide captures the notion of ontology based data integration.

query in domain specific technical terminology. The ontologies do the intermediation work for integration. By transparently integrating heterogenous data and resources, it is hoped that scientists will be able to address larger and more complex research questions than in the past, thus opening avenues for "revolutionizing" science. As we will see, such images serve a pedagogical role. They are the tools by which GEON's IT team convey the purpose of creating ontologies to earth scientists: why is the informational problem of heterogeneous concepts, data schemas and vocabularies something that is relevant to the study of the earth? We call this the problematic of interoperability.

3. Learning the problematic of interoperability

> "I've just learned how to say "ontology" and use it in a sentence!"
> Earth scientist

The first step in engaging participants in knowledge representation work was posing the problem which ontologies sought to remedy relative to the actual problems of the domain. While stovepipes, incompatibility, and legacy databases are the everyday headache of information specialists and the bread and butter of ontology efforts (Sheth, 1999), for geoscientists these problems have often been relegated to a little noticed substratum of technical work (Star, 1991). Even the principal investigators of GEON – a highly technically informed cadre – had rarely encountered their database architectures on such close terms. Conversely, computer scientists were unfamiliar with the terminologies and research agendas of the earth scientists. In this section we focus on the primary gathering of participants in the first months of the GEON project, known as the "kick-off," in November of 2002. This meeting was the initial attempt of computer scientists to introduce ontologies to earth-scientists and provided an excellent venue for observing the introductions to ontology Through this gathering, and then over time, participants articulated an understanding of the collective challenges facing the geosciences (and their data), and more specifically for "meditating the knowledge of the earth sciences" (field observation, knowledge engineer, November 2002).

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

3.1. Introducing themselves

Semantic representation and even data integration are fairly complex concepts, and ontologies are somewhat intangible pieces of software or code. The problems they seek to address and the solution they offer are not intuitively grasped in the same manner as, for example, data visualization or text search. Ideally, ontologies are infrastructure; that is, that to the extent that they work, they should be virtually invisible to the user. The software tools of ontology are not encountered in everyday work with computers (i.e., word processors, browsers or internet searches) nor have they usually been part of the core training of scientists. Introducing ontologies, then, required a fairly high order conceptual discussion about the heterogeneity of languages, the categories of scientific data collection, and the structure of databases. In turn, the knowledge and vocabularies of the domain are unfamiliar terrain to most computer and information scientists – especially when the domain was scientific or technically oriented. For example, plutons, accretion and fluvial sediments are the entities and processes which make up the language of geoscience.

Participants sought to create bidirectional learning environments: spaces for education that focused on transcending differences of method, vocabulary and even goals between the computer and earth science teams, and forming relationships for future collaboration. For example, at this meeting a geoscientist described the mechanisms for pluton formation – an igneous rock body formed at great depth and intruding from amongst other types – and a computer scientist described ontologies – an information technology for representing concepts as semantic relations. Each presentation was followed by a discussion period, allowing for participants to ask for clarification, or to draw links across their research. Such presentations were attempts to break down disciplinary boundaries, to learn of each others' research expertise and familiarize participants with accessible versions of specialist languages to facilitate sharing (Galison, 1999).

This meeting also served to begin coordinating research goals amongst earth scientists: what was the research of the diverse earth-scientists and how can individual research goals be aligned with those of the larger project? Learning was not unidirectional – while domain participants began to pick up an understanding of the problematic of ontology, the IT practitioners began to familiarize themselves with the "data-politics" of the domain (Bowker, 2000). Scientists have varying traditions for the curation and sharing of their data. They may feel possessive of their data hoping to draw out future insights, or they may feel ashamed of the quality and thus unwilling to share it with their peers (Borgman, 2007; Campbell et al., 2002; Ceci, 1988).

The extent to which a particular group is prepared to exchange their data varies substantially by discipline. For example, field scientists such as paleobotanists and metamorphic petrologists collect relatively small datasets at particular geographic sites. The intense personal involvement with the research site and with data collection may lead to unwillingness to contribute such data to a large anonymous repository. They may also feel that the data is incomprehensible or meaningless if not tied to local knowledge about a specific site.

On the other hand, instrument intensive scientists such as geophysicists have established traditions for using large arrays of remote instrumentation and the discipline has been at the advancing edge of computer science for forty years, from the first analog computers to the first expert systems (Bowker, 1994). Seismologists have a long tradition of sharing data across territories and across nations: it is in the nature of their data that it does not respect geographic boundaries. Over time geophysicists and seismologists have developed disciplinary norms that assume most data will eventually be shared (Shrum, Genuth, & Chompalov, 2007). For these scientists these varying traditions for data collection, curation and sharing can seem morally weighted – "the right thing to do". In deciding the policies for an "umbrella" data repository, at times these varying traditions became a site of explicit conflict. In creating the initial learning environments such conflicting views were given both the space for articulation and a venue for their resolution. The range of comfort with sharing their data is tempered by the "norms" of the community, and varies broadly within the geosciences. Understanding these "cultural differences"[4] is

[4] Cultural difference and norms an actors' category; participants adopted this language in order to make sense of the range of methods, vocabulary and data amongst themselves.

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

important for facilitating the process of ontology development which depends on scientists' willingness to register their data in publically available repositories.

3.2. Routine introductions

The initial phases of the routine educed the problematic of interoperability ("why integrate data") for domain members and the particular solution offered by ontologies. Geoscience participants were encouraged to see data as the property of a community (rather than as personal research material) and its interoperation as a valuable multidisciplinary resource.

These early encounters are presentations *about* ontology rather than the making of ontology. Learning ontology is a high level discussion, focused on producing a general conceptual familiarity across the disciplinary boundaries of computer science and the domain. In these initial discussions 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; this came later in the *practice* of capturing knowledge. 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. We discuss this in the next section.

