

# Connecting GEON: making sense of the myriad resources, researchers and concepts that comprise a geoscience cyberinfrastructure

Mark Gahegan<sup>\*ab</sup>, Junyan Luo<sup>b</sup>, Stephen D. Weaver<sup>b</sup>, William Pike<sup>c</sup>, and Tawan Banchuen<sup>b</sup>

\*Corresponding Author

<sup>a</sup>School of Geography, Geology and Environmental Science, University of Auckland, New Zealand.

<sup>b</sup>GeoVISTA Center, Department of Geography, Pennsylvania State University, USA.

<sup>bc</sup>Pacific Northwest National Labs, Richland, WA, USA.

Email: [m.gahegan@auckland.ac.nz](mailto:m.gahegan@auckland.ac.nz); Tel: +64 9 373-7599 ext. 88061; Fax: +64 9 373-7434

## Abstract

Simply placing electronic geoscience resources such as datasets, methods, ontologies, workflows and articles in a digital library or cyberinfrastructure does not mean that they will be used successfully by other researchers or educators. It is also necessary to provide the means to locate potentially useful content, and to understand it. Without suitable provision for these needs, many useful resources will go undiscovered, or else will be found but used inappropriately. In this article we describe an approach to discovering, describing and understanding e-resources based on the notion that meaning is carried in the interconnections between resources and the actors in the cyberinfrastructure (including individuals, groups, organizations), as well as by ontologies and conventional metadata. Navigation around this universe is achieved by implementing the idea of *perspectives* as dynamic, conceptual views (defined by SPARQL-like queries against an OWL schema) that not only act as filters, but also dynamically promote and demote concepts, relationships and properties according to their immediate relevance. We describe a means to represent a wide variety of interactions between resources using the notion of a *knowledge nexus*, and we illustrate its use with resources and actors from the GEON cyberinfrastructure community. We also closely link browsing and visualizing strategies to our nexus, drawing on ideas from semiotics to move resources and connections not currently of interest from the foreground to the background, and *vice versa*, using a new form of adaptive perspective. We illustrate our ideas via *ConceptVista*, an open-source concept mapping application that provides rich, visual depictions of the resources, cyber-community and myriad connections between them. Examples are presented that show how geoscientific knowledge can be explored not only via ontological structure, but also by use cases, social networks, citation graphs and organization charts; all of which may carry some aspects of meaning for the user.

Keywords: cyberinfrastructure, knowledge browsing, ontology, GEON, semantics, perspectives.

## 1 Introduction

Cyberinfrastructure (CI), or e-science as it is known in Europe, is beginning to have an impact across many scientific communities, offering access to remote data collections, instruments and high performance computing facilities (NSF, 2003; Hey and Trefethen, 2003). But in addition to

simply providing access to remote (e-)resources, cyberinfrastructure also offers many possibilities for improving collaboration and fostering understanding between geographically separated (non-co-located) colleagues (Finholt, 2001; MacEachren *et al.*, 2006; Brennan *et al.*, 2006, Brodaric and Gahegan, 2006). And while the more technical levels of cyberinfrastructure are by now quite advanced in terms of middleware for grid computing (e.g. [www.globus.org/](http://www.globus.org/)), the more semantic and human-oriented aspects of supporting computer-enabled, collaborative science (i.e. science involving multiple researchers not necessarily co-located nor working synchronously, Chin *et al.*, 2002) are less well developed and operationalized. The ultimate success of cyberinfrastructure depends on both of these aspects being fully realized (NSF, 2003; Berman and Brady, 2005). For example, a dataset or method contributed by a researcher into a cyberinfrastructure can only be used effectively by others if it is understood in sufficient detail to ensure the correct application and interpretation of results.

This paper describes our efforts to build geoscientific knowledge representations that can be used to facilitate community access to cyberinfrastructure resources via a knowledge portal. The ideas we describe, and the accompanying implementation in an open-source application (called *ConceptVista*) were developed to support collaborative science for parts of the geoscience and GIScience research communities, and for shared development of e-learning activities (Gahegan *et al.*, 2007). As we will see later, *ConceptVista* represents an amalgam of ideas from concept mapping (Novak, 1998), ontology (Guarino, 1998) social network and citation mapping (National Academies of Sciences, 2004), workflow representation (Cardoso and Sheth, 2003), and provenance computing. It is our contention that all of these approaches have merits to offer in the representation and communication of meaning, and that they need to be carefully integrated into a single framework. More formal knowledge—with no internal inconsistencies and strong typing can be held in ontological form (using OWL), whereas less formal knowledge, (the kind typically held in concept maps or captured as provenance) can be held in RDF form, where there may be inconsistencies or weak (even absent) typing. *ConceptVista* supports both formats, even within the same display.

To illustrate our arguments here we use the GEON project (Geosciences Network: [www.geongrid.org](http://www.geongrid.org)<sup>1</sup>) as our case study and demonstrate our ideas via many examples of the functionality described. We believe these examples demonstrate the utility of the ideas described, but note that further studies will be required in the form of user evaluations to establish their ultimate worth. As geoscience cyberinfrastructure expands, we must carefully consider—as a community—how we will facilitate: (i) discovering useful resources from the collections available, (ii) understanding these resources, and (iii) knowing how to use them appropriately.

## 1.1 Problems of supporting access to massive collections of e-resources

Since the collections of resources in digital libraries and cyberinfrastructures can become very large, strategies for dealing with a dynamic catalog, and for users to quickly *navigate* through the resources and define *perspectives* of interest are needed. These strategies must help users focus

---

<sup>1</sup> GEON is a large cyberinfrastructure project funded by the National Science Foundation in the USA to support the geoscience community. The GEON network now connects computers and institutions across the USA, and even into Asia. GEON has, at the time of writing, a growing collection of approximately 5100 datasets, in addition to many services and ontologies in its collections.

on aspects of knowledge and associated resources that are of current interest to them and thereby ignore vast areas of digital collections and associated knowledge that are not. Below are four specific challenges that arise from our work in supporting cyberinfrastructure.

### **1. The sheer volume of resources in digital collections**

Digital collections can be large, containing thousands of resources. Although these kinds of volumes cause no problems for databases, they certainly push the limits of what can be expressed with interactive graph-based browsers, as shown later in Figures 1 and 5. Section 2 describes how we have redesigned our concept browser so that it can scale to thousands of resources and connections.

### **2. The dynamic nature of the resource catalogs**

Collections are in constant flux as new resources are added and old ones retired, and as new tools and use-cases are added. So any system for visualizing the catalog must dynamically load its contents at access time. Section 3 describes one strategy to support such dynamic access.

### **3. The need to support multiple search strategies to locate useful resources**

Users of a cyberinfrastructure portal or digital library need to be able to identify potentially useful resources using a variety of search strategies and preferably utilizing their own experience, judgments and (ontological) interests as a starting point. In some situations, users may know by name exactly which resource they wish to access. In others they may only have vague clues, such as the *dataset* used by *person X* in the last *article* they published, or a *method* that will work to visualize *dataset Y*.

### **4. The need to explain what resources mean**

Resources do not explicitly carry knowledge with them of how they were made, nor of how they should be understood, or used. Yet such knowledge is often vital to would-be consumers (Heemskerck *et al.*, 2003). We need to consciously weave a knowledge infrastructure into our portals and catalogs to help address this problem (Stevens, 2003). This knowledge infrastructure should be adaptable to the specific and immediate needs of the researcher. For the GEON community, a view onto the underlying resources is presented via a customizable web portal as described by Nambiar (2006) and available at <https://portal.geongrid.org:8443/gridsphere/gridsphere>.

We consider points 3 and 4 to be closely related, in that providing multiple connections between resources helps to represent meaning and affords alternative search paths to locate useful content. Section 4 illustrates how we tackle these two related problems, describing the philosophical foundations of our ideas, details of their practical implementation and examples of what they can facilitate.

## **1.2 Contribution and connections to previous work**

The representation of knowledge in cyberinfrastructure poses significant challenges. Diverse collections of resources must be described in sufficient detail so that, (i) useful content can be found, (ii) users can make informed decisions about the suitability of resources for their tasks, (iii) users (or collection administrators) can keep track of what they do. Moreover, the things we may be interested in, and the nature of that interest, changes—sometimes often. Zaff *et al.* (1993) discuss the need to support multiple perspectives; Novak and colleagues (2003) describe a method that uses personal concept taxonomies to map between different perspectives on the

structure of a document space; Haase *et al.* (2005) use personal ontologies to describe perspectives onto bibliographic data, and also propose collaborative filtering to support management and evolution of these ontologies by suggesting changes based on other users' personal structures. McGuinness and da Silva (2003) show how the provenance of knowledge can be supported using semantics. Novak and Wurst (2004) describe the importance of capturing and sharing knowledge in collaborative work.

In our own previous work (Pike *et al.*, 2003; Pike and Gahegan, in press; Gahegan and Pike, 2006) we introduced a system called *Codex* that maintains a large collection of *unstructured* knowledge fragments (to support a science e-notebook and distributed knowledge elicitation) that are filtered according to the user's current interest, then visualized for interactive exploration. This was a first attempt at implementing simple perspective filters that we extend here in three important directions. Firstly, we add the capability to manage perspectives onto ontologies in OWL (rather than more amorphous collections of knowledge in XML). Since ontologies have a high degree of internal cohesion and often also community acceptance, it makes sense to manage them as a pooled, conceptual resource. Managing ontologies involves extensions to the notion of perspectives, and we show later how such perspectives are defined (using SPARQL) as logical views onto a knowledge collection. Secondly, and as part of our refined perspective, we develop a mechanism to temporarily promote or demote knowledge fragments between simple properties and concepts in their own right. Thirdly, we add in the ability to explicitly represent relationships between concepts that are implied via indirect connections between them. These ideas represent a much more elaborate model of a perspective, not simply as a filter, but as a way of re-presenting knowledge on the fly to suit specific user-needs, and as a way of recognizing commonalities in knowledge that might be obscured by somewhat arbitrary decisions made during their formation. Perspectives then provide a mechanism to support a more cognitively plausible focus of attention that can shift quickly, moving knowledge between the foreground and background as interests change (e.g. Albrechtsen *et al.*, 2001). Such fluidity is an important aspect of how humans reason with complexity and also connects well with ideas from semiotics that we discuss later.

