

Research Article

A Situated Knowledge Representation of Geographical Information

Mark Gahegan
*GeoVISTA Center
Department of Geography
Pennsylvania State University*

William Pike
*Pacific Northwest National Laboratory
Richland, Washington*

Abstract

In this paper we present an approach to conceiving of, constructing and comparing the *concepts* developed and used by geographers, environmental scientists and other earth science researchers to help describe, analyze and ultimately understand their subject of study. Our approach is informed by the *situations* under which concepts are conceived and applied, captures details of their construction, use and evolution and supports their ultimate sharing along with the means for deep exploration of conceptual similarities and differences that may arise among a distributed network of researchers. The intent here is to support different perspectives onto GIS resources that researchers may legitimately take, and to capture and compute with aspects of epistemology, to complement the ontologies that are currently receiving much attention in the GIScience community.

1 Introduction

You can know the name of a bird in all the languages of the world, but when you're finished, you'll know absolutely nothing whatever about the bird . . . So let's look at the bird and see what it's doing – that's what counts.

(Richard Feynman)

Foreword

This paper arises from the AAG meeting in 2005, and specifically two special sessions convened by Sarah Elwood and Nadine Schuurman to discuss the progress made towards addressing the critiques offered by many geographers concerned with the social,

Address for correspondence: Mark Gahegan, Department of Geography, Pennsylvania State University, University Park, PA, 16802, USA. E-mail: mng1@psu.edu

political and representational inadequacies of Geographical Information Systems (GIS). By design, these sessions coincided with the tenth anniversary of the publishing of several such critiques in the book *Ground Truth*, edited by John Pickles (1995). The purpose of this paper is not to explicitly tackle these critiques head on, but rather to show how progress towards better representational methods in computer and information science can help to bring to light many aspects of meaning and epistemology that currently lurk in the background of many GIS applications.

1.1 *The Problem*

In GIS and most other computational systems, the comprising datasets, methods, conceptual models and other resources are usually treated as objective entities. However, they were created for – and are applied to – specific situations that affect their objectivity in many ways. These situations are generally not captured or represented at all, except perhaps in the minds of the people involved in their creation and use, or after the fact in journal articles far removed from the actual processes of inquiry. And when data producers and data consumers do not share such situational knowledge, the chances of misunderstanding and misapplying resources increase. Davis and Simmonet (1991) sum up the situation thus: “... *the processes and data used to generate the cartographic information are usually unknown or irretrievable*”.

Many of the social critiques of geographic information systems and science take issue with the often-assumed objectivity of data and methods, pointing out that such resources are in fact value-laden artifacts that perhaps say as much about their creators as they do about the world (Latour 1987, Pickles 1999, Harvey 2000, Schuurman 2000). Furthermore, the use of these resources, if unquestioned, can lead to misplaced and unwarranted trust in their validity and suitability at best, and the adoption of some unwritten agenda at worst. In Pickles' (1995) original critique, these themes of unrepresented agendas, blind faith in technology and data subjectivity crop up many times (e.g. Veregin 1995, McHaffie 1995, Goss 1995). Yet from the ensuing tension highly useful dialog between critical theorists and GIScientists has emerged (see, for example, the special issue of *Cartography and Geographic Information Science* published in 1995, NCGIA 1998, and Schuurman 2002).

Like any other technology, GIS has its limitations and is open to misuse and abuse. Some of these problems are a consequence of larger social systems that GIS exist within, and others are a direct result of the impoverished descriptions of data and other resources that GIS and other computational technologies currently offer. We focus our attention here on the latter set of problems since they are within our reach as GIScientists, while also acknowledging the importance of the former. While the solutions we propose focus on computational techniques, such approaches could ultimately serve as catalysts for change in the power structures between researchers and public by providing a model for representing and sharing the different meanings and outcomes each attributes to representations of places. Indeed, efforts such as ours to represent places conceptually rather than through GIS data alone might offer ways to integrate public and expert views on a problem.

In GIS-based settings, analysts explore complex and voluminous data resources, and combine them in various ways to synthesize new understanding. These activities both utilize and produce knowledge that for the most part remains unrecorded, residing only in the volatile memory of the analyst(s). In current GIS databases there is simply

no place for a more knowledge-oriented description of resources, nor any means to represent user-experiences or points of view that pertain to them.

Taylor (1990), as quoted by Pickles (1995) sums up the situation thus:

Knowledge is about ideas, putting ideas together into integrated systems of thought we call disciplines. Information is about facts, about separating out a particular feature of a situation and recording it as an autonomous observation . . . The positivist's revenge has been to retreat to information and leave their knowledge problems – and their opponents – stranded on a foreign shore. But the result has been a return to the very worst sort of positivism, a most naïve empiricism.

Such a situation is undesirable for many reasons, including the following:

1. It is not possible to question the knowledge formed in any deep way, nor clearly connect knowledge, via causal mechanisms, to the information from which it derives.
2. Analysts must play the role of institutional memory, though they may forget what they have learned, or may be unavailable when their expertise is needed. As a result, important background (tacit knowledge) is easily overlooked because it is not made present within the investigative environment. The *a-priori* knowledge, evolving understanding of the situation, working hypotheses and likely outcomes go unrecorded, so the system itself remains open, ambiguous and highly subjective – though it may appear to be highly objective given that there is no represented basis for doubt. This absence of understanding plays an important role in reducing the utility of information to others, whether intentionally or not, and consequently affecting their agency.
3. It is difficult to reconstruct an understanding of the social setting that gave rise to the outcomes, thus it becomes problematic to assess the role of existing organizations and power structures or to critique the situation that gave rise to the outcomes.

In short, the existing state of affairs promulgates the notion of outcomes created in a vacuum, then presented to potential users as objective truth, devoid of any useful epistemology (though perhaps increasingly rich in ontology). It is our intention, through the work described below, to bring knowledge and information in GIS more closely together.

1.2 A Possible Way Forward

Significant progress has been made recently in three areas that we argue here form a more complete picture of meaning. The first area is formal top-down knowledge that captures aspects of meaning relating to domains of discourse and associated tasks (Gruber 1995, Guarino 1997, Fonseca et al. 2002). The second is problem-solving environments and workflow management systems that can represent the mechanical aspects of *how* resources were constructed and used (Boisvert and Rice 1996, Takatsuka and Gahegan 2002, Sanchez and Langley 2003). Finally, and perhaps most importantly, research into the more subjective, situational aspects of meaning has also made significant breakthroughs, especially in terms of computational representation (Sowa 2000a, b; Fischer and Ostwald 2001; de Moor et al. 2002). It is our *aim* to bring these three perspectives into some kind of alignment, based on the premise that humans reason by combining context and procedural knowledge with more formal domain and task

knowledge. In a real sense, one might view this aim as an attempt to combine computational approaches to both epistemology and ontology, to better represent meaning. As a consequence, we might hope to give a voice to different individuals and communities, specifically by representing a wealth of different perspectives that might be taken on geospatial information.

In the results that follow later, we show that some aspects of the situations surrounding creation and use of resources can be harvested, remembered, mined, visualized and applied to help communicate some of these missing aspects of meaning, and to complement the more objective, top-down knowledge that might be provided by computational ontologies. In our efforts we have been influenced by the writings of Charles Sanders Peirce, particularly with regard to semiotics and pragmatics (Peirce 1877, 1905, 1931), and Alfred North Whitehead, and his writings on the philosophical and systematic foundations of knowledge representation (Whitehead 1929, 1933, 1938).

Adding in situational aspects of meaning to those represented by ontologies results in an intricate web (a *nexus* in Whitehead's terminology) of relationships linking people, methods, data, places, times, concepts, tasks and more that can be built up over time as resources are created, modified and used by communities of researchers. In this way, resources are contextualized in a manner that reflects, albeit to a limited extent, similarities and differences in the way they are understood by their users, *through* the ways they are used.