This first step of the routine occurred at the founding meeting of the GEON project. Months later, as participants began to develop ontologies, they came to look back at this kickoff meeting as a crucial pedagogical moment: "it's at the kick-off that I first learned about data like a computer science person, and how to talk to my colleagues about data like that" (interview, geoscientist, March 2005). They realized that any participant in an ontology development project would require this basic level introduction: for computer scientists an introduction to domain language and concepts, and for geoscientists a primer on data, interoperability and ontology. In this sense, the initial kick-off that we describe in this section was not part of the routine we are outlining, *per se*. Rather, participants later identified the kick-off and its two-way educational presentations as a part of the routine by which they would regularly initiate the process of ontology development.

Over time, the step we have called learning the problematic of interoperability came to be routinized within GEON through the pedagogical presentations of computer scientists, and also in the meeting format which was set forth at the kick-off. For example, computer scientists created a set of elaborate slides for teaching geoscientists about ontology and knowledge representation (of which Fig. 1 remained a staple). These slides were used repeatedly in virtually all GEON presentations and by many diverse project participants. On each occasion that a new group of geoscientists sought to integrate their data by creating an ontology, the event began with a bidirectional forum where "science questions" were elaborated and knowledge mediation was explained.

Eventually it was the core GEON geoscientists themselves who began to explain ontology to their colleagues. They had come to adopt the problematic of interoperability as an endogenous concern for the geosciences: seeing their domain as disciplinarily balkanized, their datasets as incompatible and ontologies as one promising solution. Data are not a foreign consideration for scientists, but they are usually conceptualized as the raw material of research rather than as community resources which need to be interoperated. "Thinking in databases" generated a new orientation to data for participants.

These participants had gained sufficient familiarity with ontology and with the pedagogical presentations that they could take-up the task themselves. As one domain participant described the task: "these meetings are pretty straightforward now, we always get everyone talking. The computer scientists talk about ontology, we talk about our science, we go back and forth. This was harder at first, but I think we're getting the hang of it, and the [geoscience] community is also becoming more familiar with the idea of ontology," (interview, geoscientist, March 2004). As the routine became familiar to participants through repetition and facilitated by material support – e.g., a stock set of slides and a regularized meeting format – it was less of an effort to go through this pedagogical process.

4. The practice of knowledge capture

In the ensuing stage, which we call the practice of knowledge capture, a significant gap was revealed between abstract understandings of ontology (as in "learning" with the first step of the routine)

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

and a know-how for proceeding in the capture of knowledge. First, this section focuses on the hands-on learning in concept-space workshops as domain participants came to speak their knowledge in a form amenable to modeling. For participants, domain knowledge is somewhat naturalized and taken-for-granted (sometimes described as tacit within knowledge representation circles Sowa (2000)); it must be "made strange" for members to be able to reflect and articulate it in forms amenable to formal representation. Ontology languages require the formulation of knowledge in particular ways e.g., parsed into entities, operators and logical relations. We traced participants' activities as they enacted a division of labor between domain scientists and knowledge modelers. Nominally, domain experts are responsible for speaking their knowledge, while ontologists encode it: they sought to carefully demarcate the "form" of knowledge from its "content". We identify this as a form of *boundary work* (Gieryn, 1999).

Second, this section focuses on how participants learned a practical skill-set to engender working agreements in the face of disputes about community knowledge. As geoscientists began to build ontologies, what had passed as the tacit accord of experts broke down in attempts to render knowledge explicitly. Disagreements about the details of earth science had to be resolved in discussion. Over time a set of techniques emerged for resolution of disputes. These techniques were often suggested by computer scientists (rather than domain participants) who had previous experience with the practice of knowledge capture. The practice of ontology became routinized in a "recipe" that served to convey the skill-set of articulating knowledge in the language of code, and for resolving epistemic conflict.

4.1. Concept-space workshops

The GEON IT team initially called their particular method for knowledge acquisition concept-space workshops. During these workshops geoscientists and IT experts co-located for two to three days and sat in conference rooms "hammering out" formalized representations of geological knowledge. Because GEON is geographically distributed, knowledge capture was conducted in punctuated bursts of face-to-face interactions, followed by extensive revision via e-mail and video-conferencing.

This method for knowledge acquisition is in the lineage of expert elicitation (Meyer & 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 to be a *socially distributed* phenomenon, rather than localized in particular individuals, texts or archives. From this conceptualization of knowledge, it follows that the methods of acquisition must be tailored to the configuration of knowledge in a domain; open ended methods such as interviews or focus groups are used to triangulate community opinion. These methods involve a collective practice for knowledge acquisition rather than an individual interaction between "formal knowledge" (e.g. textbook) and modeler. It is particularly apt method for scientific knowledge, where the determination of truth itself is often a site for debate, controversy, and tension (Collins, 1981). This conceptualization of knowledge and knowledge capture was (initially) held primarily by the ontologists. As we will see, they organized the concept-space workshops to reflect this model.

Concept-space workshops in GEON became a location for both debate and for building working agreements. The IT team hoped that through these discussions geoscientists would come to an accord on knowledge and thus become spokespeople of community knowledge (Ribes & Finholt, 2008). As they went about modeling the domain, the ontologists sought to maintain a division of labor between those who speak their knowledge and those who captured it.

4.2. A division of labor around the knowledge of a domain

A driving understanding within knowledge representation circles is that participants can only be expected to use or contribute to the ontology if they see their worldview in it. An ontology must lay on the topology of scientific knowledge like a plaster cast, faithfully tracing dips and indentations, capturing a surface to support what is technically necessary for data interoperability or information navigation. Ontologists sought to enact a division of labor between scientists and modelers to ensure the ontology reflected domain understandings. The workshops themselves were structured to ensure that the responsibility of knowledge determination did not force a *de facto* shift onto the shoulders of the ontologists.

Domains carry with them epistemological traditions: core knowledge, ways of knowing, and criteria for specifying what is considered new knowledge. While in the received understanding of science the myth of a single scientific method abounded, more recent research in the history and sociology of science has uncovered a plethora of domain specific methodologies, trials for the establishment of knowledge, and great shifts over time in these methods and criteria (Galison & Stump, 1996). How each group *knows* is domain specific, tempered by its particular instrumentation (Clarke & Fujimura, 1992), history of representation (Lynch & Woolgar, 1990) and methods of communication (Knorr-Cetina, 1999).