## 2 Browsing massive collections of resources

Browsing large collections of data and knowledge requires a visualize-on-demand strategy. We assume that in most situations users wish to see only a small portion of the entire set of resources and associated knowledge at any given time. *ConceptVista* therefore only creates displays for those parts that are currently the focus of attention. Thus we avoid the problem of having to visualize the massive collections of concepts and instances that are common in resource catalogs and domain ontologies. When an ontology is initially opened, its entire contents are loaded into the supporting OWL knowledge model (Web Ontology Language: <http://www.w3.org/TR/owl-features/>; McGuinness and Harmelen, 2003) but only the upper-level concepts are displayed; i.e., those which connect to the concept 'thing' or do not have any more general forms. From this starting point, a user can search through the graph of concepts and relations in several ways, such as: (i) by submitting an RDF or OWL query, (ii) via real-time navigation using the expansion and localization functions built into the graph layout algorithm used for display (based loosely on the open-source TouchGraph™ algorithm: [www.touchgraph.com](http://www.touchgraph.com)) or (iii) by using a perspective filter (described below in Section 4). Figure 1 shows a screenshot of the application, with the graph layout on the right. Likewise in the ontology tree view, (shown on the left side of Figure

1) only tree paths currently expanded by the user need to be loaded into the tree structure for display.

As an example of scalability, the entire OpenCyc general ontology ([www.opencyc.org/](http://www.opencyc.org/)), containing over 62,000 terms, is loaded in about two minutes on a modest laptop computer (2GHz Pentium processor with 1GB of RAM). It can be navigated with ease in real time. The browser in Figure 1 contains just a small part of the large AGU geoscience topic map<sup>2</sup> focused on hydrogeology (which we converted to ontological form, adding in cross reference links to form a graph from the original hierarchy: <http://www.agu.org/pubs/indexterms/browse.html>). The entire topic map contains 1200 distinct terms, organized into thematic and regional headings and is already in use within GEON and the wider geoscience community. Note though that even the small fraction of the topic map shown in Figure 1 is difficult to grasp when displayed this way, with all concepts and relations visible; better browsing strategies are needed to hide unnecessary detail. Our solutions to this problem are described in Section 4.

### 3 Dynamically accessing CI resource catalogs

In a large cyberinfrastructure, it is not practical to hold resource catalogs and related metadata locally, as these are large and subject to regular change. Likewise, if use-case data is being captured as the CI is used, then this too will need to be accessed when needed. Some ontologies, concept maps and supporting information may be available locally, but information about CI resources, needs to be accessed remotely. Such catalogs may be stored centrally, or in many cases may be distributed across the CI, and accessed via GRID Services protocols. (In the case of GEON, the GRID services are deployed and managed using a system called Rocks; see Sacerdoti *et al.*, undated). So any tools developed to browse such knowledge must make requests to the appropriate remote catalogs as well as perhaps supplementing with content from local sources.

Like other similar tools for ontology creation and browsing, *ConceptVista* imports a variety of formats developed specifically for the representation and sharing of computational knowledge, such as OWL, RDF and KIFF. If CI catalogs hold information in this form, then they can be imported straightforwardly. But to connect with the GEON catalogs requires a mechanism to translate the current display request (which is usually made to the underlying OWL model) into an SQL database query and to then convert the results of the database query into concepts and relations that can be visualized. Information about GEON resources and community members is stored in two PostgreSQL databases, hosted on two different servers (*geon06* for resources and *geon03* for people). The GEON portal uniquely identifies users in the database by their email addresses. When a user submits a resource to the portal, it assigns a unique identification number to the resource and stores various personal metadata on the *geon03* server, including the contributor's email address. We use JDBC (Java Database Connectivity technology <http://java.sun.com/javase/technologies/database/>) to make remote SQL requests to these two database servers. Since it is more helpful for users to identify GEON personnel by using their names rather than their email addresses, the system has to query against the *geon06* server to first

---

<sup>2</sup> Topic Maps are a kind of domain ontology usually restricted to hierarchy. Their use predates computational ontologies to some extent, and there are well-established tools and standards for their representation and navigation (e.g. see: <http://www.topicmap.com/>).



match users to their email addresses. It then uses these addresses to query against the *geon03* server to retrieve resources that GEON users have contributed to the portal and translates the results back to triple form, as used by OWL.

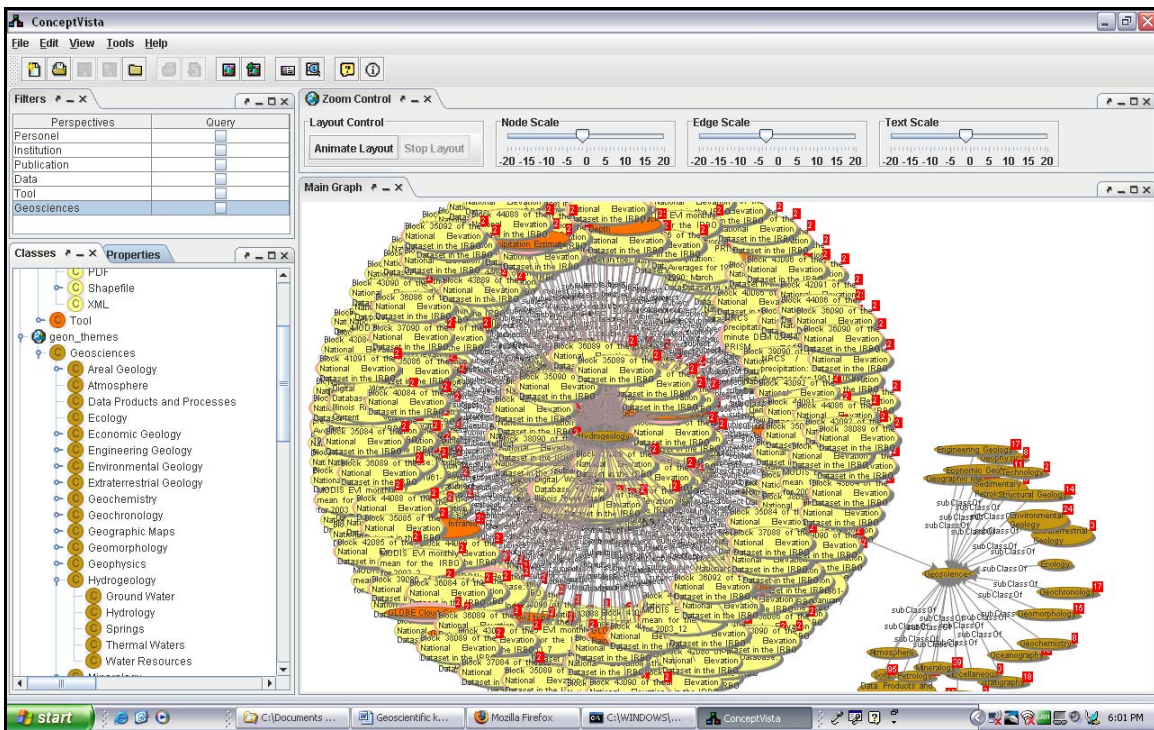


Figure 1. A ConceptVista screenshot showing a small part of the AGU geosciences topic map describing the knowledge realm of hydrogeology (at the center of the large cluster). Other major terms in the topic map are shown at lower right. The left side of the application provides a hierarchical (tree) view onto the same knowledge.

For a given GEON resource, a database query returns its identification number, title, type, subject and description. There is a small latency introduced into browsing by the constant interaction with the remote GEON catalogs, but this way the most current information is constantly available. It is also possible to download the contents of the catalogs once only at load time, if it is safe to assume they will not change for the duration of the session (not a valid assumption if one plans to submit a resource).

The GEON cyberinfrastructure classifies resources into 4 categories and 18 subcategories as shown below:

1. Data: shapefile, database, XML, GeoTIFF, CUAHSI data, GMT raster, ASCII, PDF, Excel, WebURL, NetCDF, comments, PowerPoint and Word.
2. Service: ordinary web service and Web Mapping Service (WMS)
3. Tool: source code and binary file
4. Ontology: ontology and OWL file

Resource descriptions are then translated into RDF triple clauses (against a schema held in *ConceptVista*) and imported, at which time a concept map is constructed. The database catalogs

provide rich metadata about GEON users (names, email addresses, organizations and publications) and resources (titles, topics, descriptions, URLs and contributing users).