We demonstrate these ideas via a web portal (named Codex) that forms a gateway to resources for groups of researchers working together. With the onset of national and international cyber-infrastructures and collaboratories (also known as e-science initiatives in Europe) to support collaboration between researchers, we are now in a position to access more data, methods and other resources than ever before (Finholt 2001, NSF 2003). But another way of stating this is that we are now further removed from the creators of resources we use than has previously been the case. Simply making e-resources available widely does not insure their appropriate use (or even their use at all). Better research will result only if these large distributed networks of researchers can be formed into communities that can exchange understanding as well as outcomes. Cyber-infrastructure, like related efforts in public-participation GIS (e.g. Carver et al. 2001) underlines the need for better representation and communication of understanding, and it is these needs that have motivated our own research efforts. We draw inspiration from prior efforts to create systems that represent inquiry, and especially the vision of Vannevar Bush (1959) that predates computational approaches, but also notably the recent work of Buckingham Shum et al. (2000). We then extend these models by incorporating representations of both the data resources in use and the social and knowledge-creation circumstances of their use. In this paper we draw specific examples from two cyber-infrastructure projects of which we are a part, the Geosciences Network (GEON; <http://www.geongrid.org>) and Human Environment Regional Observatory (HERO; <http://hero.geog.psu.edu>).

2 Challenges to Representing Geographical Meaning

We begin by acknowledging that the concepts we create to help us understand the world say as much about us as they do about the world. But rather than using this inherent subjectivity as a reason to abandon quantitative approaches, we regard it instead as an inspiration to improve our science so that it becomes better 'situated' within communities

and work practices; making explicit some of those subjective aspects and providing what we believe is a more conceptually flexible and open foundation for collaborative geographical enquiry.

We further acknowledge that the conceptual structures we use to understand the world are themselves not static; they shift in response to the development of our own personal understanding, through the process of community debate, and the formation of consensus. Sowa (2002) shows that this insight is in fact far older than the twentieth century philosophy of science, to wit:

According to Heraclitus, *panta rhei* – everything is in flux. But what gives that flux its form is the *logos* – the words or signs that enable us to perceive patterns in the flux, remember them, talk about them, and take action upon them even while we ourselves are part of the flux we are acting in and on.¹

Implied in this insight are four important challenges in the representation of geographical meaning. Firstly, the world is changing, so concepts must either adapt accordingly or become obsolete.² Secondly, we as individuals and groups are also constantly changing, so our needs, goals understanding and experience – i.e. our bases for constructing concepts – are also in flux (Brodaric et al. 2004). Thirdly, we use words or signs to stand for (encode) concepts, but there is no guarantee that concepts will be understood in the same way by all parties during communication (Reschner 1978, Noth 1990). Fourthly, we need to keep track of the conceptual structures we construct and use since they are the key to understanding our data and other outcomes.

2.1 *Not Just Ontology*

In our attempts here to better represent meaning, we here take an approach that conceives of meaning as having three strands, comprised of syntax, semantics and pragmatics, after the ideas originally proposed by Peirce. Syntax details the mechanics of knowledge representation, with details of schema and encodings; semantics describes the conceptual structures and their inter-relationships, along with associated constraints and rules; pragmatics describes the situations surrounding the creation and use of meaning. We show later how these three strands can be independently represented, then woven together into a single thread.

Current research into the use of computational ontologies as a representation of geographical semantics has led to promising results (e.g. Rodriguez et al. 1999, Agarwal 2004) and some immediately useful knowledge structures that – usually in a theoretical sense – describe some of the rich typologies common in geographical systems. A good example is the set of SWEET ontologies, produced by NASA (<http://sweet.jpl.nasa.gov>), and describing Earth realms (e.g. sea floor, atmosphere layer boundary), physical phenomena (e.g. climate change, time interval) and various measurement systems (e.g. coordinates, spatial distribution).

Despite claims that computational ontologies are simply a “*specification of a conceptualization*” (Gruber 1993), for the most part, these current ontological approaches tend to be treated as a formal representation of ‘Truth’; that is, some agreed-upon or imposed semantic structures that are assumed to be ‘true’, and about which debate, proposed alternative structures and change are not supported (Marcos and Marcos 2001). Yet within the discipline of geography (and many other disciplines no doubt), knowledge can be uncertain, incomplete, evolving, contradictory and oftentimes

contested. Computing with a multiplicity of uncertain knowledge structures is a vital step towards progress in representing meaning in GIS. Preserving these often-important differences is also a vital step in giving a voice to the different perspectives and interpretations that can legitimately be made around some map or other GIS resource.

Indeed, it is our contention that, far from being a “*silver bullet*” (Fensel 2001), ontologies by themselves solve only a part of the problem of succinctly representing and communicating the meaning of resources. Perhaps one could also argue that concentrating solely on ontological knowledge in GIScience might result in a worsening of the problems described by Pickles and colleagues, in the sense that more objectivity may tend to re-enforce the belief that resources can always be taken at face value, or may fuel the argument that GIS resources are not critically examined (or even critically examinable) by their users.

2.2 Managing Concepts That Emerge From Situations and Evolve

In summary, there has been relatively little investigation to date of how conceptual structures emerge from practice and how they can reflect the evolving understanding of a situation by individuals and communities. This contrasts with a growing recognition in the computer science and artificial intelligence communities that tools to support knowledge representation must account for the situated work practices of their users (following from research into situated cognition, e.g. Barsalou and Medin 1986, Clancey 1994) and accommodate the dialogical, interactive nature of exploration (Schultze and Boland, 2000, Nake and Grabowski. 2001). So we aim here to extend the current focus on ontologies to include the syntactic and methodological details residing in workflows, and the pragmatic aspects of meaning carried within social networks and use-cases. In this regard we set ourselves two objectives.

Our first objective stems from the fact that computational systems to support social, interpretive science must do more than just enumerate a set of ontological commitments – they should also reflect how it is that we come to make those commitments. A more complete and examinable record of work conducted to produce some shared outcome will help others to assess both its usefulness (to them) and its inherent biases, as well as providing one strand to the elusive thread of meaning. Our second objective arises from the many critiques of science that rightly question the dissociation of researcher from research (e.g. Feyerabend 1988); we therefore aim to support critical examination of knowledge in light of the actors involved – the producers and consumers of the GIS resources. This second objective leads us to connect resources with the social networks of actors, their communities and associated organizations that represent the social dynamics and power structures involved in the creation and use of resources.

Before we proceed, three caveats are in order. Firstly, by no means do we claim our representations of these aspects of meaning are complete. But we do claim that the structures we describe can accommodate additional concepts, relations, descriptions, and attributes as required by their users to better represent their specific needs. In addition, note that although the representation mechanism is that of formal logic, this does not mean that magically we have discovered how to capture and represent the wealth and depth of human understanding within a computer. Formal logic takes its name from an emphasis on the ‘form’ of a representation, that is, its structure as a surrogate for real understanding.

Secondly, the goal here is not to build a machine representation of knowledge that allows a computer to function as a geographer, but rather to string together a number

of ontological and situational cues from which a geographer might synthesize a more complete understanding, and in doing so, be in a position to interpret and use GIS resources more wisely.

Thirdly, one important assumption we make at the onset is that researchers, by and large, do strive to present to each other their deep, conceptual understanding. True, there are sometimes those who prefer to guard their knowledge (or parts thereof), but for the most part, journal articles and conference presentations are (often) excellent examples of researchers laying out rich ideas for their colleagues to try and grasp (typically using a combination of rhetoric, logic and mathematics as the conveyors of meaning). So it is our contention that the sharing of knowledge is a well-practiced community norm; we are simply trying here to improve the act of communicating, situating it within the processes of science (interaction, debate and argument) in a more convenient manner than the article – so that the packaging supports more efficient searching and browsing strategies for those wishing to engage the material.

3 What Were We Thinking?

The philosopher and mathematician, Alfred North Whitehead (1934) noted that unattended knowledge, “. . . *keeps no better than fish.*” The first step towards more effectively sharing concepts is therefore to somehow remember them. That is not to say that aspects of meaning are not currently reported; pieces may appear in field reports, published articles, presentations, even scripts and log files created by commercial GIS and in many other outlets. However, in these forms the details can be difficult to elucidate, and their connection with sharable outcomes such as datasets and methods to which they apply is tenuous or even non-existent. In short, the conceptual details are simply not accessible enough. Even in the form of the traditional scientific notebook these details would not be readily searchable without a good deal of investment in text pre-processing and mark-up.

From this perspective our research can be interpreted as developing a modern, computer-based representation of collaborative science, one that captures and communicates a rich characterization of the scientific process (e.g. <http://www.discoverynet.org>) via a knowledge portal. As such it extends earlier efforts that address only the sharing of data and methods (Myers et al. 2001). As well as keeping an account of tasks performed by individuals, and a reproducible record of activity, a portal shared by a community can provide answers to questions such as:

- Who first introduced this concept?
- What methods and datasets have been used to synthesize or signify this concept?
- Which individuals and groups have applied this concept, and to what problems?
- Do the recorded conceptualizations of two individuals using the same dataset agree?
- Are they also consistent with the conceptualization of the originator of the dataset?