In turn, because knowledge representation has a language of its own (often called predicate logic, although in fact, it is a specific breed of code, such as the Ontology Web Language (OWL)) the domain must learn to speak their knowledge in a language accessible to machine encoding. While the actual work of encoding is that of the ontologists, the knowledge must first be articulated in ways that can be parsed in the language of logic. Learning the practice of ontology is thus a fitting process between the epistemic conventions of the domain and the demands of formal modeling.

This is the basis for a boundary work (Gieryn, 1983; Ribes & Finholt, 2008) of epistemic granularity between information technology and its (geoscientific) content. Boundary work refers to the material arrangements and practical work of *participants themselves*, which sustain a demarcation or distinction of importance to members. In this case, participants sought to sustain the integrity of scientific knowledge: scientists wish to accurately convey their knowledge, and ontologists wish to faithfully capture that knowledge. Most importantly the ontologists had a strong orientation towards not affecting the content of knowledge, rather, they sought to represent it. Ontologists would regularly raise both hands and show empty sleeves, demonstrating no interest in intervening in the knowledge affairs of the domain they were representing. More than a rhetorical position, we found this to be a learned skill, enacted again and again in the discussions of concept-space workshops.

4.3. The 'form and content' of knowledge

A key practical skill in knowledge modeling is an enactment of the separation of form and content, i.e., semantic relationships and knowledge. 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 [of building an ontology], 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. (field research, computer scientist, October 2003)

The separation of form and content is mirrored in a division of labor between IT and domain experts. In this relationship, IT experts map the domain knowledge while sustaining distance from scientific knowledge itself. Since the act of naming objects in the world is invariably political in science, particular attention had to be paid to the actors' own terminology. In a sense, ontologists are very concerned with actors' categories.

In the excerpt below (in which the speaker is referring to the slide presented in Fig. 1, above) a computer scientist is addressing his ontology colleagues; the quote encapsulates the division of epistemic labor in ontology work between IT and domain:

> 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 plutons 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. (field research, computer scientist, October 2003)

In this extract the ontologist is first expressing a relative ignorance regarding the details of geochemistry or geophysics, while also marking-out some familiarity with the domain: he knows, to some extent, the nature of a pluton and how to read the nomenclature of chemistry. For ontology modelers, knowledge acquisition requires a comfort with the language of the domain, but specific details are to be left to experts. This ontologist was instructing his colleagues to leave aside the particular content of the scientific knowledge and to focus on the *relations* between concepts and the specific connections to database schemas: plutons and uranium-lead become predicates connected by a query which in turn must access particular values in at least two databases. While in practice, a clean division between specifying content and creating representation is difficult to maintain, this remains the normative goal of modelers. It is believed that to the extent that this division of labor fails in practice, the final ontology might itself be placed in question by domain members: "just how much does this ontology represent our knowledge?" (field research, geoscientist, February 2005).

4.4. Debating the content of knowledge

In concept-space workshops what had appeared to be a shared epistemic umbrella – "the geosciences" – began to break down into a finely grained mapping of disciplinary differences. It is this identification of difference within the knowledge domains, and its transformation into explicit discourse accessible to machine encoding that primarily characterized the learning-by-practice of ontology building. This revelation of difference also shifted participants' orientation to data, knowledge and their domain.

In the following quotation, a knowledge modeler encapsulated many of the initial encounters with such a practical learning:

> We've had people here from geochemistry, 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 [develop] ontologies, of what are the things they care about, what is important for them, what are [their] analytic methods, how do all these things work together, and how [they] can 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... (interview, computer scientist, December 2003)

The modeler neatly summarized many of the principles that guided action within GEON. As the speaker noted, the task of building ontologies began to reveal to scientists themselves their internally diverging knowledge commitments. To a great extent, this often came as a surprise to the domain participants: while geoscientists did take their various disciplines to be heterogeneous (and thus expected some divergence), they also had expected some agreement about "the basics" (field research, geoscientist, March 2004). However, knowledge capture sessions are specific forms of interactions that demand a higher order of specificity about technical details than most scientists are used to within multidisciplinary discussions. Sharing findings at a conference – even a highly multidisciplinary geoscience conference such as the Geological Society of America (GSA) – does not usually require an explicit discussion of, for example, textbook knowledge or representational conventions.

For example, in one concept-space workshop, a debate emerged over the particular usage of the color red in a map legend. This may seem trivial to those unfamiliar with the geosciences, however, within the room the discussion was quite heated. Across the geosciences there are various standardized legend color schemes, and members of subdisciplines may be quite used to particular representational conventions when reading visualizations. Some of these conventions are tied to state entities such as the United States Geological Survey (USGS), or to similar regulative bodies outside the US. Thus, in determining the spectral band of a particular shade of red, there was 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 have excluded another.

In data integration, exclusion or mismatched categories can lead to complications in queries, data visualization or representation, e.g., leaving out an important dataset during an ontology-enabled search. It could also mean excluding entire subdisciplines that have relied on particular conventions of representation. At conferences or other specialized meetings of geoscientists, discussions of such details can be set aside, however, when encoding ontologies, detailed granularity becomes crucial and momentary agreements can become programmed commitments.

Having never explicitly discussed such details, domain experts found it challenging to produce agreements at such fine scales of granularity. Modelers encouraged these discussions, however, so as to be able encode a conclusion in the ontology and facilitate the automated integration of data. For ontology development, tacitly held differences should be articulated for the purpose of overcoming them in a single unifying ontology: "if we want users to be able to find data, we need all possible logical semantic relations mapped out," (interview, computer scientist, Oct.2003). In this particular case, the dispute was resolved technically (see below): two representational schemas for color legends were included in the ontology, allowing users to choose their convention of preference.