The RDF representations constructed from these information fragments can then be used to visualize, for example: (i) all resources a user has contributed, (ii) resources connected to specific geoscience topics, (iii) common areas of interests among users, and many additional relationships as shown later in later figures. Furthermore, using a built-in semantic search engine that connects to various web search and information catalogs such as Google, Amazon, CiteSeer and Wikipedia (Gahegan *et al.*, 2007), external content and descriptions of individuals, organizations, publications and resources can be discovered and added to the RDF representation to provide additional signifiers of meaning for the various concepts represented (e.g. Li *et al.*, 2002). URIs (Uniform Resource Identifiers: <http://gbiv.com/protocols/uri/rfc/rfc3986.html>) are used to link concepts to these additional discovered signifiers, and a native web browser built in to the application using Java Desktop Integration technology (<https://jdic.dev.java.net/>) displays the URI contents. Figure 2 shows an article that has been discovered and then associated with its author, with the article itself appearing in the embedded browser on the left (the figure caption explains how to read the graph, or the tree view shown in Figure 1 provides a key).

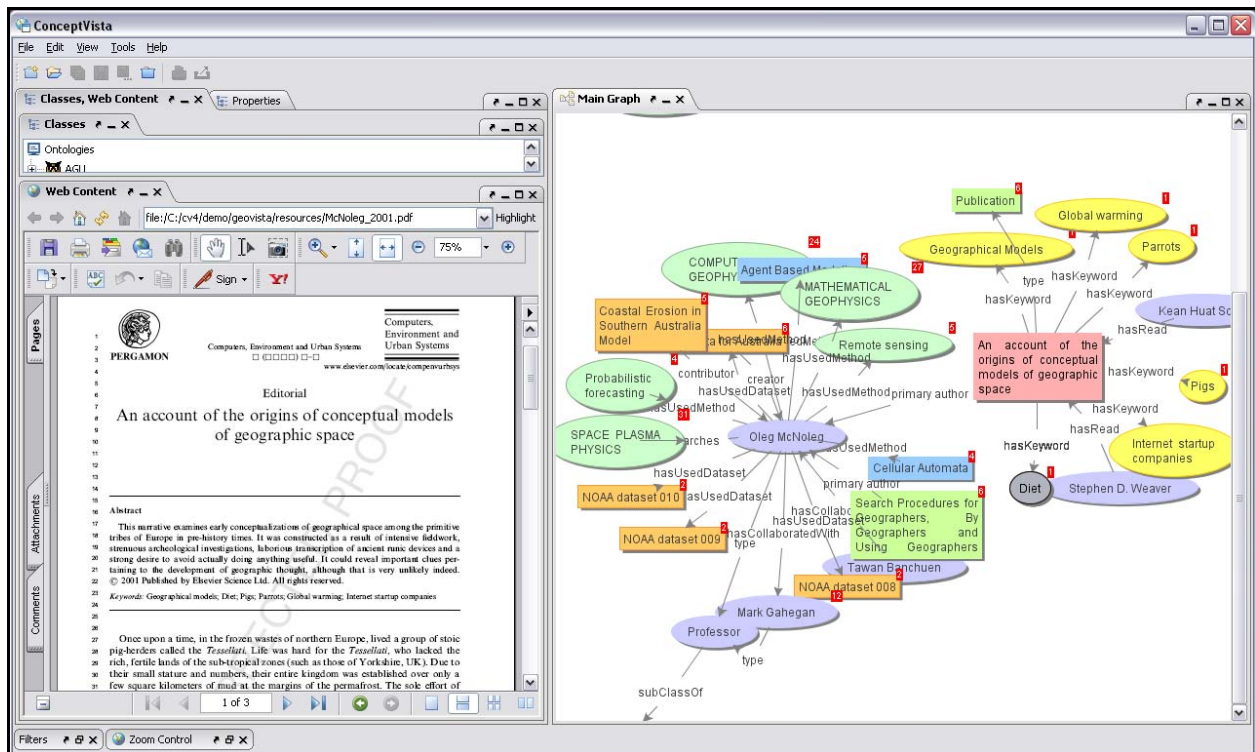


Figure 2. A cluster of interconnections defined around a researcher and an article they authored (“O. McNoleg”, “An account of the origins of conceptual models of geographical space.”) The graph at right shows datasets contributed by the same researcher (orange squares), domain terms of interest to the author (green ellipses) methods used by the author (blue rectangles), keywords relating to the article (yellow ellipses) and its readership (gray ellipses).

## 4 Augmenting ontologies to support multiple access paths and richer semantic descriptions

In this section we present theoretical arguments as to why we believe ontology is not sufficient to describe resources and support the communication of meaning between researchers in a cyberinfrastructure. Note though that the ideas we describe are in part implemented using ontological tools and languages for representing and reasoning with knowledge that includes both situational and ontological aspects. But we also propose enhancements based on (i) a deeper representation of knowledge that integrates situational elements and (ii) *perspective filters* that provide temporary, task-specific views onto this knowledge and the underlying resource collections that it describes.

### 4.1 Are domain ontologies sufficient?

Ontologies describing the upper-level structures of science domains are by now quite well established (Guarino, 1998; Noy and Hafner, 2000; Sugumaran and Storey, 2002). Two well known examples in the geoscience domain are the EarthRealm series developed by NASA/JPL (Raskin and Pan, 2007; <http://sweet.jpl.nasa.gov/ontology/earthrealm.owl>) and the AGU topic map described above. Clearly, resources can be registered to such ontologies as a form of topic indexing. But before relying solely on ontologies to provide thematic access to resources, let us first pause and examine the nature of geoscientific knowledge, to better understand what it is we strive to represent (Reynolds *et al.*, 1996). Ontological knowledge has already been shown to have great potential benefit in many aspects of geoscience research (Fox *et al.*, 2006), and most especially for helping resolve schema-level inconsistencies arising in data integration (Sheth, 1999; Wache, 2001; Rodriguez and Egenhofer 2003). Ontology has also proved useful as a registration and markup language, providing common terms by which resources might be tagged using a consistent set of keywords (e.g. Buckingham-Shum *et al.*, 2000). GEON has successfully used both of these ideas, by providing an ontology-based map integration service (Lin and Ludäscher, 2003) and utilizing a derivation of the AGU topic map as an upper-level ontology by which geoscience themes are tagged. We argue here that such uses of ontology are certainly valuable, but by themselves they do not provide strong support for discovering and understanding CI resources.

Informally, we note in passing that e-shopping vendors such as *Amazon*<sup>TM</sup> use recommender systems (Ansari *et al.* 2000) to help connect their users with potentially useful resources; for example *LivePlasma*<sup>TM</sup> (<http://www.musicplasma.com/>) uses association rules—emerging from purchasing behavior—to build connections between resources, instead of a defined ontology. Kim and Fox (2004) address collaborative filtering for use in digital libraries, grouping users by interests that are defined dynamically; that is, use-cases and provenance information provide a viable alternative to domain ontology in that they capture what users have actually done (and which may at times conflict with what they might say or believe). And while there is no certainty that the connections so formed are semantically reliable, they appear to work well in most cases and prevent users from having to confront a potentially-overwhelming and possibly contentious set of concepts and relationships. We use this example to make the point that use-cases (situations of use) also carry important aspects of meaning if appropriately analyzed. Knowing what was done or used, by whom and for what purpose, and adding this to ontological knowledge allows us to better understand our resources (Solomon *et al.*, 1999; Chin and Lansing, 2004; Carroll, 2005). Situations are important to knowledge representation because



they explicitly reproduce the enactment that is part of selecting and reasoning with a set of concepts or resources (Barsalou, 2002). Pike and Gahegan (2007) argue for the explicit recognition of the *situational* nature of geological knowledge and Brodaric and Gahegan (2007) show experimentally how some ontological categories in the geosciences are themselves situated within a spatio-temporal context: in other words their existence is entangled with some situation, and their meaning explicitly tied to it.

It is well known—metaphysically and practically—that there are limits to philosophical ontology and the understanding it can convey (Lemke, 1997; Frodeman, 2003; Sowa, 2002); and in fact we argue that a good deal of useful geoscientific knowledge is not ontological in nature, but rather situational (Clancey, 1994), and thus more closely aligned with epistemology—i.e. how things are known. In a subject where concepts and relationships are defined as much by interactions between experiments, observations, motivations, methods, researchers, locations and time as they are by theory, the representation of meaning becomes a multi-faceted web of interactions between these different components (theory, data, methods, tools, places, times, people, organizations and so forth) and importantly, this meaning is in flux.

The tools we describe below are designed to support, explore and communicate such situated knowledge—that is, knowledge whose relevance changes according to the situation at hand. At first glance, this idea may seem to add confusion to ontological clarity, and complexity to any interface, but in fact the opposite is true in practice because humans reason using very adaptive, contextual strategies, regularly changing their focus of attention. We draw our inspiration in part from the writings of Alfred North Whitehead, a mathematician and later a philosopher of science in the early half of the 20<sup>th</sup> century. Whitehead's life's work began by searching for formal clarity in harmonization theories for mathematics with his colleague and friend Bertram Russell (Whitehead and Russell, 1997: originally published 1910-13) and ended with a deep understanding and respect for the fluidity of theory, and his challenging of scientific systems that are ontologically rigid<sup>3</sup> and thus not easily amenable to change, perspective and discovery (Whitehead, 1929; 1933; 1938). In Whitehead's terms, we deal with complex reasoning by moving concepts and relations between the sharp clarity of immediate focus and the 'penumbral background' within which ideas are more vaguely represented and understood (Sowa, 2002). It is our ability to make such transitions that enables us to cope with the overwhelming complexities of the world we encounter, and all that we know about it—we do not try to deal with it all at once!