Section 4 gives some examples to illustrate how we represent these kinds of questions.

3.1 A Nexus of Meaning

Working at the meta-level we begin by identifying the elements of scientific analysis as we conceive of them (Figure 1). This diagram represents our current conceptualization

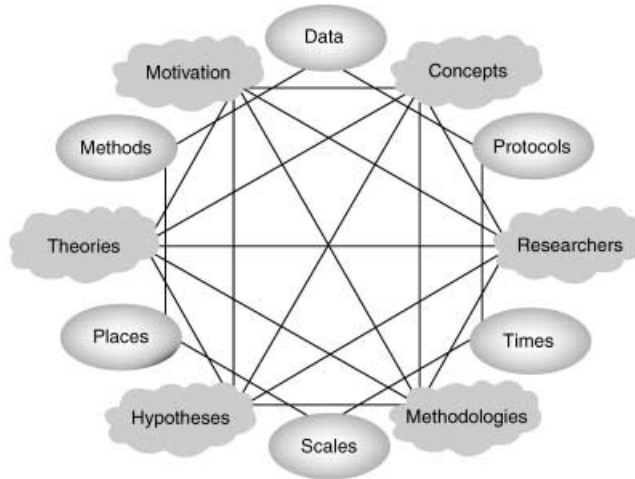


Figure 1 A *nexus* (Whitehead 1938) of the conceptual structures used in our work. Oval shapes denote concepts that can be represented (at least in part) by existing metadata standards; cloud shapes intimate that others are largely conceptual and currently require a richer form of representation, such as a narrative

– itself a compromise between what we would like to be able to represent, and what may be practical, and while comprehensive, it is surely not exhaustive. To borrow a term from Sowa (2002), these elements of inquiry can be seen as a kind of ‘knowledge soup’, some parts of which are clearly identifiable and describable (the ‘vegetables’) while others are not (the ‘broth’), yet together they define the flavor. At the present time, we typically split these ingredients out into two sets: those elements we can describe formally – these become meta-data, and those that are more amorphous and difficult to describe and communicate – these are usually left out of any computational or analytical ‘system’ though as noted above they may appear in other places, such as reports and academic papers.

Different communities of researchers may use their own conceptual structures to synthesize their own unique nexus, which better suits their specific needs. Table 1 below gives a more formal definition of each of the conceptual structures that appear in Figure 1, and provides examples of the values they might take. These structures arise in part from study of the science process (e.g. Popper 1959, Kuhn 1962, Buckingham-Shum et al. 2000) and also from our own experience as collaborators, and recognize the practical needs (and limitations) of our own efforts to capture and exchange information and knowledge.

From a philosophical perspective many have argued that the use of fixed protocols and paradigms can ultimately act as a barrier to scientific progress (Whitehead 1933, Kuhn 1962). It is for this reason that we focus our efforts at the meta-level (Michalski 1993) since the elements of inquiry used in science, rather than their specific instances, appear to be much more stable – for example some of the conceptual structures defined below in Table 1 were originally coined by Aristotle.

The system described later explicitly represents some of these conceptual structures, where currently practical. What we currently lack, however, are means to capture and

Table 1 Some of the conceptual structures that form a part of the science process

Knowledge ingredients	Definition (role)	Examples
Data/Files	Measurements, observations, and anything used to signify concepts.	Census data, images, photographs, newspaper articles, video clips, interviews.
Methods	Analytical tools such as GIS functionality, statistical procedures or web services.	Overlay, buffering, clustering metrics, rank correlation coefficients, text transcribing and summarizing.
<i>Concepts</i>	The mental devices used to understand and communicate different ideas that help to structure the domain of discourse.	Ecosystem, natural hazard, landuse categories, poverty, vulnerability, adaptive capacity, resilience.
Protocols/Tasks/ Workflows	Formalization of some task, e.g. by a set of data and methods arranged in a workflow to accomplish some task.	The calculation of indices, e.g. poverty, vulnerability; land use change assessment.
Scales	The scale, or range of scales over which some resource or conceptual structure might apply.	Local, regional, global, 1:100,000.
Times	The time, or range of time over which some resource or conceptual structure might apply.	Quaternary, 2006, late 20th Century, current.
Places	The place, or places within which some resource or conceptual structure might apply.	Study sites, typically defined to match either natural boundaries or administrative units such as counties.
<i>Hypotheses</i>	The question that drives a particular analysis exercise, whose truth will ultimately be evaluated.	That the rate of forest planting exceeds that of harvesting in the Pacific northwest, that the USA is a net carbon sink.
<i>Methodologies/ Philosophy</i>	The specific philosophical approaches or methodologies that direct the choice of how theories and hypotheses are framed and evaluated.	Typically these might span a range between quantitative and qualitative forms of analysis, from parametric statistics to narrative.
<i>Theories</i>	Belief structures that describe links between concepts, methods, data and the systems they represent, within a specific philosophy.	Anthropogenic climate change, plate tectonics.
<i>Researchers</i>	The individuals, teams and organizations involved in analysis, with their own unique experience and understanding.	Individual HERO and GEON collaborators, lab groups, international research communities.
<i>Motivations</i>	The explicit and implicit motivating factors, goals and intentions that shape all of the above aspects.	Research interests, funding sources, politics, economics.

represent the more abstract, but arguably more important elements toward the bottom of the list. Workflows or protocols that describe the data and methods used might allow us to reconstruct how a particular resource was constructed, and concept maps can give insight into how the dataset is understood by its creators, but not why it was constructed – which would require representation of the researchers' motivations and social settings.

4 Supporting Different Perspectives in Practice

The knowledge structures in Table 1 each provide a unique **perspective** onto a collection of resources. By holding some resource of interest constant (such as a map), users can change between these different perspectives, to examine different aspects of the map, such as how it was made or who has used it. At the time of writing our Portal, called 'Codex' after the style of historical science notebooks such as da Vinci's, provides a central access point to six specific perspectives (we plan for later versions to support additional points of the nexus shown in Figure 1 and Table 1):

- **Researchers** – the individuals and groups who create or apply resources accessed through the Portal. Each person who touches a resource leaves a mark, revealing patterns of use that constitute a user's sphere of interest; a kind of social network (Schweizer 1997).
- **Concepts** – descriptions of abstract ideas used by researchers. Concepts might include "vulnerability," "flood", or "drought". Sets of concepts specifically include both informal concept maps and more formal ontologies.
- **Data resources (typically files)** – that express something about a concept. Files could include spreadsheets, text documents, images, audio clips, maps, or other data formats (quantitative or qualitative) that connect observations or measurements to the cognitive structures represented by concepts.
- **Tools** – the methods used to analyze data. Tools could include GIS operations, predictive models, interviewing instruments, or statistical tests.
- **Places** – geography is fundamental to integrative human-environment and geoscience research, and places help researchers define the locations and scales under study, whether described as bounding polygons or as place names. Place may also account for large differences in epistemology between researchers, so plays a major role in defining the nature of the defined concepts.
- **Tasks** – people, concepts, files, tools, and places are linked together through tasks that might describe a workflow process, an experimental procedure, or a problem-solving approach. One goal central to the Human Environment Regional Observatory (HERO) project previously mentioned is the development of tasks to inform understanding of human-environment interactions, specifically land use/land cover change and assessing vulnerability to extreme climatic events.

The portal model gives each user and community a customized view onto different resources and conceptual structures; users have personalized workspaces that serve as repositories for their own concepts, data and other resources, while groups of researchers can establish a common workspace across which resources and conceptual structures are constructed and shared (see Figure 2). The Web interface makes a user's workspace accessible through any workstation, in any location, enabling a researcher's work to follow him or her while always providing a point of access to progress made by collaborators.