However, if no agreement can be readily achieved, no clear definitions can be encoded (at least, not without the intervention of a computer scientist in domain knowledge). In the example below, drawn from a concept-space workshop in 2003, geoscientists were discussing their preferred labelling scheme for geologic time. While portions of the geological timeline have been standardized, within particular disciplinary communities important differences remain. The excerpt below was part of discussion over whether to use the terms "Proterozoic" or "Algonkian". These terms can refer approximately to the same period (2600-600 million years in the past) although on occasion of the terms also vary in meaning, e.g., many geologists have used the term Algonkian to refer to the late-Proterozoic period, and, most importantly, the terms are of differing importance within geoscientific institutions and subdisciplines. Put briefly, the United States Geological Survey uses the term Proterozoic while Algonkian is more commonly used within the subdisipline of geology[5]:

> Listen, we have a specific set of science questions we'd like to address here. Yes, you're right, the most common vocabulary in USGS is Proterozoic but to address these questions we [geologists] don't need their data. (field research, geologist, February 2004)

The discussion continued: one set of geoscientists argued that the ontology should be centered on the most common terminology within the broader geoscience community and its institutions (e.g., USGS; i.e., thus being relevant to the greatest majority of geoscientists), while the other group argued that definitions should mirror that of the subdiscipline driving the scientific research question, in that case, geology.

Given that the ontologists did not wish to intervene in scientific debates, but also required some consensus in the domain community to do their modeling work, this kind of controversy or contentiousness represented a significant derailment for building the ontology.[6] Once the conflict had been identified and had continued for several minutes, one modeler in the room availed himself to the domain by offering strategies for going on in the work of ontology development.

4.5. 'Going on': agreeing on the knowledge of the domain

In responding to controversies – the inability of domain scientists to agree on a particular semantic definition or relations between definitions – knowledge representation experts often suggested several possible solutions to be employed. While for geoscientists disagreement about method or findings are common, discussions of terminology are somewhat more esoteric, though certainly not altogether rare. In contrast, elaborating the concepts of a domain is the everyday work of knowledge modelers. Having engaged in multiple sessions of acquisition (within the geosciences and without) the ontologists were familiar within fine grained disagreements and could suggest modes of resolution. Below

[5] Geology refers specifically to the study of the history and structure of the earth's crust.

[6] In one case during a concept space workshop for hydrology that we did not attend, participants reported that the meeting broke down altogether as the present scientists were unable to agree and the discussion became heated.

we describe four strategies in forming working agreements that we have observed modelers suggesting in practice, we illustrate each relative to the example above of the geologic time-line:

(i) *Decrease granularity*: deal with the issue at a higher conceptual level, where the domain has established a stronger consensus. In the example above, one participant suggested leaving aside the specific labelling of time, and instead to rely on the broader term Precambrian which encapsulates both the Algonkian and Proterozoic periods.
(ii) *Working agreement*: encourage the domain experts to form a temporary working consensus in order to continue the ontology development process at hand. The quote above is an example: the speaker suggests a solution relative to the "science questions" geologists would like to address, rather than to a more abstract goal of serving the entire community of geoscientists.
(iii) *Rain-check*: leave the problem aside for a later time; experts may be able to resolve the issue with a re-examination of evidence, review of the literature, further consultation with experts or, in the long term, production of new evidence. In the case described above the conclusion of the discussion simply left the question of terms and periods "up in the air," and the concept-space workshop continued to other topics. To some extent this makes it difficult for the modelers to create drafts or prototype ontologies, however, the process of development is iterative and often takes months.
(iv) *Represent the uncertainty*: this is a technical solution within semantic modeling which affords encoding and representing disagreement, uncertainties, ambiguities or ambivalences. Within GEON this was a *promised* capability of ontologies; however, since it was still an experimental option, we never saw it employed in practice. We expect that expressing domain knowledge as a form of uncertainty is itself a practical skill – and one perpetually controversial amongst scientists.

The techniques for articulating knowledge and then resolving difference are a skill-set. Both the identification of differences, understanding their implications, and knowing what kinds of resolutions are possible was learned in practice. This skill was learned through the practice of ontology building itself: the communicative work between IT and domain – rather than through having read, for instance, ontology guides (c.f., Noy & McGuinness (unpublished manuscript)) which domain specialists often found alien, impenetrable and irrelevant.

In summary, the second step of the routine – learning acquisition and the practice of ontology building – refers the practical skills of bringing knowledge into discourse and of learning techniques to establish agreement necessary for representation. A division of labor was structured into the workshops themselves and enacted by participants: domain scientists sought to convey the details of their knowledge as ontologists encoded it. The activity was filled with conflict over the details of domain knowledge. However, the primary orientation of participants was to complete a working ontology rather than a coming to a definitive resolution.

In efforts to integrate heterogeneous knowledges (such as in a multidisplinary setting seeking to bring together the geosciences) it was not simply a matter of properly capturing knowledge but also a question of whose knowledge to capture. As we saw in the example above, deciding which temporal classification to encode also meant debating whether institutional agreement (e.g., USGS' use of Proterozoic) or scientific research agendas (e.g., geology's use of Algonkian) should drive ontology development. Representing community knowledge meant discussing "what community", "what institutions" and for "what purposes".

The ability to transform technical domain knowledge into a discursive form is a skill which was only learned in doing, and similarly, strategies for forming working agreements were also acquired. Learning these skills and engaging in these activities transformed participants' orientation towards knowledge. Even though many of the participants in this modeling activity dedicated their lives to education and research, university pedagogy and scientific inquiry have distinctly different qualities than formalizing knowledge for machine readable modeling. Similarly, while scientists may be adept at discussing the state of the field, disagreeing on findings amongst themselves, and bringing evidence to bear on current controversies, modeling required a qualitatively different expression of knowledge using the languages of predicate logic.

The practice of knowledge capture was routinized primarily in set of rhetorical skills of the participating computer scientists and a meeting format which structured a division of labor. A geoscientist described this routinized activity as a "recipe":

> Here is the recipe that we've currently applied with some success:
> One, you lock up scientists for 2-plus days; add some CS [computer science] 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; then, have other scientists visit us, 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. (field research, geoscientist, March 2005)

Ontologists rove from domain to domain, repeating knowledge capture workshops again and again (in GEON, working with seismologists, geophysicists and so on; outside GEON, working with biologists, ecologists and also non-scientists). Over time they learned a set of skills for articulating domain knowledge and resolving disputes which they could impart to domain participants. As we see in the excerpt, and we explore more thoroughly in the next section, part of this "recipe" included outreach to the community ("have other scientists visit us").