As Whitehead explains it, understanding (scientific or otherwise) resides in a *nexus* of interactions (Figure 3) that connects the threads between what we believe, what we observe, how we reason and experiment, and potentially many other sociological and personal factors. We can see similar ideas in many of the more contemporary critiques of science (Latour, 1987; Pickles, 1995; Gould, 2000). So, while ontology describes the concepts and relations within a domain (or more philosophically speaking, what can be true, or what exists) epistemology describes how we find out what is true, (more broadly, the nature of our knowledge and the way it is constructed). Thus, along with ontology, epistemology carries important meaning within a knowledge community; indeed, they are perhaps best considered as two sides of the same coin and it could

---

<sup>3</sup> For example: “*Systems, scientific and philosophic, come and go. Each method of limited understanding is at length exhausted. In its prime each system is a triumphant success: in its decay it is an obstructive nuisance.*” (Whitehead, 1933)

be argued that the representation of one without the other is the metaphysical equivalent of a one-sided coin! Aspects of epistemology can be captured in workflows, in provenance meta-data and even in the interactions between people. These aspects can, and are here, represented in graph form. In the future, it may be possible to directly embed tools specifically designed to describe social networks or scientific workflows in the nexus as described, but for now, we abstract the details into a form that is commensurate with domain knowledge we deal with.

Note that some of the nodes in the nexus can be represented as information, such as places, times, scales, and values (data). Other elements are more abstract in nature, such as concepts and people, and need to be represented in descriptive knowledge. Of course in Whitehead’s original vision, the nexus was a mentally-held device, with meaning deriving from the *web* created by the values and connections that we understand among data and knowledge artifacts. However, supported by progress in representing both information and knowledge (Sowa, 1999), it is now possible to characterize many aspects of the nexus within a digital environment, and to reason more formally about the nodes and their connections.

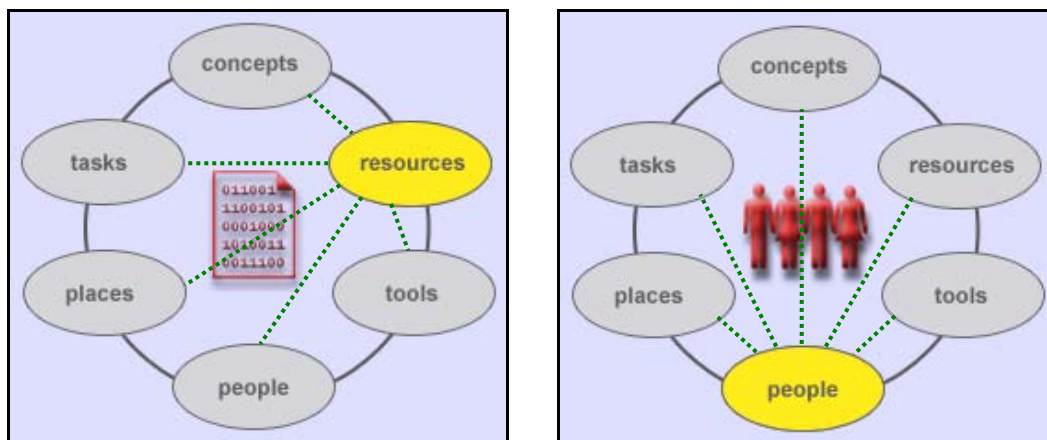


Figure 3. A simplified nexus of relationships that taken together help to define geoscientific meaning. The two alternative views show that any node can be the focus of attention, and any other aspects can be used to help signify its meaning. The left figure shows a resource as the focus of attention, the right figure shows a person.

The nexus in Figure 3 does not show all possible kinds of nodes, (for example we could also include theories, models, institutions, motivations and many others), nor all connections between nodes (though they may be connected in a variety of ways). But we assume that the more complete the nexus, the more aspects of meaning might be represented. By itself, whether ontological or epistemological, no single one of these web threads is strong enough to carry the essence of meaning. And if too many connections are removed, the structure collapses—that is, meaning is lost (Yapa, 1996). Perhaps only a subset of all connections need to be described in order to understand or use a resource appropriately, but it is difficult to predict *a-priori* which pieces might be vital for some particular situation, since it depends on what the user already knows, and what they wish to infer.

The nexus can be regarded from a semiotic perspective (Ogden and Richards, 1927; Rescher, 1978; Sowa, 2000) where the nodes change their semiotic role depending on which is the current subject of interest. Semiotics is a theory that helps explain the process of using signs or symbols to stand for, or explain, an idea—the *interpretant* (Baker, 1999). Specifically, if we are interested in the meaning of a particular dataset, we can use the neighborhood of the graph

around this dataset to interpret its meaning, with related nodes and their connections acting as the signifiers (sign vehicles). In this regard, the approach we describe is fundamentally different from current metadata approaches because it makes no conceptual distinction between the resources in a collection and the various metadata elements that describe them. In more traditional approaches, the *objects* of interest are always the resources, described by various metadata and ontological structures: whereas in our approach a resource can also be viewed as part of the metadata that helps to describe an individual or a user-community, or that helps to explain a concept. Thus, no single kind of node is always the focus of inquiry (or the subject of explanation), but rather by switching the roles of sign vehicle and object we shift perspective, and in doing so move some new notion into the foreground and others into the background.

To give a practical example of this semiotic turn, we might begin to understand some contributed dataset by knowing such things as when it was produced, and what location (where) it describes: we would find such fields in most metadata formats. Going further, a catalog or portal might add ontological richness to these metadata by connecting the resource to a series of ontological themes, processes or sub-disciplines that help to categorize the domain of interest. However, we might also wish to understand the interests of a researcher by examining the resources they have contributed, or have used; from this perspective the dataset becomes a property of the researcher, rather than the other way round. The need to reorder subjects and objects, or to truncate concepts to be simply attributes can render current representations of knowledge in triple form rather cumbersome to use. Hence the need for perspectives that can operate on an OWL-like knowledge base, to support a more flexible representation of knowledge to the user.

Importantly, these changes in perspective are not permanent in our implementation, the default state being the ontologies and metadata defined by the cyberinfrastructure conceptual design, but their temporary imposition helps us address a different set of questions about a cyber-community. The next sub-section describes how these ideas can be implemented.

## 4.2 Implementing Perspectives

We first make a distinction between a perspective defined across a whole ontology—a *global perspective*, and one defined over a (typically small) subset of it—a *local perspective*. Technically, a perspective can be defined via an RDF query; to which end *ConceptVista* supports a query structure similar to the SPARQL (<http://www.w3.org/TR/rdf-sparql-query/>) RDF query language. More specifically, an RDF query is essentially a set of ungrounded triple patterns (RDF statements that include variables), where the values of variables can be further constrained by logical filters.

### 4.2.1 Global perspectives

Consider the following SPARQL query:

```
PREFIX rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:    <http://www.w3.org/2000/01/rdf-schema#>
PREFIX theme:  <http://www.geovista.psu.edu/cV/themes.owl#>
SELECT      ?x
WHERE (?x    rdf:type          rdfs:Class
       ?x    rdfs:subClassof  theme:Geosciences.)
```

Here “rdf” and “rdfs” are namespace prefixes for W3C’s RDF and RDFS languages, and “theme” is the namespace prefix for an ontology that defines subjects of interest within the geosciences. Executing this query on the ontology will return all the resources that are sub-

classes of the resource “theme:Geosciences”. Precise definition of a global perspective based on the query is as follows. Let  $T$  denote the original ontology (which is a set of RDF statements in the form of [subject, predicate, object]),  $Q$  denotes an RDF query,  $S(Q)$  denotes the set of subject constants in  $Q$  (empty in the above example, since all subjects are variables),  $P(Q)$  denotes the set of predicate constants in  $Q$  (including “rdf:type” and “rdfs:subClassOf” in the example), and  $O(Q)$  denotes the set of object constants in  $Q$  (including “rdf:Class” and “theme:Geosciences” in the example). If we use  $R$  to denote the set of resources returned by the query, then a global perspective is a set of RDF statement, denoted by  $PS$ , which satisfy the following conditions.

For any statement [subject, predicate, object] in  $PS$ :

1. if predicate  $P(Q)$  then subject  $R \cap S(Q)$  and object  $R \cap O(Q)$ ;
2. if predicate  $P(Q)$  then subject  $R$  and object  $R$ ;
3. [subject, predicate, object]  $\in T$ .

In other words, the global perspective is a subset of the original ontology, which includes all statements suggested in the query pattern (with variables replaced by results of the query), plus all statements that connect any two resources in the query result. For example, suppose the result of the above query includes two concepts, “theme:Ecology” and “theme:Geophysics”, both of which are direct sub classes of “theme:Geosciences”; then the following statements should be included in the global perspective because they can be obtained from the original query pattern by substituting “?x” with “theme:Ecology” and “theme:Geophysics” respectively.

```

theme:Ecology      rdf:type          rdfs:Class
theme:Ecology      rdfs:subClassof  theme:Geosciences
theme:Geophysics   rdf:type          rdfs:Class
theme:Geophysics   rdfs:subClassof  theme:Geosciences

```

Now if the original ontology contains the following statement that connects “Ecology” and “Geophysics”:

```

theme:Ecology      theme:connects_to4 theme:Geophysics

```

then it should also be included in the global perspective because both “theme:Ecology” and “theme:Geophysics” are in the result of the query. In this sense a global perspective is more than simply filtering the ontology by class because we make the assumption that if two concepts are both relevant then their connections are too. By contrast, the following statement:

```

theme:Ecology      theme:connects_to theme:Geography

```

should not be in the perspective given that “theme:Geography” is not contained in the results of the RDF query.