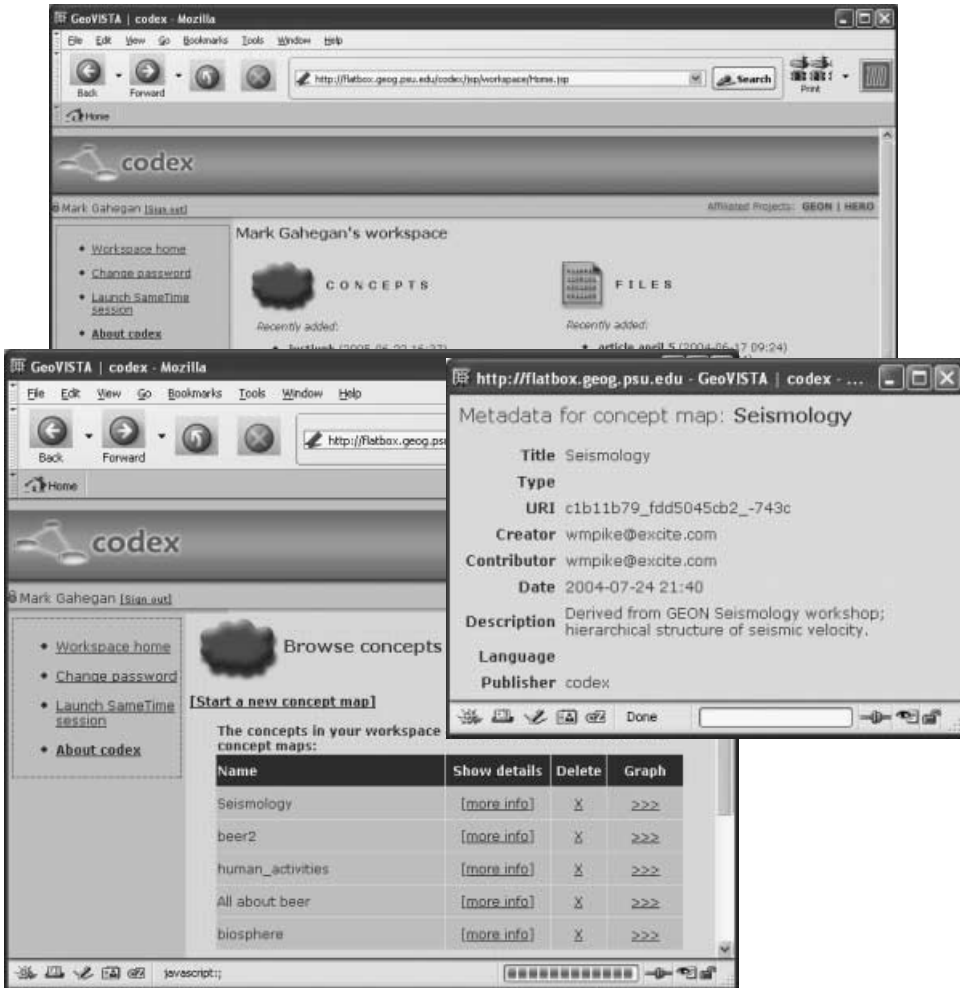


Figure 2 Parts of the user interface within the Codex knowledge portal showing (from back to front): (1) those concepts and other resources recently added to a user's knowledge workspace; (2) larger concept maps (collections of concepts and associated relations); and (3) example metadata for one selected concept map (seismology). Users select which resources they wish to work on. Each user or group of users sees a different view onto the underlying knowledge depending on their interests and memberships

4.1 Changing Perspective: Moving Concepts Between the Foreground and Background

Humans employ an attention mechanism to deal with the vast amounts of conceptual knowledge we each carry, in Gestalt terms the knowledge of immediate interest is said to be in the foreground and the less relevant details are abstracted away into the background. Semantic neighborhoods are supported graphically within the Codex concept browser by setting a 'locality' measure that is applied to the concept currently in focus (the object of inquiry). Setting the locality measure of '1' shows only the set of concepts

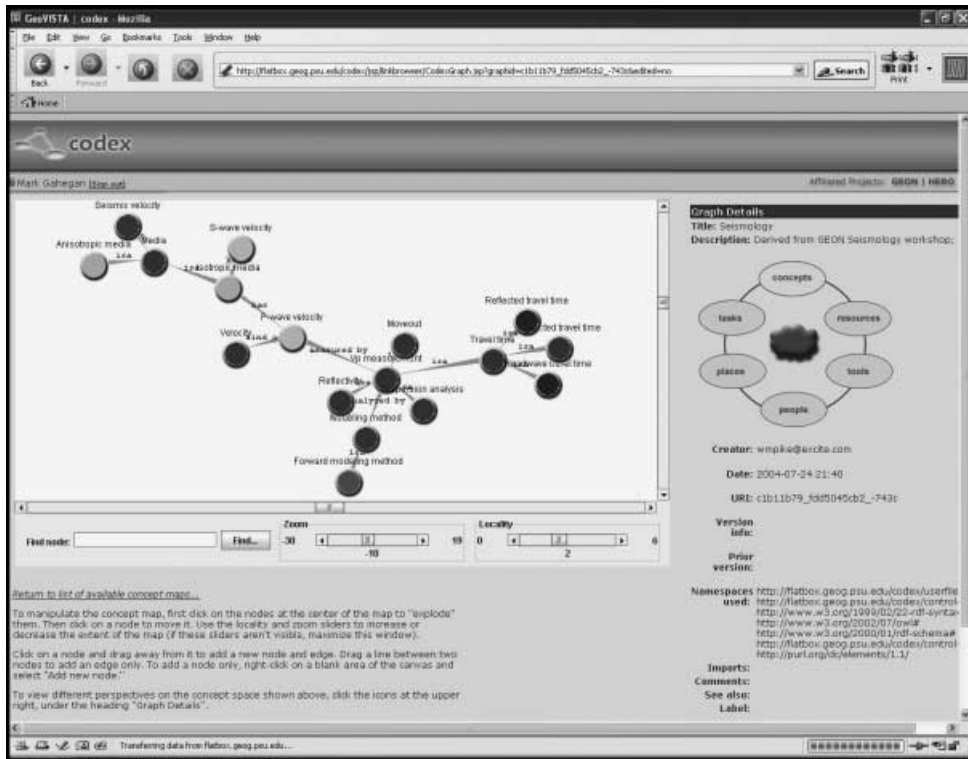


Figure 3 The Codex web portal, showing an ontological perspective defined onto the concept of Seismic Velocity. Concepts are depicted as circles of varying colors and relationships between them are represented by the links. The nexus, by which users change perspectives, appears in the top right of the portal; important information about the concept currently in focus appears below the nexus

directly connected to this object via some relation; a locality of '2' shows this set of concepts, along with all the concepts to which they themselves are directly connected, and so forth. As the user changes focus to another concept, the concept map changes dynamically to bring the selected node into the semantic foreground and expand its neighborhood by the locality setting, whilst at the same time contracting the neighborhood of the previous focal point.

Within this neighborhood another device operates to remove certain kinds of relationships – a *perspective filter* – that projects out certain conceptual structures into the conceptual foreground and hides others. Although a perspective filter can be arbitrarily complex, we specifically define filters for the six perspectives outlined above for convenience of moving quickly between them. Figure 3 shows our limited nexus displayed graphically.

To give an example from within GIS practice, it is common to associate metadata with resources such as datasets. These metadata might include the names of people and organizations involved in the production of the resources, scale applicability, date, and so forth. So, before using a map of land cover change we can check these metadata attributes to ensure compatibility with other data we intend to combine it with.

But when questioning the social network that describes the prior use of the map, these metadata attributes themselves become the concepts at the center of our investigation: for example, the concept of a person or organization, their motivations and intentions, and the original concepts become signifiers of the social dynamics. We call this shift of attention a *perspective*, and have designed Codex to support the definition and application of many alternative perspectives onto the same conceptual knowledge base, providing users with many ways by which to explore and come to understand the resources and concepts it contains.

4.2 Examples of Changing Perspective

In the examples that follow we demonstrate a variety of perspectives that can be constructed for a given map resource, including: a conceptual (ontological) understanding of what the map portrays, the map's user community, the methods, datasets and work-flows that together define how the map was made, and the tasks for which the map has been used (see Figure 4). Each of these perspectives provides a different take on how the map can be understood, supporting a range of quite different questions.

To give a concrete example, the concept of seismic velocity – the rate at which a seismic wave travels through geological structures – is of interest to researchers within the Geosciences Network (GEON) project mentioned above. Seismic velocity can be conceptualized, modeled and applied in a variety of settings, and these different perspectives convey aspects of meaning that might be useful in different settings. Figure 3 shows the conceptual understanding of Seismic Velocity as understood in an ontological sense by a group of expert GEON researchers.³ It also shows the Codex tool in operation, with our limited nexus at the upper right hand side, from which users can shift through the various available perspectives.