5. Engaging the community

We have described how participants in knowledge acquisition workshops evolved a set of techniques for educating newcomers to the problematic of interoperability, and the particular solution offered by ontologies; we have also traced as they learned how to speak in the language of knowledge representation and to resolve epistemic disagreements for the purpose of modeling. These practices became routinized through repeated enactment and by designing material resources which facilitated their repetition (such as meeting formats or slide-sets). While first two steps were oriented to those directly participating in ontology development, the third step in the acquisition routine was outwardly oriented to a broader group of future users: the domain community.

Because expert communities are usually considered the arbiters of domain truths (Epstein 1996), in ontology development the front stage work of formalizing knowledge, forming temporary pragmatic consensuses, or representing uncertainties, had to be coupled with the backstage work of securing consent, building alliances, and holding standards in place within a community. Community engagement had to be routinized, that is, made into an integral and iterative complement to the activities of knowledge modeling.

In GEON this activity was called community outreach, engagement, or contribution. Participants asked of themselves: "how could the community of geoscientists become involved in the process of knowledge capture?" In this section we trace the emergence of the third step, as participants sought to invent outreach techniques: engaging the community by using media such as a website/portal; soliciting participation by inviting geoscientists to ontology workshops, and; reshaping that community in face-to-face interactions at geoscience society meetings and through publication.

5.1. *The importance of community*

Ontologies themselves are not useful without a considerable investment on the part of its user community. They are a connective media, linking heterogeneous resources such as data and tools. In order to be useful for data integration, ontologies need the data which is held in the hands of practicing researchers and scientific institutions. If the community did not contribute its resources and time (e.g., data and updates on concepts or knowledge) the ontology would remain only a sophisticated knowledge representation without any resources from which to draw. Ontologies gain value as they obtain greater use by the community. Moreover, knowledge is not static. Ontologies will require continuous care if they are not to become outdated. As knowledge, terminology or concepts change within the scientific community, a once-accurate ontology could become obsolete:

> 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

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

> help science, and so they have to somehow match that [science's] mobility, rather than somehow holding it back by being too solid. (field research, geoscientist, March 2004)

Science is not only a repository of knowledge, it is also an evolving, contested and shifting set of propositions, language and methods. Without maintenance (an iterative process once again involving ontologists and knowledgeable domain members Graham & Thrift (2007)) even the most cautiously developed ontology may become useless over time.

In these senses – uptake, data contribution and maintenance – the larger participation of the domain community becomes seminal in defining the success of any ontology endeavour.

5.2. From representing knowledge to representing community

As they formed working agreements and debated geoscientific knowledge in concept space meetings, GEON participants had begun to think not only of the scientists present at a given workshop, but also of the larger community of future users for the ontology. While we describe the practice of ontology and engaging community as steps in a process of development, in practice, within the emerging routine these activities were largely co-temporal: the disagreements and conflicts which inevitably emerged in the practice of ontology alerted domain participants as to the necessity of gaining a broader consent than was possible from the people attending concept-space workshops.

Participants had come to recognize the importance of continuously returning to the broader community as they went about representing collectively held knowledge. In GEON's concept-space workshops they were developing a computable semantic map of geoscience knowledge. At these gatherings the participating individual domain scientists were standing-in as representatives of a larger epistemic community. For this "representativeness" to hold, the individual geoscientists had to ensure that the larger collective would stand behind (and contribute and use) their knowledge representations.

While in the first two years of the project modeling activities were described as concept-space workshops, following the third year (as outreach to non-participants geoscientists had become more clearly important) these events were renamed *community-based* ontology development to emphasize the increased effort to draw in and represent the geosciences rather than immediate GEON participants. This renaming was not a purely a rhetorical move intended to convince geoscientists that the ontologies reflected their knowledge and practices; rather, it constituted a processual evolution in the routine.

Within cyberinfrastructure development projects (large-scale scientific infrastructure) there is often a concern that the projects may be "top-down" – that is, that they are imposed upon the community. "Dumping" an ontology, developed in closed rooms "by committee," onto scientific practitioners was one example of such top-down development. It was assumed that other geoscientists would receive the products of such a process unfavourably and participants regularly expressed a sense of responsibility to their community. Community engagement became a way for GEON and its ontologies to shed its image of exclusivity.

One of the ways they found to do this was to make the ontologies public, that is, to make them accessible to debate and contribution by non-GEON geoscientists:

> This [ontology] can be something that geologists debate, contribute, and then they feel better. And they may feel more comfortable accepting those ontologies if it's not something that we usurpers [...] are dumping onto them. If it's something that they are discussing themselves, it can only help. (field research, geoscientist, December 2004)

In the excerpt above a geoscientists argued that opening the ontology to public debate would lead to a greater sense of participation and less chance of off-hand rejection. In the workshops, the small selection of geoscientists which came together to build ontologies had begun to think of themselves as "a sample" of the larger geosciences community and had come to see that the work of community enrolment was partially that of making the sample representative of the larger body:

> ontologies are not just one time things, obviously, so as a group [of geoscientists], how do we follow up? Clearly these [ontologies] can be exposed on the [GEON] website, but then we need to get a real engagement going. We might even bring-in some representatives from the community [to

workshops]. Like I said, put it up there, let the community look at it. We need some process. (field research, geoscientist, December 2004)

The excerpt evidences an understanding on the part of participants that engaging the community was important but also a reflexivity on the need to routinize the activity: "we need some process".

Part of the regularized routine of developing an ontology came to include posting drafts of the concepts spaces online on the GEON website/portal, and then encouraging colleagues to visit and comment on them. The most enthusiastic (and prestigious) geoscience colleagues were invited to the concept-space workshops themselves, to act as participants. Ontology development was moved out from closed rooms of computer and geoscientists and participants sought to make it a public endeavour. In doing so, the number of people working on ontology development was greatly expanded. As with the initial GEON principal investigators at the beginning of the project, these new participants were unfamiliar with the technical and practical details of ontology, necessitating a return to learning ontology and practice.