A global perspective is analogous to a ‘view’ in a relational database since it constrains the content of the display to a theme of current interest, and the resulting structure is virtual; that is, a temporary reinterpretation to suit a particular purpose. Figure 4 shows a concept map of multiple ontologies, without the use of any perspective filters. As shown by the figure, when the number of concepts and relations increases, the readability of the display can decrease severely and a user with particular interests (for example LiDAR datasets) may find it challenging to tell where

---

<sup>4</sup> Here we define the term loosely to indicate a user-designated connection between two thematic areas of science, though other, more formal descriptions may be possible.



the relevant concepts and resources are, and how they connect together. This complexity seems typical of what we might expect from cyberinfrastructure data portals. Common solutions to this problem involve either summarizing large clusters of concepts as ‘clouds’ (Reeve *et al.*, 2005) or using a ‘depth control’ to limit the nodes displayed to only those close to a focus node (for example all nodes connected to the focus node, or connected to a neighbor of the focus node—i.e. a depth of two). While both are useful (and we implement the latter), they are not clearly connected to a user’s current interests, though they may sometimes suffice. Defining a perspective allows the type of connections, not merely their syntactic distance from the focal point, to drive the display.

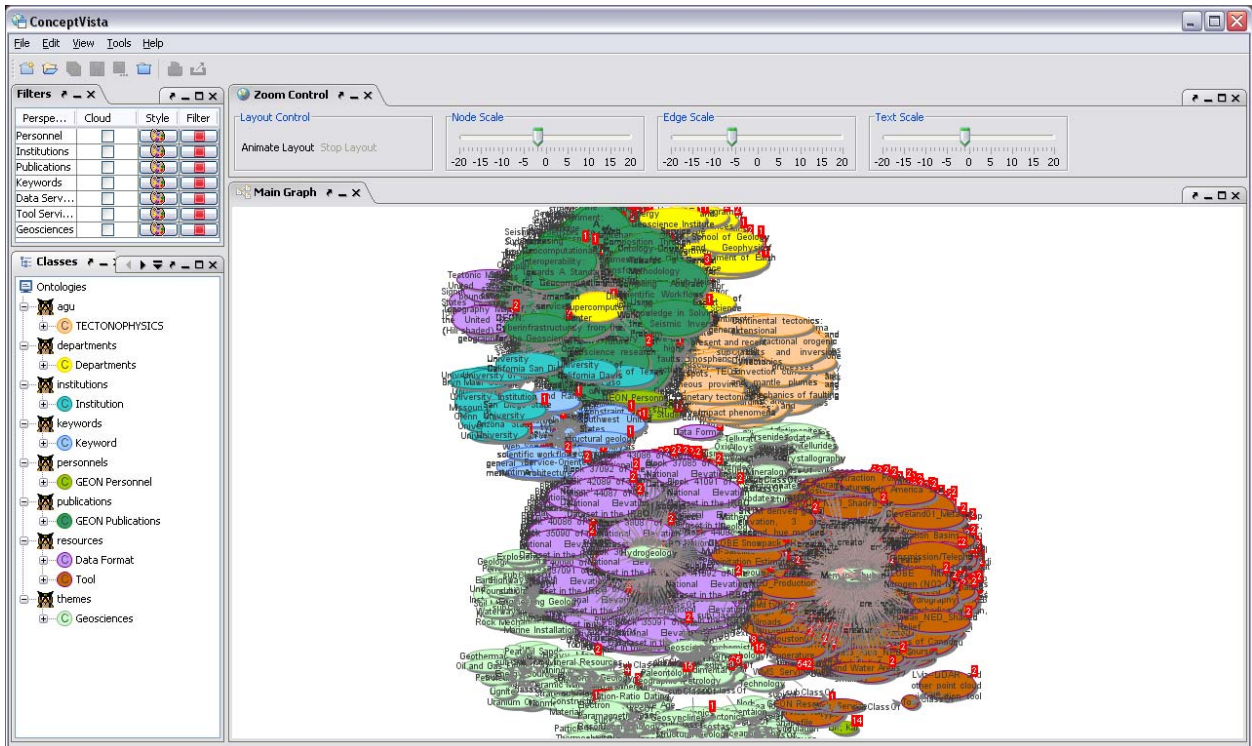


Figure 4. The conceptual universe of GEON shown without any structure imposed. Major themes and their color coding are listed in the left panel.

Figure 5 shows a concept map that uses a global perspective to filter the original display, which now contains fewer nodes, specifically only those corresponding to an interest in the supported GEON data formats mentioned in Section 3 and all instances of the ‘Shapefile’ format (datasets). By contrast, Figure 6 uses a different perspective filter to show only the GEON research team and their respective institutions (Chin *et al.*, 2002). This figure—akin to an organization chart—is an example of the semiotic turn (described above) whereby metadata describing resources (people and institutions) can be treated as concepts in their own right and thus become the subject of inquiry.

Global perspectives acting on classes are quite trivial to define; there is no reason why each OWL class in an ontology cannot be filtered in this way by default. However, there is also no reason why more complex perspectives cannot be created and serialized for later use—the upper left panel in Figure 5 shows the global perspective filters we have defined for this example application.



RDF queries. When the user applies a local perspective to a selected concept, the “?this” variable is automatically substituted by the concept, and the revised query is submitted to the query engine. The rules to determine the content of a local perspective are similar to those for global perspectives. For example, consider the following query:

```
PREFIX rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX people:  <http://www.geovista.psu.edu/people.owl#>
SELECT      ?this ?x
WHERE (?this  rdf:type      people:Personnel
        ?this  people:hasUsedDataset  ?dataset.)
```

This query aims to find all persons (instances of “people:Personnel”) and the datasets they use. Now suppose a user who is browsing the concept map selects a concept representing a particular person. Replacing the reserved variable “?this” with the selected person and executing the query will generate a set of statements that constitute the local perspective *around* the person. Compared with standard RDF queries that generate fixed results over a given ontology, local perspectives allow users to quickly define locally-customized views through interaction; a right mouse click on a node reveals the local perspectives that can be applied. As a practical example of how perspectives can be used, Figure 7 shows the results of applying a local perspective called ‘Contributions’ around GEON researcher ‘Kai Lin’. The figure shows *datasets* (dark blue), *articles* (dark green) and *tools* (brown) contributed to the cyberinfrastructure, with all other connections to ‘Kai Lin’ removed.

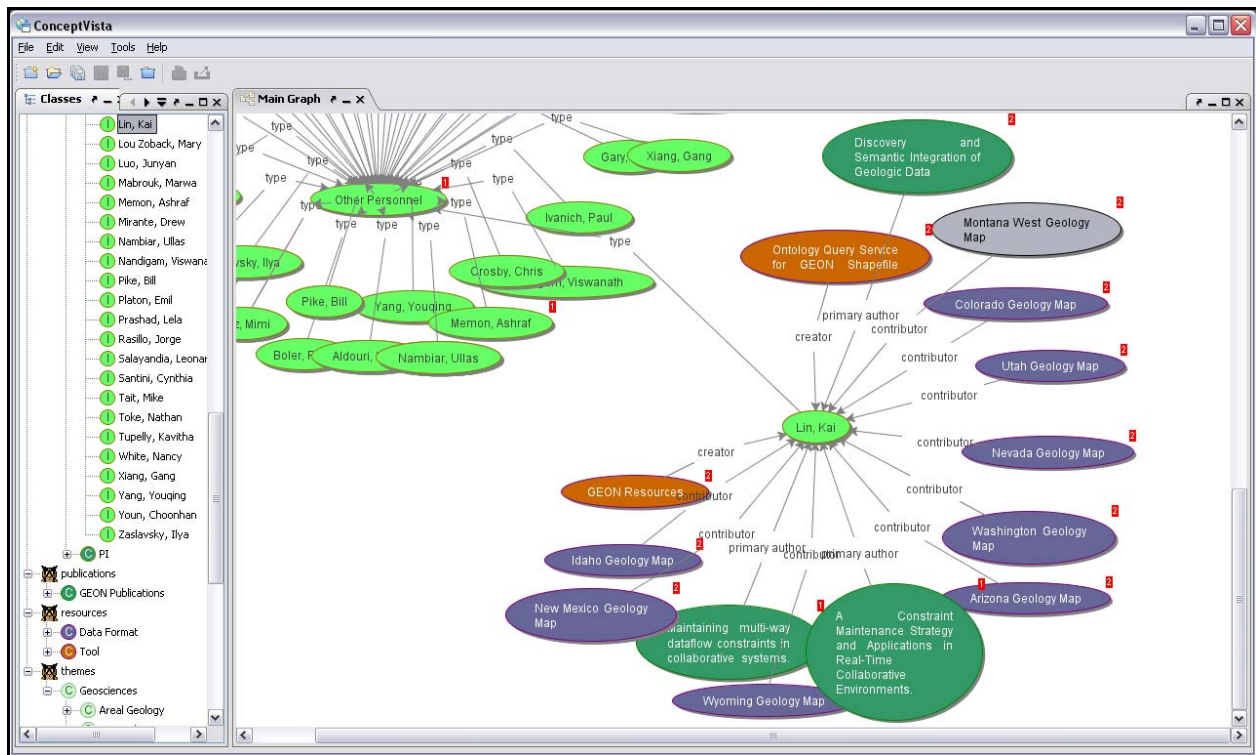


Figure 7. A local perspective constructed around GEON researcher Kai Lin, showing the resources he has contributed to GEON.

Perspectives, local and global, are defined as serialized queries using a script similar to the examples shown above. Thus they need be defined only once (typically by a knowledge engineer or a domain scientist with suitable experience). Once defined, a perspective can be



loaded into *ConceptVista* (automatically, via a project file loaded at startup) and appears as part of the list shown in the upper right of Figure 5 in the case of a global perspective, or in a list box accessed via the right mouse button for a local perspective. Selecting the perspective simply runs the appropriate script and updates the current view to display the results. In our current implementation the user can choose between filtering out all aspects of the current display outside of the perspective, or specifically highlighting those aspects that fall within it—using a shaded convex hull (as shown later in Figures 9-12).