Four additional perspectives are shown in Figure 4, all centered on the same seismic velocity concept, but highlighting different aspects of its meaning and use. The Codex web portal interface is cropped out of these figures to save space. Figure 4a shows the resources used to create a seismic velocity map, Figure 4b the user community of seismic velocity – its social network. Figure 4c describes the tasks in which seismic velocity plays a part, while Figure 4d shows the various tools that seismic velocity employs in its calculation. The user navigates from one perspective to the other simply by clicking on one of the bubbles shown in the nexus control in Figure 3.

4.3 Perspectives and Semiotics

The above transitions between perspectives can also be considered from a semiotic point of view. Traditional GIS use fixed sets of sign systems to convey their objects, but in Codex the relationship between a sign vehicle (representamen) and the object it represents is more flexible, allowing concepts, resources and situations to play interchangeable roles, by taking on different semiotic positions in the triad that links the object of inquiry (Figure 5), the means by which the object is signified and the interpreted meaning we draw from it (Ogden and Richards 1927, Sowa 2000b). The Codex representation of a file, for example, can be both representamen and object. In the former case, the file is a signifier for a set of resources in its extension (derivative data, perhaps), or its user community (using this file is a characteristic they all share). In the latter, the file itself is an extensional example for an abstract concept (such as seismic velocity

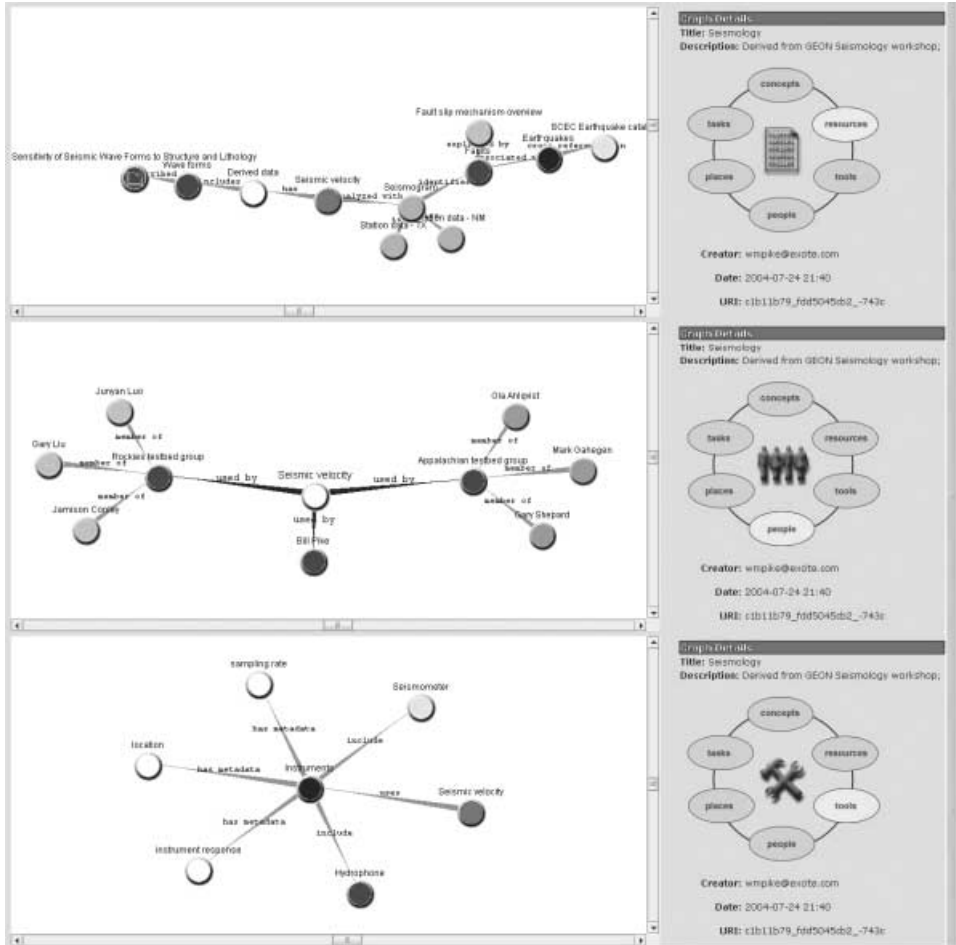


Figure 4 Three additional perspectives defined around the concept of seismic velocity: (top) resources used to define seismic velocity; (middle) the user community of seismic velocity; (bottom) the various tools that seismic velocity employs in its calculation. The right hand panel in each perspective shows the nexus control as it appears in the web browser. The user changes perspectives by clicking on the icons around the nexus

above). Through reification – treating an abstraction as if it were just a simple concept – even the interpretant can be used as representamen or object in another concept (Baker 1999). Thus a use-case, for instance, can be treated as a signifier in a meta-cognitive act. Having produced some new knowledge, a researcher might want to say something about it; reification of interpretant makes it possible to “make a statement about a statement.” But it is important here to also support different interpretations of meaning, so Codex makes it possible for researchers to describe different situations for the same resources, which can help to identify boundary objects useful for making associations between perspectives (Harvey and Chrisman 1998).

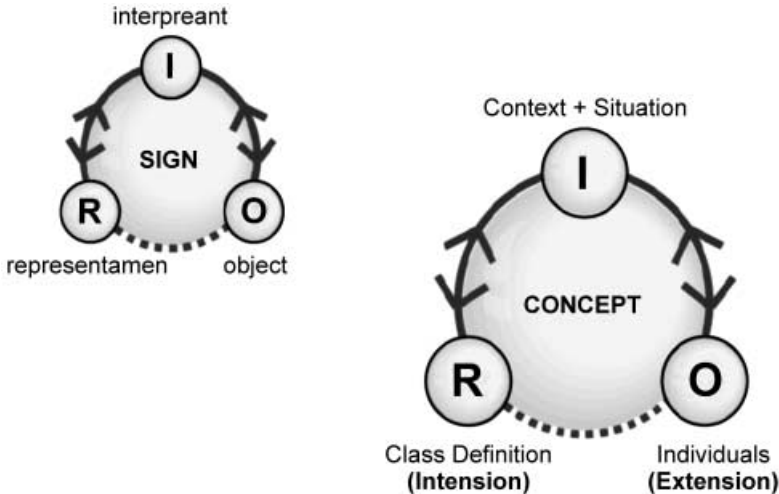


Figure 5 Semiotic view of concept structure in Codex. Because of the interconnectedness of concepts in Codex, they can play different roles (assume different positions) in the semiotic triad depending on whether they are being viewed as an instance of some concept, or as a signifier of some aspect of a different concept

4.4 Examples of Supporting Different Points of View

This next set of examples, shown in Figure 6, derives from a Research Experience for Undergraduates (REU) workshop held as part of the HERO project. The four concept maps shown below were created by teams (separate Codex communities), working remotely, each with a different emergent view of how the concept of transportation planning in central Pennsylvania could be understood, given the planned development of a new interstate highway (I-99). The four views show some commonalities and differences: (a) emphasizes effects on water quality, whereas (b), (c), and (d) show connections to land use planning; (b) and (c) both connect the development of the highway with suburbanization; only (d) notes any historical connections.

5 How it Works: A Description of the Codex Implementation

Codex represents every resource and conceptual structure as some kind of Concept object (upper case C) using the Web Ontology Language (OWL). So each of the six types listed in Section 4 (users, concepts, data, tools, places, and tasks) are specializations of this Concept object. Concepts each have properties such as their name, a text description of what they represent, and other attributes that users may add to suit their purpose. Where a Concept represents an instance of a resource such as a dataset or a map, then the Concept description can contain (as a property) a Uniform Resource Identifier (URI) that links to the physical resource somewhere in cyberspace.