5.3. A return to learning ontology and practice

Attempts to engage the community returned GEON practitioners full circle to learning the problematic of interoperability, the first step in the routine. Here the tools and methods initially deployed by ontologists – e.g., the slide-sets (see Fig. 1) – came to be used by geoscientists themselves in processes of outreach to their domain communities. More than contributing, the community had to understand the value and practice of ontology as did GEON participants.

By traversing the first two steps – learning and practicing knowledge capture – GEON's geoscientists had been transformed. They had learned to see the disciplines of the geosciences as siloed, their databases stovepiped. Their thinking had been shaped by the necessity to express knowledge in the formalisms of ontology: predicates, logical operators, and database schemas. They had come to understand that ontologies required community support and participation. And finally, they came to believe that to participate, the entire community would have to learn what they understood of ontology; to do so they would have to traverse the routine. That is, GEON participants came to realize that while for them ontologies had become a valuable endeavour, for geoscientists "of the community in general," ontologies remained esoteric and that this would have to change. Ontology development becomes a missionary project.

The problematic of interoperability, the purposes of integration, or the future benefits of multidisciplinary data sharing were understood only within the group of domain practitioners who had traversed the learning aspects of the routine. Similarly, the investment and implications of modeling knowledge were understood only by those who had done so in practice. The broader community would have to be informed of ontology, the problematic of interoperability, and the work in GEON: "This is the opportunity to reach out to the geological community as a whole, [...] And that's the challenge, just how to lay this stuff on, 'this is what is going on', that this [ontology] is under development." (field research, geoscientists, March 2005).

Even more challenging, the ontologists of GEON (now including both computer and "converted" geoscientists) faced a more disciplinarily grounded problem. To those earth scientists outside the narrow sphere participating in knowledge capture, ontology remained not only esoteric but also "not geoscience": a subject appropriate to information technology or computer science but tangential to the study of the earth.

This is a commonly articulated problem within collaborative ventures between computer and domain scientists: while both seek to engage in research, what is considered a contribution to science for one group is not for the other (Lawrence, 2006; Ribes & Bowker, 2008; Ribes & Finholt, 2007). Thus, for earth scientists, ontologies were not a contribution to knowledge about the earth, or as GEON participants often stated it, not "something new about the Rockies." For instance, there was a concern that graduate students in the geosciences who worked with GEON were not receiving adequate disciplinary training if they focused on ontology development; similarly, junior faculty were concerned about spending too much time on data integration and not enough on basic research. This would reflect poorly in, for example, tenure review.

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

In response, part of community engagement activities within ontology development were to transform the community itself by proliferating educational and research materials about ontology. GEON participants published a special issue in a geoscience journal, and established a division at the Geological Society of America, which focussed on "geoinformatics" (the intersection of information technology with earth science research). Such activities were attempts to educate (and thus transform) the geoscience community on the importance, value and scientific contribution of activities such as ontology development.

The return to learning ontology was different in one marked respect: the drivers of these activities were geoscientists themselves. Using the very slides that computer scientists had used to teach them about interoperability, data integration and ontology, geoscientists ventured to their national society meetings and conferences presenting to their colleagues. They published articles in earth science journals extolling the virtues of multidisciplinary integration and the dangers of analyses based on isolated datasets. These "champions" of ontology were able to speak the language of cutting edge geoscience research, but parsed those science problems into matters of data integration and ontology enabled queries.

6. Discussion: traversing the acquisition routine as experience

Routines are not only a matter of repeated practice; they are also necessarily cognitive and experiential. Changing *ways of doing* has consequences for participants' orientation to the objects of those activities. In their excellent study tracing the routinization of polymerase chain reaction (PCR) through standardization into a kit, Jordan and Lynch argue that technical "micro" objects became available to everyday vernacular language. They describe how routinization in a kit changed the human experience of doing PCR. The formerly arcane practices of the lab – acting on "micro" objects which were not visible to the naked eye – could thereafter be described as straightforward written instructions on the PCR kit, rendering them accessible to "macro" action (Jordan et al., 1998). This phenomenological element has rarely been discussed in the literature on routines, which treats them as meso-level organizational phenomena (Vaughan, 1999) rather than as a part of everyday sense making and world making (Weick, Sutcliffe, & Obstfeld, 2005). Paraphrasing Goodwin (1994), a routine, as a learned, repeated and taught set of practical and material arrangements, generates a relevant *object of knowledge* (in this case, paradoxically, knowledge itself), a *domain of scrutiny* (the textbooks, journals, and community of the geosciences) and a set of *discursive practices* (knowledge acquisition, representation, and debate) being deployed within a *specific activity* (ontology development). Temporality (Preda, 2006) is also shaped by the routine, which emphasizes repeated iteration over years to achieve a durable knowledge representation.

Developing – and for latecomers, traversing – the acquisition routine initiated a phenomenological transformation for participants, translating how they viewed knowledge and data in their fields and how they viewed the broader domain community. The activity of capturing knowledge made new objects available for consideration while old objects took on new meaning. In this paper we have focused on how geoscientists came to see knowledge and data through the informational paradigm of computer science. Furthermore relations to "the broader community" of the domain took on a new significance: no longer only colleagues in a shared pursuit of knowledge but now also potential users (or, more worrying for them, non-users) of an emerging semantic computational tool.

In the first step of the routine participants learned to communicate across their expertise: domain participants began to understand "what is an ontology" and why it is relevant to scientific work. We called this learning the problematic of interoperability. *Enacting the problematic of interoperability reconstituted science through the eyes of its data.* No longer were data only an individual researchers' raw materials, rather, they became a community resource. But these data were *not yet* a community resource until they were interoperated, able to move seamlessly across disciplinary, institutional and technical barriers. Participants came to see that in the geosciences this was not the case, that their data was divided by heterogeneous categories and database schemas.