#### 4.2.3 Folding properties in and out of concepts

When a local perspective is applied, we do not entirely remove concepts and relations that do not meet the query criteria. Instead, we *fold* the unwanted portions of the immediate graph neighborhood inside of the node currently in focus, as if these properties were now only internal to the node, and thus denoting the idea of shifting focus and abstracting away peripheral ideas into the background. The properties of current interest—those within the perspective—become external to the node, i.e. concepts in their own right. Figure 8 shows how this process works.

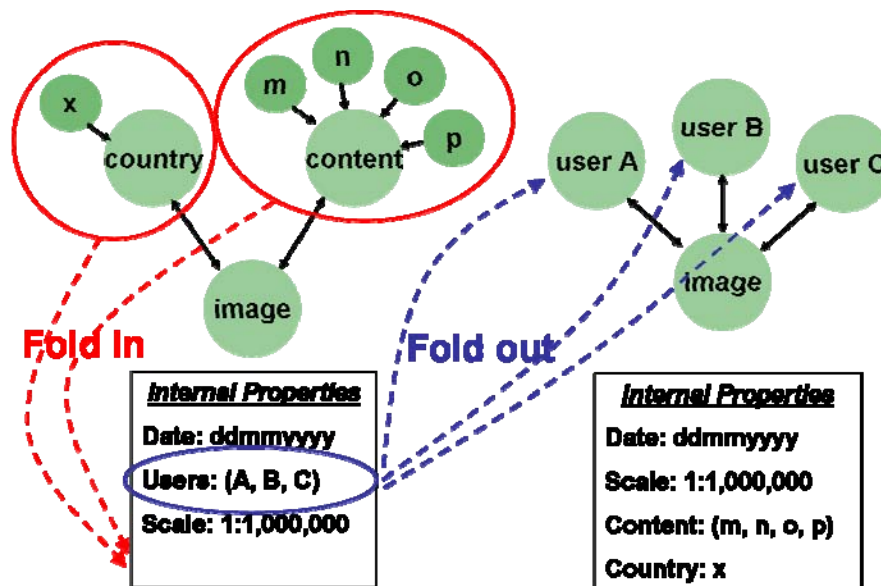


Figure 8. An example of shifting perspective on an image resource, from examining its metadata to examining its user community. Terms are folded in or out of the internal properties of the image as needed, in response to a perspective shift initiated by the user.

We maintain a table showing which properties have been internalized, and which have not, for the subject of the perspective, as Figure 9 shows. To help ease the cognitive burden of moving from one perspective to another, the folding and unfolding of concepts is animated in the visual interface (aspects to be removed are progressively contracted into the focal concept and new aspects are progressively expanded). Since we can calculate the properties contained within a perspective (as described above), we can also calculate its obverse—all properties left out of a perspective—by calculating all properties connected to a local concept and then subtracting those included in the current perspective.



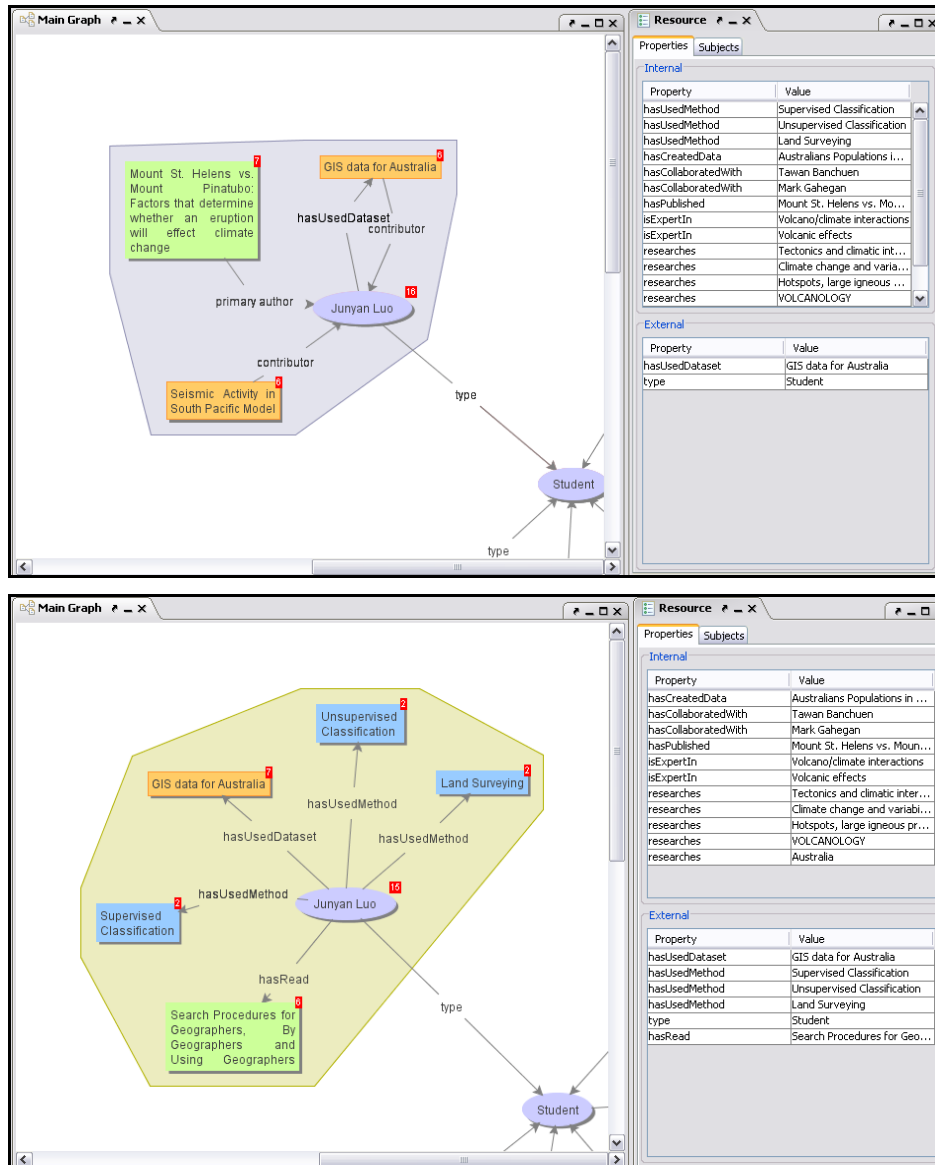


Figure 9. Two perspectives constructed around an individual ('Junyan Luo'). The upper screenshot describes the contributions of the individual, the lower one the resources they have used. The right panel in the application shows how the local filter described above folds various properties in or out depending on which perspective is being used. Properties folded in temporarily become internal to the individual and are removed from the graph<sup>5</sup>.

### 4.3 Comparing Perspectives

As above, a local perspective can also be as arbitrarily-complex as needed. And since many perspectives can be displayed concurrently, we can easily explore many kinds of complex relations between nodes in the graph, to reach a deeper understanding of the resources in a

<sup>5</sup> For illustrative purposes, Figures 9-12 depict added details (richness), beyond what is currently gathered and supported by the GEON cyberinfrastructure, for a small fragment of a research community.

cyberinfrastructure. As an example, Figure 10 allows us to ask the question: “Which resources contributed by A have been used by B?”

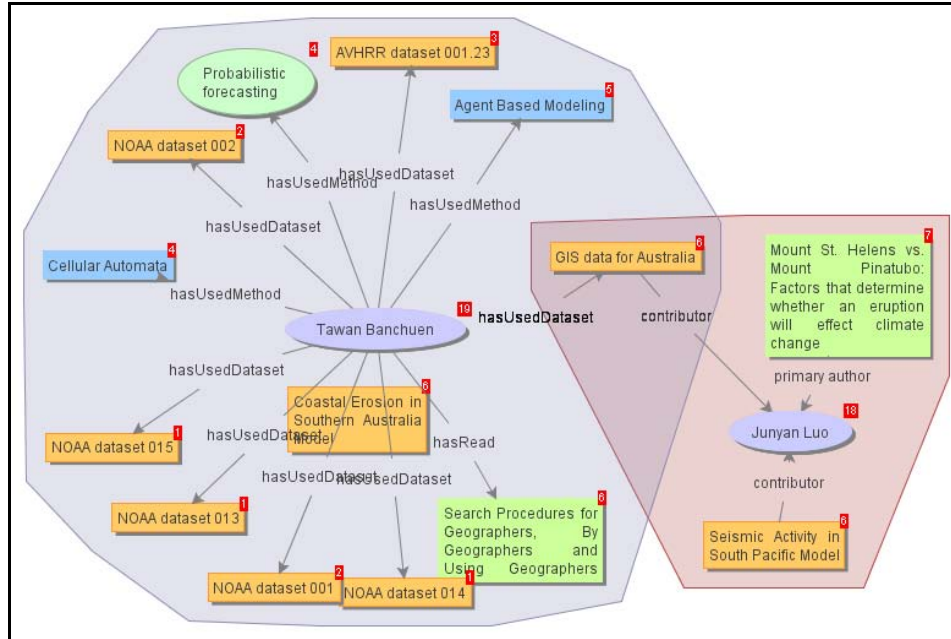


Figure 10. Two perspectives showing what one researcher has used (blue hull), and another has contributed (pink hull). Their intersection shows that one dataset ‘GIS data for Australia’ was contributed by one and used by the other.