The definitions of Concepts and their connections with one another are stored in a relational database on the Codex server. And whereas it is current practice in ontology

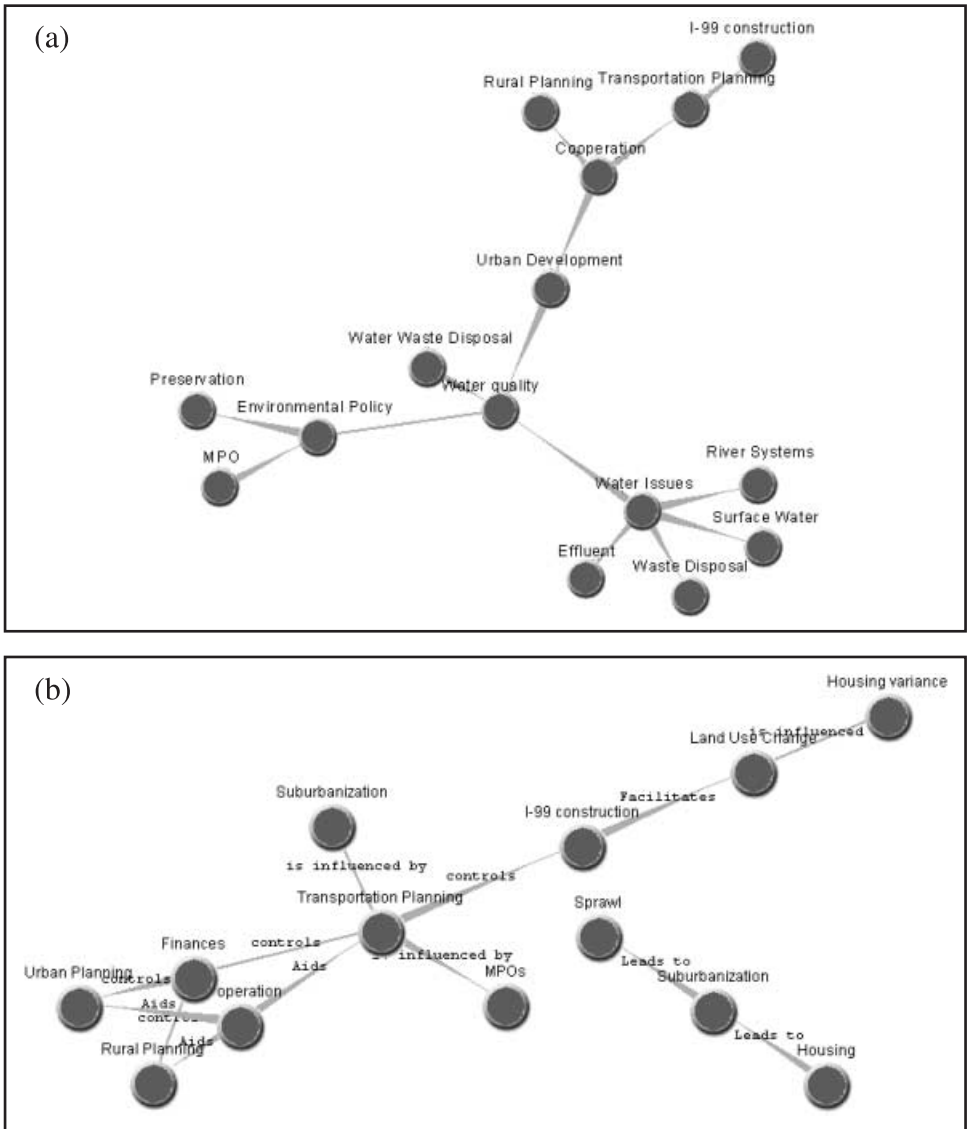


Figure 6 Four different interpretations on the same concept of land use planning, as taken by four different Codex communities. See text for full details

representation to use a triple to define a relationship between two Concepts <concept_A, relation_X, concept_B> here we extend this notion to deal with a multiplicity of possible relationships that are not always true, or may change through time, or may differ from one researcher to another. So each triple is defined within a *situation*, which is created by Codex each time Concepts are linked together into a larger structure to achieve some goal. A situation is thus a wrapper that surrounds an activity. This allows us to state that <concept_A, relation_X, concept_B> applies in some situation(s) but not in others.

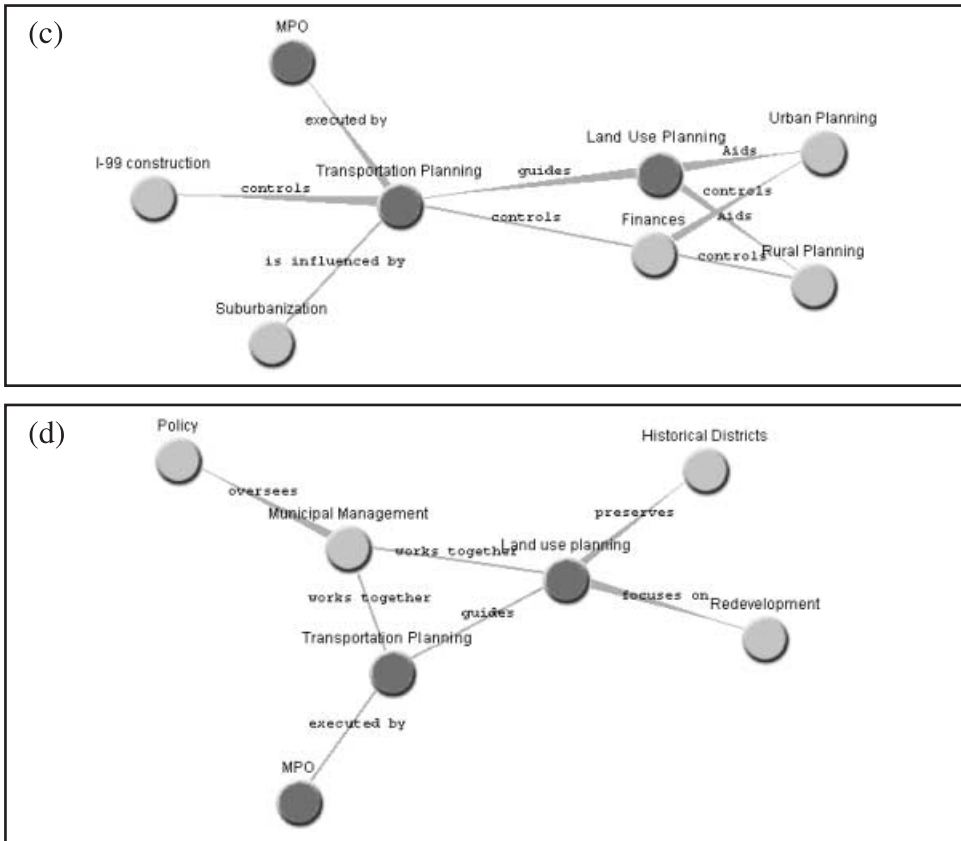


Figure 6 Continued

Some properties of a Concept can be unique to a given situation, allowing the same basic Concept to be re-purposed for a specific task or role. It also allows for two or more alternative representations of the same Concept by different communities of individuals. So for example a community of scientists may describe global climate change as a notion caused in part by anthropogenic factors and likely to have a devastating effect on various ecosystems whereas others may define global climate change as a local anomaly in the climate record of limited consequence. Both may share the same definition of global climate change as: ‘periodic fluctuations in the Earth’s average temperature and climate’.

We can divide the properties of a Concept into two sets: properties that are invariant metadata, such as when it was created, how it was created and by whom – we refer to this as the *context*, and properties that may change depending on the role that a Concept is currently playing – its *situation*. Codex does not represent ontologies (or social networks) as its basic unit of information, but rather deals with individual Concepts and with knowledge fragments that relate one Concept to another by a typed relation. The top-down structure of ontologies is at odds with the more emergent view of science as a network of connections structured from the bottom up and does not readily support the idea of alternative or contested meaning that may occur in some settings. In Codex,

an ontology would be defined as a situation of community agreement surrounding a set of concepts and the relations that join them (filtering out all the files, tasks, users, tools and places).

Concepts may be linked together explicitly in the Codex browser (via a user action) or implicitly captured from workflows and other such structures imported directly from the computational systems, such as GIS, being used (something we hope to be able to achieve in the future). Situations surrounding such activities can be examined after their creation, as a codified memory of the research process, or as a way for one researcher to better understand the work of another. So a researcher who wishes to make use of a dataset registered in Codex, but is uncertain as to how best to utilize it, could query for situations surrounding the past use of the dataset, to gain insight into how others have used it previously. The same researcher might also want to see if any colleagues have used it previously, or whether it has any record of being used for the current task at hand. These situational knowledge fragments, created as Concepts are used and combined, can leave a trail that is useful to researchers who come along later.

A *perspective* is then a set of situations, filtered according to certain conditions to emphasize some aspects of meaning whilst hiding others. Thus the perspective shown in Figure 4b (the user community of the seismic velocity concept), can be formed by selecting all situations encompassing seismic velocity and filtering out those that do not involve users. For ease of use we pre-define some perspectives as Figure 3 shows. Others perspectives can be defined and applied as needed. Perspectives are implemented as queries against the relational database held on the server, with the resulting set of Concepts and relations sent back to the Codex browser.