The practice of knowledge acquisition sharpened this sense of disorder (Berg & Timmermans, 2000). It was in the practical learning, or learning-by-doing, that domain scientists began translating their knowledge into inherited categories, logical operators, and predicates. Geoscience knowledge

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

itself began to appear messy and disorganized from the perspective of the requirements for modeling. Ontologies have their own epistemology: what and how the computer can "know" is very particular, limited by the availability description logics and the extant level of formalization. This second step of the routine involved a *reapprehension*[7]: a double movement of the domain specialists reflecting on their own knowledge base and beginning to rework it terms of a language of ontology; in turn, IT experts had to establish a sufficient familiarity with the domain knowledge such that they could offer themselves as a resource for this translation, all this while skirting the difficult line of intervening in domain knowledge.

This practice, articulating knowledge in the technical languages of modeling, revealed to domain participants an entire landscape of nuanced disciplinary difference. Previous to exposure through the explicit practice of knowledge capture, these divergent beliefs or commitments were often held unproblematically. Developing a knowledge representation forced them to confront endless nuanced disagreements and to see the data of their field as disordered.

The third step in the routine, engaging the community, placed the success of an ontology well outside its development and instead in its acceptance by a community of "knowers". This community was no longer simply their intellectual colleagues but also a potential user base (Woolgar, 1991), a target for technological uptake (Ribes & Finholt, 2008). As this routine emerged, the definition of success for an ontology came to stretch beyond a narrow definition of technological development ("is the ontology complete?") to its uptake and usage by domain practitioners ("is the routine sufficiently representative of the community?"). In doing so "the broader community" became important as part of an outreach project: a community which itself had to be engaged and transformed such that they would use and contribute to ontologies.

The third step of the routine returned participants to the first step, only now it was the geoscientists themselves who were teaching their colleagues the problematic of interoperability. Traversing the routine had made "lay ontologists" out of geoscientists (Epstein, 1996). They had come to know their data in ways that resembled the orientation of computer scientists. In order to engage the community in using ontologies and registering their data, members of the community had to understand the value of sharing data, and of ordering them through ontologies.

7. Conclusion

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 of ontologies required the simultaneous invention of (routinized) techniques for education, dissemination and application. In a stripped down technical understanding of knowledge acquisition, engineering ontologies involves taking domain knowledge, formalizing this knowledge into a machine computable format, and encoding it into machine language. The resulting ontologies become infrastructure, invisibly supporting the work of users in their queries and collaboration. But in the practical work of developing ontologies, participants engaged in a much broader range of activities than any such technical definition revealed and ontologies were never quite so transparent.

We traced the work of participants as they simultaneously developed ontologies and also a routine to support future modeling. This routine was not planned but emergent (Ashforth & Fried, 1988; Cohen, 1991), it was shaped in response to the repeated encounters with difficulties in the pedagogy of ontology, the practice of knowledge acquisition and engaging the community. The routine was also more than simply a way getting things done: it proved transformational for those who traversed its path. The practices of the routine changed the orientation of participants towards their knowledge and communities: domain knowledge became that which had to be articulated in the language of modeling and colleagues those who had to be enroled in the use and maintenance of ontologies.

[7] Our use of this term is greatly indebted to the work of Karin Knorr-Cetina (2003), who reformulates the phenomenological concept of 'apresentation' in terms of the practical activities of bringing close that which is far, and the philosophy of science concepts 'representation' and 'reference' as a matter of situated experience. Clearly, scientists often spend a great deal of time looking over, manipulating and rendering their data; the term reapprehension is intended to emphasize how they come to see data and knowledge anew as a question of informational order, and then seek to act on its organization *as* information.

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

Information technologists have always had as part of their output the goal of producing tools – e.g. computing power, visualization, data management and storage. In this sense ontologies were no different: they are a tool for integrating data, and enabling the user to execute sophisticated searches and queries. But building ontologies required something that few applications had demanded before: the users had to be enroled in its development by contributing their knowledge and then reaching out to their community. The success of ontologies became predicated on the communities' participation in their development, use and maintenance. A community which itself first had to understand the benefits of ontology. In its wake, ontology development left behind a domain more acutely aware of its data infrastructure and in greater need of further information organization.

References

Ashforth, B. E., & Fried, T. (1988). The mindlessness of organisational behaviors. *Human Relations, 41*, 305–329.

Atkins, D. E. (Chair) (2003). Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. In *National Science Foundation*.

Berg, M., & Timmermans, S. (2000). Orders and their others: On the constitution of universalities in medical work. *Configurations, 8*, 31–61.

Berkley, C., Bowers, S., Jones, M. B., Ludascher, B., Schildhauer, M., & Tao, J. (2005). Incorporating semantics in scientific workflow authoring. In *Proceedings of scientific and statistical database management, SSDBM'05* (pp. 75–79). Santa Barbara, CA.

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

Borgman, C. L. (2007). *Scholarship in the digital age: Information infrastructure, and the internet*. Cambridge, MA: MIT Press.

Bowker, G. C. (1994). *Science on the run: Information management and industrial geophysics at Schlumberger, 1920–1940*. Cambridge, London: The MIT Press.

Bowker, G. C. (2000). Biodiversity datadiversity. *Social Studies of Science, 30*(5), 643–683.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, Mass.: MIT Press.

Button, G., & Sharrock, W. (1994). Occasioned practices in the work of software engineers. In M. Jirotka & J. Goguen (Eds.), *Requirements engineering: Social and technical issues* (pp. 217–240). New York: Academic Press.

Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A., et al (2002). Data withholding in academic genetics: Evidence from a national survey. *JAMA, 287*(4), 473–480.

Ceci, S. (1988). Scientists' attitudes toward data sharing. *Science, Technology and Human Values, 13*(1–2), 45–52.

Clarke, A., & Fujimura, J. (Eds.). (1992). *The right tools for the job: At work in twentieth-century life sciences*. Princeton, NJ: Princeton University Press.

Cohen, M. D. (1991). Individual learning and organizational routine: Emerging connections. *Organisation Science, 2*, 135–139.

Collins, H. M. (1981). Knowledge and controversy: Studies of modern natural science. *Special Issue of Social Studies of Science, 11*(1).

David, P., & Bunn, A. (1988). The Economics of gateway technologies and network evolution: Lessons from electricity supply history. *Information Economics and Policy, 3*, 165–202.