The multiple perspectives in Figure 11 show the conjunction (and disjunction) of areas of interest between five researchers, based on ontological tags associated with each person. Tags such as these might be assigned directly or harvested from the production and use of other tagged resources through time, such as keywords from published articles, metadata tags from datasets used, and so forth. In both of these examples, we might also think of the intersecting nodes as being a kind of ‘boundary object’ that represent a point of common knowledge or experience that is shared between two researchers (Harvey and Chrisman, 1998).

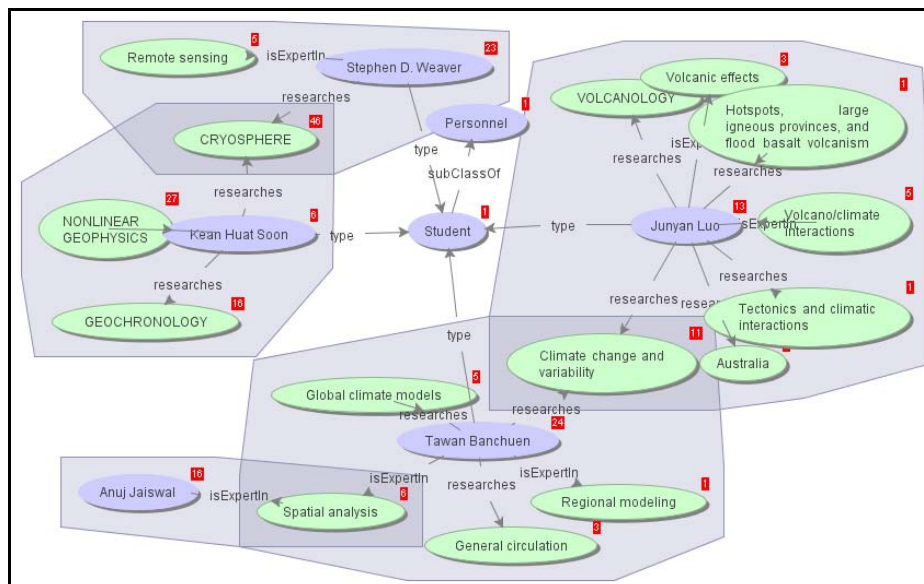


Figure 11. An overview of the research interests (green nodes) of five researchers (mauve nodes). Overlapping topics of interests between researchers are shown by intersecting convex hulls, each representing the ‘research interest’ perspective for a given researcher.

As a final set of examples, consider a dataset (or a person, a method, or some other resource) that is described by metadata and conceptual tags from a domain (or task) ontology. We have shown that perspectives allow us to hide or promote aspects of metadata to make them the subject of investigation, as opposed to being simply a property of a resource. Thus we gain access to further insights about resources by exploring their user community (and *vice versa*), or some other aspect of their descriptions. We can go further, however, and compare how resources are defined (ontologically, top-down), and how they are used (emergent, bottom-up). That is, the resources we use and create, and the way they in turn are used, by whom, and for what purpose, also help to create a kind of emergent meaning. To some extent, this bottom-up meaning represents an evolving, *pragmatic* view (Buckingham Shum, 2006) onto knowledge in flux as resources are used, gain acceptance within a community, and eventually are retired or superseded. Figure 12 shows some of this ‘drift’. The dataset in focus (‘Coastal Erosion in Southern Australia’) is registered to two research themes (‘Coastal Processes’ and ‘Modeling’) that do not appear to match the research interests of its users. Whereas we might expect such mismatches to occur, if they happen frequently we might consider modifying the research themes connected to the dataset, or to the researchers (Haase *et al.*, 2005).

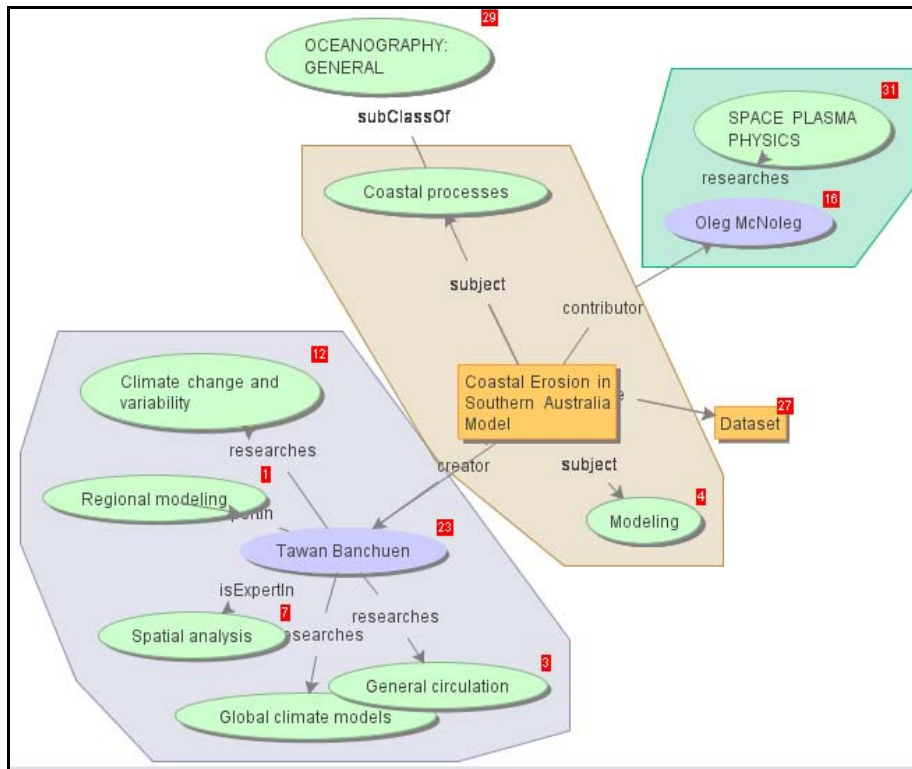


Figure 12. Mismatch between thematic description of a resource and its use by a community. See text for details.

### 4.3.1 Deriving perspectives from implicit relationships

The idea behind derived perspectives is to explicitly represent relationships between concepts that are indirectly linked through some intermediary concepts. For instance, if two keywords are associated with the same article or dataset, then there is an indirect relationship of co-occurrence between the two resources, or between their creators. A derived perspective allows such relationships to be created virtually and used as needed, without any permanent changes to the underlying knowledge structure. As with the SQL Group-By clause, derived perspectives provide a temporary field that does not exist in the schema, but which may be useful to the user.

During the capture or creation of ontologies or concept maps, it is typical for somewhat arbitrary decisions to be made that will affect which relations are captured, and which are not. For simplicity, and to avoid confusion among the users, not all of the possible relations between concepts are represented. However, these missing relationships can be useful, for example to help a user to see that two different concept maps are semantically equivalent, or to reorganize one ontology so that it maps more readily to another. There is often a need to see past minor differences in depth and detail to better emphasize the commonalities.

We implement derived perspectives by temporarily creating the needed, missing triples and storing them in a separate ontology (which one might think of as a special kind of mapping ontology between the current knowledge schema as is and how the user would like to think of it). If needed, a separate instruction can be used to roll these derived statements into the original ontology—in which case they are no longer represent a perspective, but part of the permanent knowledge record. Beyond the basic parameters required for an RDF query, a derived perspective requires three more terms: (i) a ‘Group by’ operator, whose results will be used to infer an indirect relationship between other concepts, (ii) a name/label for the new (derived) relationship and (iii) identified concepts, between which the derived relationship should be applied.

Conceptually, the syntax of a derived perspective can be defined as follows:

```
SELECT      ?y, ?z
WHERE (?x rdf:rel_type1 ?y) and (?x rdf:rel_type2 ?z) group by ?x add
relationship rdf:new_relationship
```

Examples:

1. Select ?author, ?keyword  
where (?author rdf:authorOf ?article) and (?article  
rdf:hasKeyword ?keyword)  
Group by ?article add relationship "interested\_in"
2. Select ?keyword1, ?keyword2  
Where (?article rdf:hasKeyword ?keyword1) and  
(?article rdf:hasKeyword ?keyword2)  
Group by ?article add relationship "co-occur"

The original representation of GEON publications, shown in Figure 13, captures a minimal set of relationships between articles, authors, and keywords. However, from this original representation it is difficult to observe clear connections between authors and their implied interests (the articles act as intermediaries). By applying the derived perspective in example 1



above, we are able to temporarily remove articles from the representation and instead promote the indirect relationships between authors and keywords, shown in Figure 14.