One other interesting point about perspectives is that by filtering situations in various ways (using different relational operators) we can explore for agreements and conflicts within a knowledge community. For example we can define a perspective as being:

- All situations involving a concept (the union) – to examine the breadth of understanding in a community;
- Those situations which are shared among a set of users (the intersection) – to look for commonalities on which to build a community consensus; or
- Those that are unique (the complement of the above) – to show areas of discord or individuality.

Figure 7 is a simplified OWL snippet that illustrates how a typical resource in Codex blends semiotic aspects of intension, extension, and context. The long alphanumeric strings that begin with the letter *c* uniquely identify each resource, property, and version thereof; coupled with the storage protocol in Codex, this string provides a globally unique path through which any application can directly access any unit of information in Codex.

In this example, the resource in gray is an abstract concept (i.e. a direct member of the Codex class *Concept*) called “Earthquake risk”. This title and accompanying description offer syntactic meaning. The structure of the concept encodes its semantics. Here, earthquake risk is intensionally defined through a single property, which is a cast of a “Distance decay” concept defined elsewhere in the user’s workspace. This decay function has been given an alternate title to use when it acts as a property: “varies-WithDistance” supplies a verb form for an otherwise nominative concept. The target of this property is a particular instance of the concept “Geographic area” (“Fault zone” is in the extension of things that are geographic areas). Including only intensional properties

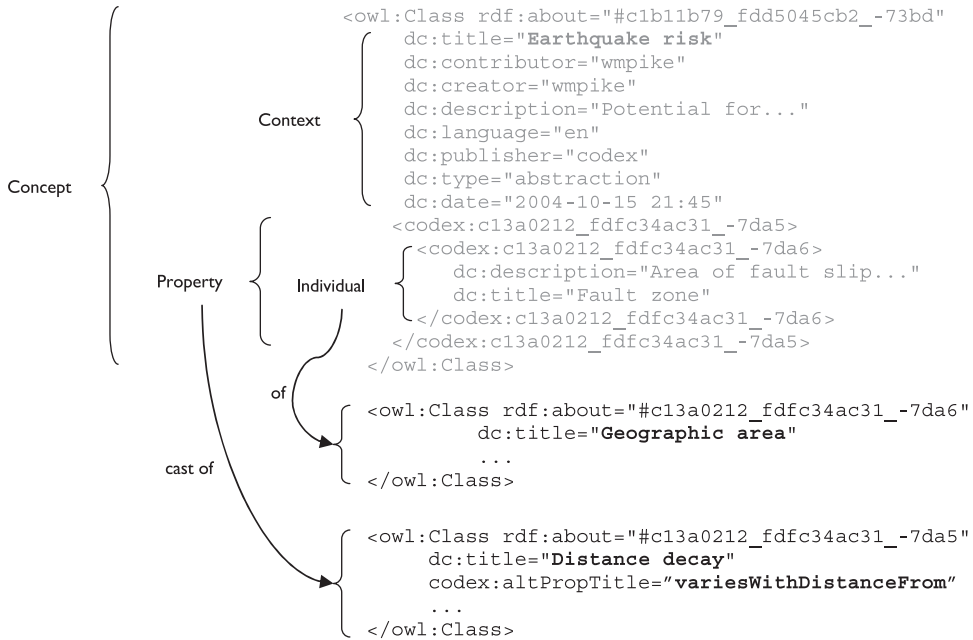


Figure 7 Simplified OWL implementation of a representative concept in Codex. See text for a full explanation

in Codex resource definitions creates parsimonious knowledge representations. If resource *A* relates to resource *B* such that *B* is part of *A*'s intension ($A \rightarrow B$), *A* is the relationship's domain and *B* is its range. The Codex definition of *A* includes only those properties for which *A* is the domain. In the example of Figure 7, "Earthquake risk" is the domain and "Fault zone" is the range of the relationship "variesWithDistance". There may be a resource *C* for which *A* is range ($C \rightarrow A$), but *A*'s representation is not aware of this relation. For instance, there may be a concept "Insurance rate" that is defined in part through a relationship to "Earthquake risk". But because "Insurance rate" is not an intensional part of "Earthquake risk" ("Earthquake risk" is the range, not the domain, of this relation), Codex does not encode insurance as part of risk. The alternative is to encode each relationship twice, once in the domain and once in the range, but Codex avoids this to reduce the likelihood of an inference engine entering a loop while traversing a resource set.

6 Conclusions and Further Work

Although the social critique of GIS is now well-established, little has changed within the inner workings of GIS to address the key issues raised; for example by offering support for alternative perspectives. However, the last several years have seen great strides made towards the computational infrastructure needed to provide access to digital resources; thanks to the success of the Internet, digital libraries and warehouses, web portals and grid computing, resources are widely available and can be downloaded as needed. The

greatest barrier to uptake and reuse now is not availability or access, but our ability to describe the products of our endeavors (resources such as learning activities, datasets, methods, and workflows) so that they can be identified and used effectively and appropriately by others. We must develop successful ways to communicate what the GIS resources we use mean to us. Ontologies and metadata certainly help address this problem, but it is our contention that the meaning, appropriate use, and roles played by resources can be conveyed most effectively by combining a variety of approaches that contextualize resources from different perspectives (Zaff et al. 1993). Such a strategy would integrate use cases, usage patterns and user communities with concept maps, ontologies and other more formal semantic structures. The examples shown above demonstrate one way to tackle this difficult problem.

We must also bear in mind that what resources mean to us may not be the same as they mean to others, for many legitimate reasons. Consequently we argue here for the support of alternative conceptualizations of the geographic concepts and resources that we compute with, and connecting them to a variety of situational knowledge in order to help the geographer reach a deeper understanding – if they so choose. To this end we have demonstrated support for the capture and reuse of situations, and the definition of alternative perspectives, based around these situations. The representations shown are not perfect or complete, but they do provide additional insight that has potential to redress some of the critiques of GIS as promoting an objectivist, positivist approach to geography (whether intentional or not).

Like all computational systems, Codex is not above abuse. To be of use to a community, it requires some kind of central placement among that community, so it raises the question as to whether communities will indeed place their trust in such a tool. But if they do, it does empower a community with the means to create and manage their own knowledge resources, as opposed to having to accept ontologies handed down from elsewhere. Codex communities currently self-identify; access to the knowledge and data contained in a shared workspace is controlled by that community. It is certainly possible to grant public access to these knowledge spaces as well, although professional culture barriers would have to be addressed. The current solution does at least provide a measure of audibility to analysis results. There is also the possibility that widespread access to Codex workspaces could allow closer interaction between researchers and the public. Both communities could engage in exploration of each other's knowledge structures and understand how observations about a problem of shared interest connect to concepts and beliefs.

As noted above, we intend to extend Codex to represent (aspects of) additional elements from our knowledge nexus. Some of these elements are very abstract indeed, and do not easily lend themselves to a formal representation in description logic. Indeed it seems to make most sense currently to represent them in the words of the user, simply as text descriptions. If users maintain a profile that contains descriptions of theories being applied and methodological and philosophical standpoints taken (which they are prompted to revise at regular intervals) then that would enable these elements to be added to new situations automatically. Other situational elements relating to more immediate properties of the investigation, such as the hypothesis under investigation or immediate goal, must be provided by the user at the concept's time of creation. Another possibility is to try to infer such aspects later from how conceptual resources are used in practice, though this by no means a straightforward undertaking. We also intend to investigate alternative ways of visualizing concepts and other resources, and to try to support the idea of changing perspectives with graphical animation.

Notes

- 1 A more pithy version, also attributed to Heraclitus, is: “*You cannot put your foot in the same stream twice*”.
- 2 As an aside, C.S. Peirce is sometimes credited with the observation that: “*The textbook is the funereal urn of a discipline.*” Perhaps the modern equivalent of the textbook for us is the computational ontology, for without the ability to evolve and adapt in response to conceptual changes, ontologies will effectively lock us into the “static, lifeless, purposeless world” (Sowa 2002) in which discovery, contested meaning and re-conceptualization are not facilitated.
- 3 These details were gathered at a GEON workshop dedicated to the creation of ontological knowledge underpinning geoscience resources.