Egyedi, Tineke (2001). Infrastructure flexibility created by standardized gateways: The cases of XML and the ISO container. *Knowledge, Technology and Policy, 14*(3), 41–54.

Epstein, S. (1996). *Impure science: AIDS, activism and politics of knowledge*. University of California Press.

Feldman, M. S. (2000). Organisational routines as a source of continuous change. *Organisation Science, 11*(6), 611–629.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly, 48*(1), 94–121.

Forsythe, D. E. (1993). Engineering knowledge: The construction of knowledge in artificial intelligence. *Social Studies of Science, 23*, 445–477.

Forsythe, D. E., & Buchanan, B. G. (1989). Knowledge acquisition for expert systems: Some pitfalls and suggestions. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 435–442.

Gaines, B. R. (1989). Social and cognitive processes in knowledge acquisition. *Knowledge Acquisition, 1*(1), 39–58.

Galison, P. L. (1999). Trading zone: Coordinating action and belief. In M. Biagioli (Ed.), *The science studies reader* (pp. 137–160). New York: Routledge.

Galison, P. L., & Stump, J. D. (1996). *The disunity of science: Boundaries, contexts and power*. Stanford: Stanford University Press.

Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review, 48*(6), 781–795.

Gieryn, T. F. (1999). *Cultural boundaries of science: Credibility on the line*. Chicago: University of Chicago Press.

Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606–633.

Graham, S., & Thrift, N. (2007). Out of order: Understanding repair and maintenance. *Theory Culture and Society, 24*(3), 1–25.

Gruber, T. R. (1993). A translation approach to portable ontology specification. *Knowledge Acquisition, 5*, 199–220.

Guarino, N., & Welty, C. (2000). A formal ontology of properties. In R. Dieng & O. Corby (Eds.), *Knowledge engineering and knowledge management: Method, models and tools. 12th International Conference (EKAW 2000)* (pp. 97–113). Berlin/New York: Springer.

Hanseth, Ole, Monteiro, Eric, & Hatling, Morten (1996). Developing information infrastructure: The tension between standardization and flexibility. *Science, Technology and Human Values, 21*(4), 407–426.

Horrocks, I., Patel-Schneider, P. F., & van Harmelen, F. (2003). From SHIQ and RDF to OWL: The making of a web ontology language. *Web Semantics: Science, Services and Agents on the World Wide Web, 1*(1), 7–26.

Jordan, K., & Lynch, M. (1998). The dissemination, standardization and routinization of a molecular biological technique. *Social Studies of Science*(28), 773–800.

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001

Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, Mass.: Harvard University Press.
Knorr-Cetina, K. (2003). From pipes to scopes: The flow architecture of financial markets. *Distinktion, 7*, 7–23.
Latour, Bruno (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, Mass.: Harvard University Press.
Lynch, M., & Woolgar, S. (Eds.). (1990). *Representation in scientific practice*. Massachusetts: MIT Press.
Lawrence, K. A. (2006). Walking the tightrope: The balancing acts of a large e-research project. *Computer Supported Cooperative Work, 15*(4), 385–411.
Mackay, H., Carne, C., Benyon-Davies, P., & Tudhope, D. (2000). Reconfiguring the user: Using rapid application development. *Social Studies of Science, 30*(5), 737–757.
Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent Systems, 16*(2), 72–79.
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., & Booker, J. M. (2001). *Eliciting and analyzing expert judgment: A practical guide*. Philadelphia, PA: Society of Industrial and Applied Mathematics.
Noy, N. F., & McGuinness, D. L. (Unpublished manuscript). *Ontology development 101: A guide to creating your first ontology*.
Preda, A. (2006). Socio-technical agency in financial markets: The case of the stock ticker. *Social Studies of Science, 36*(5), 753–782.
Ribes, D., & Finholt, T. A. (2007). Tensions across the scales: Planning infrastructure for the long-term. In *Proceedings of the 2007 international ACM SIGGROUP conference on Supporting group work, Sanibel Island, Florida, USA* (pp. 229–238).
Ribes, D., & Finholt, T. A. (2008). Representing community: Knowing users in the face of changing constituencies. In *Proceedings of the 2008 ACM conference on Computer-Supported Cooperative Work (CSCW)*.
Ribes, D., & Bowker, G. C. (2008). Organizing for multidisciplinary collaboration: The case of the geosciences network. In G. M. Olson, A. Zimmerman, & N. Bos (Eds.), *Scientific collaboration on the internet*. Cambridge: MIT Press.
Sheth, A. P. (1999). Changing focus on interoperability in information systems: From system, syntax, structure to semantics. In G. Egenhoer & K. Fegeas (Eds.), *Interoperating Geographic Information Systems*. Boston: Kluwer Academic Publishers.
Shrum, W., Genuth, J., & Chompalov, I. (2007). *Structures of scientific collaboration*. Cambridge, MA: MIT Press.
Sowa, J. (2000). *Knowledge representation: Logical, philosophical, and computational foundations*. Boston, MA: MIT Press.
Star, S. L. (1991). Invisible work and silenced dialogues in knowledge representation. In I. V. Eriksson, B. A. Kitchenham, & K. G. Tijdens (Eds.), *Women, work and computerization*. Elsevier Science Publishers B.V., North-Holland.
Star, S. L. (1999). The ethnography of infrastructure. *American Behavioral Scientist, 43*, 377–391.
Star, S. L., & Ruhleder, K. (1994). Steps towards and ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. In *Proceedings of the conference on computer supported cooperative work* (pp. 253–264). Chapel Hill: ACM Press.
Vaughan, D. (1999). The role of the organization in the production of techno-scientific knowledge. *Social Studies of Science, 29*(6), 913–943.
Waterman, D. A. (1986). *A guide to expert systems*. Reading, MA: Addison-Wesley.
Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science, 16*(4), 409–421.
Woolgar, S. (1991). Configuring the user: The case of usability trials. In J. Law (Ed.), *A sociology of monsters*. Routledge.

Please cite this article in press as: Ribes, D., & Bowker, G. C. Between meaning and machine: Learning to represent the knowledge of communities. *Information and Organization* (2009), doi:10.1016/j.infoandorg.2009.04.001