References

- Agarwal P 2004 Contested nature of place: Knowledge mapping for resolving ontological distinctions between geographical concepts. In Egenhofer M, Freksa C, and Miller H (eds) *Geographic Information Science: Proceedings of the Third International Conference (GIScience 2004)*. Berlin, Springer-Verlag Lecture Notes in Computer Science No 3234: 1–21
- Baker V 1999 Geosemiosis. *GSA Bulletin* 111: 633–45
- Barsalou L and Medin D 1986 Concepts: Static definitions or context-dependent representations? *Cahiers de Psychologie Cognitive* 6: 187–202
- Boisvert R and Rice J 1996 From scientific software libraries to problem solving environments. *IEEE Computer Science and Engineering* 3: 44–53
- Brodaric B, Gahegan M, and Harrap R 2004 The art and science of mapping: Computing geological categories from field data. *Computers and Geosciences* 30: 719–40
- Buckingham Shum S, Motta E and Domingue J 2000 ScholOnto: An ontology-based digital library server for research documents and discourse. *International Journal on Digital Libraries* 3: 237–48
- Bush V 1959 *Memex II*. Cambridge, MA, MIT Archives, Vannevar Bush Papers, MC78, Box 21
- Carver S, Evans A, Kingston R and Turton I 2001 Public participation, GIS and cyberdemocracy: Evaluating on-line spatial decision support systems. *Environment and Planning B* 28: 907–21
- Clancey W 1994 Situated cognition: How representations are created and given meaning. In Lewis R and Mendelsohn P (eds) *Lessons from Learning*. Amsterdam, North-Holland: 231–42
- Davis F W and Simonet D S 1991 GIS and remote sensing. In Maguire D J, Goodchild M F, and Rhind D W (eds) *Geographical Information Systems*. Harlow, Longman: 191–213
- de Moor A, Keeler M, and Richmond G 2002 Towards a pragmatic web. In Priss U, Corbett D, and Angelova G (eds) *Conceptual Structures: Integration and Interfaces: Proceedings of the Tenth International Conference on Conceptual Structures (ICCS 2002)*. Berlin, Springer-Verlag Lecture Notes in Artificial Intelligence No 2393: 235–49
- Fensel D 2001 *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*. New York, Springer
- Feyerabend P 1988 *Against Method*. London, Verso
- Finholt T A 2001 Collaboratories. *Annual Review of Information Science and Technology B* 36: 73–108
- Fischer G and Ostwald J 2001 Knowledge management: Problems, promises, realities, and challenges. *IEEE Intelligent Systems* 16: 60–72
- Fonseca F T, Egenhofer M J, and Agouris P 2002 Using ontologies for integrated Geographic Information Systems. *Transactions in GIS* 6: 231–57
- Goss J 1995 Marketing the new marketing: The strategic discourse of geodemographic information systems. In Pickles J (ed) *Ground Truth: The Social Implications of Geographic Information Systems*. New York, Guilford: 130–70
- Gruber T R 1993 A translation approach to portable ontologies. *Knowledge Acquisition* 5: 199–220
- Gruber T R 1995 Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies* 43: 907–28

- Guarino N 1997 Understanding, building, and using ontologies. *International Journal of Human-Computer Studies* 46: 293–310
- Harvey F 2000 The social construction of geographical information systems. *International Journal of Geographical Information Science* 14: 711–3
- Harvey F and Chrisman N 1998 Boundary objects and the social construction of GIS technology. *Environment and Planning A* 30: 1683–94
- Kuhn T S 1962 *The Structure of Scientific Revolutions* (Second edition). Chicago, IL, University of Chicago Press
- Latour B 1987 *Science in Action*. Cambridge, MA, Harvard University Press
- Marcos E and Marcos A 2001 A philosophical approach to the concept of data model: Is a data model, in fact, a model? *Information Systems Frontiers* 3: 267–74
- McHaffie P H 1995 Manufacturing metaphors: Public cartography, the market and democracy. In Pickles J (ed) *Ground Truth: The Social Implications of Geographic Information Systems*. New York, Guilford: 113–29
- Michalski R S 1993 Beyond prototypes and frames: Two-tiered concept representation. In Mechelen I V, Hampton J, Michalski R S, and Theuns P (eds) *Categories and Concepts: Theoretical Views and Inductive Data Analysis*. New York, Academic Press: 173–202
- Myers J, Mendoza E, and Hoopes B 2001 A collaborative electronic notebook. In *Proceedings of the IASTED International Conference on Internet and Multimedia Systems and Applications*, Honolulu, Hawaii
- Nake F and Grabowski S 2001 Human-computer interaction views as pseudo-communication. *Knowledge-Based Systems* 14: 441–7
- National Center for Geographic Information and Analysis (NCGIA) 1998 Initiative 19: GIS and Society. WWW document, <http://www.geo.wvu.edu/i19/>
- Noth W 1990 *Handbook of Semiotics*. Bloomington, IN, Indiana University Press
- NSF 2003 *Revolutionizing Science and Engineering through Cyberinfrastructure*. Washington, DC, National Science Foundation
- Ogden C and Richards I 1927 *The Meaning of Meaning*. New York, Harcourt Brace
- Peirce C S 1877 Fixation of belief. *Popular Science Monthly* 12(November): 1–15
- Peirce C S 1905 What pragmatism is. *The Monist* 15(2): 161–81
- Peirce C S 1931 *The Collected Papers of Charles Sanders Peirce*. Cambridge, MA, Harvard University Press
- Pickles J 1995 Representations in an electronic age: Geography, GIS, and democracy. In Pickles J (ed) *Ground Truth: The Social Implications of Geographic Information Systems*. New York, Guilford: 1–30
- Pickles J 1999 Arguments, debates, and dialogues: The GIS-social theory debate and the concern for alternatives. In Longley P A, Goodchild M F, Maguire D J, and Rhind D W (eds). *Geographical Information Systems: Principles, Techniques, Applications, and Management* (Second edition). New York, John Wiley and Sons: 49–60
- Popper K 1959 *The Logic of Scientific Discovery*. London, Hutchinson
- Reschner N 1978 *Peirce's Philosophy of Science*. London, Notre Dame
- Rodriguez M A, Egenhofer M J, and Rugg R 1999 Assessing semantic similarities among geospatial feature class definitions. In Vckovski A, Brassel K E, and Schek H-J (eds) *Interoperating Geographic Information Systems*. Berlin, Springer Lecture Notes in Computer Science No 1580: 190–202
- Sanchez J and Langley P 2003 An interactive environment for scientific modeling and discovery. In *Proceedings of the International Conference on Knowledge Capture*, Sanibel Island, Florida
- Schultze U and Boland R J 2000 Knowledge management technology and the reproduction of knowledge work practices. *Journal of Strategic Information Systems* 9: 193–212
- Schuurman N 2000 Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography* 24: 569–90
- Schuurman N 2002 Reconciling social constructivism and realism in GIS. *ACME: An International E-Journal for Critical Geographies* 1: 73–90
- Sowa J 2000a *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA, Brooks/Cole
- Sowa J 2000b Ontology, metadata, and semiotics. In Ganter B and Mineau G (eds) *Conceptual Structures: Logical, Linguistic, and Computational Issues*. Berlin, Springer-Verlag Lecture Notes in Computer Science No 1867: 55–81

- Sowa J F 2002 Signs, Processes and Language Games: Foundations for Ontology. WWW document, <http://www.jfsowa.com/pubs/signproc.htm>
- Schweizer T 1997 Embeddedness of ethnographic cases. *Current Anthropology* 38: 739–60
- Taylor P 1990 Editorial comment: GKS. *Political Geography Quarterly* 9: 211–2
- Takatsuka M and Gahegan M 2002 GeoVISTA *Studio*: A codeless visual programming environment for geoscientific data analysis and visualization. *Computers and Geosciences* 28: 1131–44
- Veregin H 1995 Computer innovation and adoption in geography: A critique of conventional technological models. In Pickles J (ed) *Ground Truth: The Social Implications of Geographic Information Systems*. New York, Guilford: 88–112
- Whitehead A 1929 *Process and Reality: An Essay in Cosmology*. New York, Social Science Book Store
- Whitehead A N 1933 *Adventures of Ideas*, New York, Macmillan
- Whitehead A N 1938 *Modes of Thought*, New York, Macmillan
- Zaff B, McNeese M, and Snyder D 1993 Capturing multiple perspectives: A user-centered approach to knowledge and design acquisition. *Knowledge Acquisition* 5: 79–116