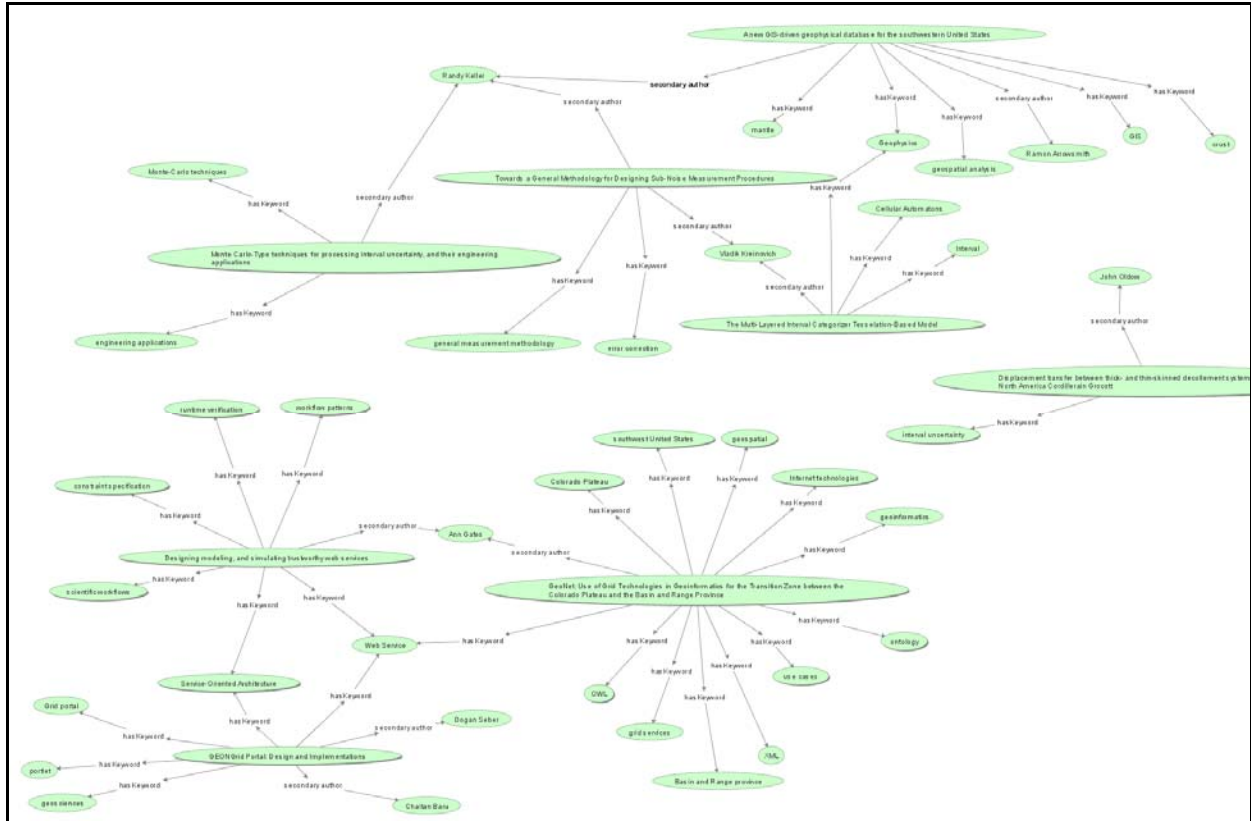


Figure 13. Overview of GEON articles, authors and article keywords, as currently represented with no perspectives defined.

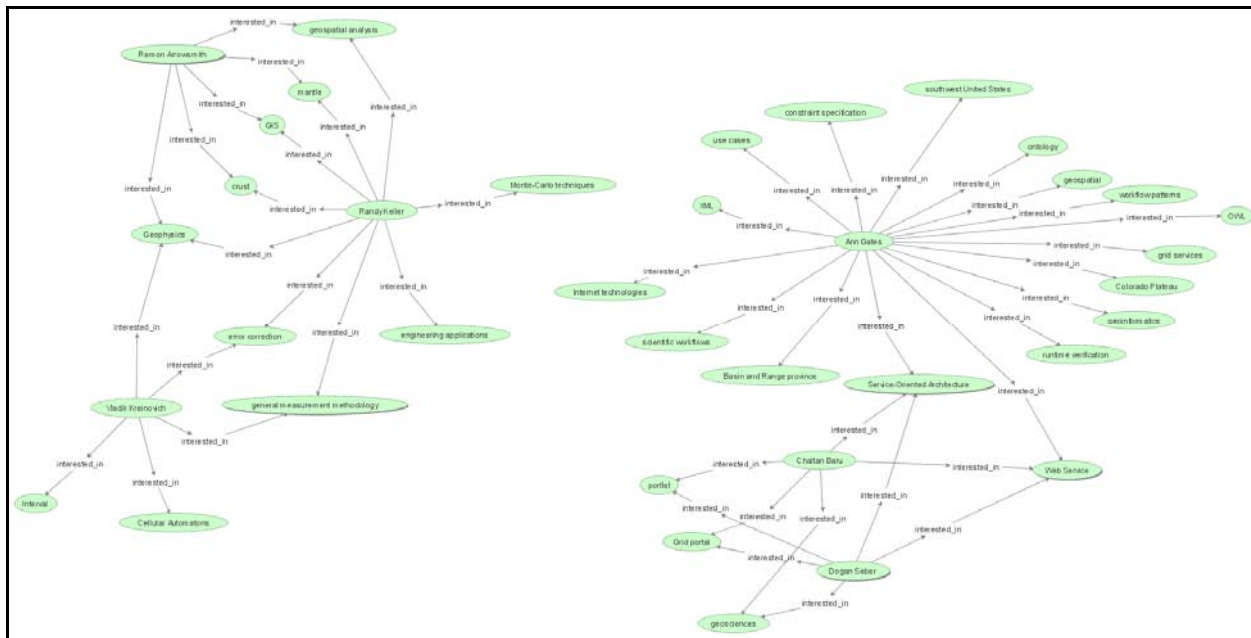


Figure 14. A derived perspective on the author/keyword/article schema, removing articles and temporarily connecting authors directly with the keywords used to identify their articles.

Based on these new relationships, a division between two groups of authors is clearly visible. Further, one can also observe that “Randy Keller” and “Ramon Arrowsmith” appear to have many common research interests. Similarly, by creating a derived perspective on keyword co-occurrence, defined in example 2 above and shown in Figure 15, one can clearly observe different clusters of topics, and that “web service” and “service oriented architecture” form important common links between researchers (because they show many connections).

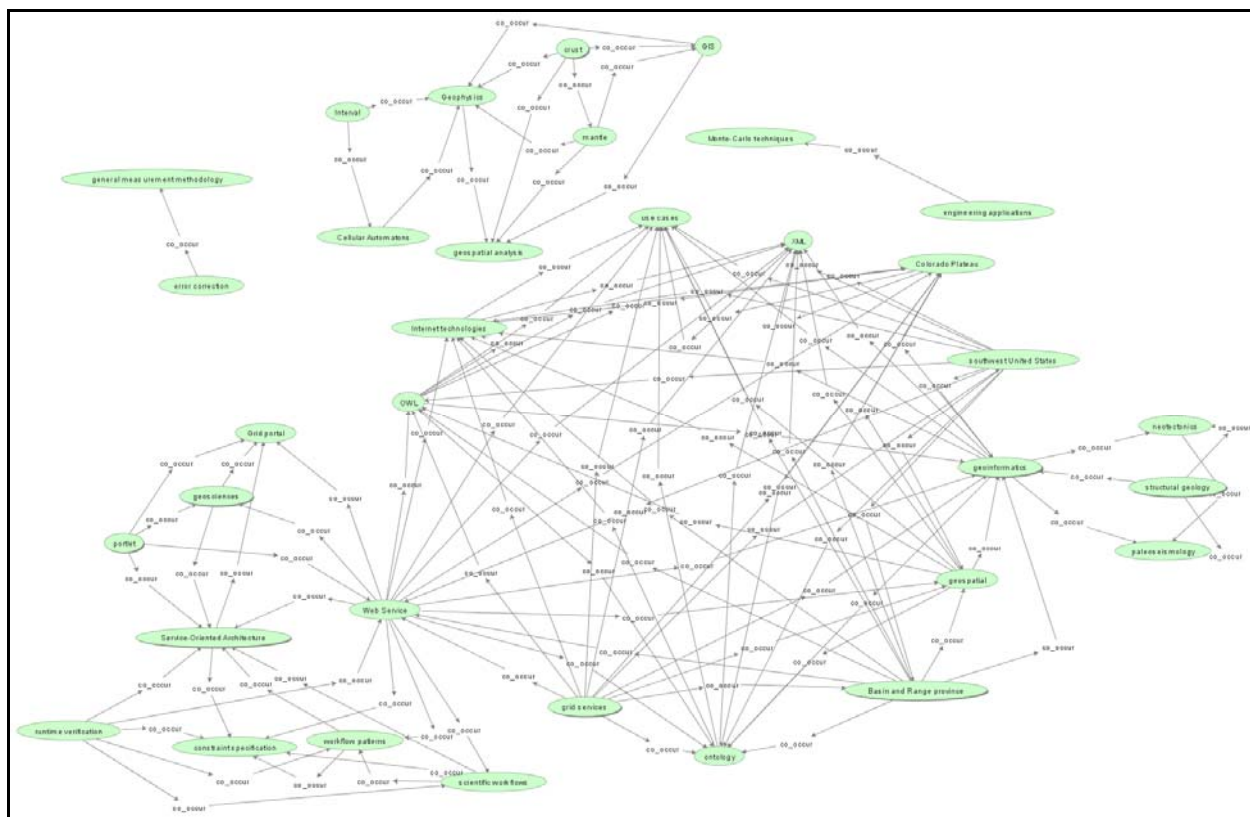


Figure 15. A derived perspective to create a keyword co-occurrence graph for keywords used in GEON publications.

## 5 Discussion and Conclusions

*“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)

Any knowledge domain that relies only on static descriptions runs the risk of stagnation (Sowa, 2002). Use-cases, provenance data, social networks and workflows, when added to ontology, provide a rich structure that allows us to see what our resources and researchers actually do. This helps us to understand some of the more pragmatic aspects of their meaning and definition. We give some examples above of how richer contextualization of resources, through the use of perspectives, might facilitate the study and comparison of different aspects of meaning.

Perspectives as defined here have several specific uses within our knowledge visualization strategies:

1. Providing the means to create useful relationships that are missing from some ontology—without changing it.
2. Comparing one ontology to another. If two ontologies are represented at different levels of abstraction, a derived perspective can hide extra details in one ontology to better match the other.
3. Allowing concepts and relationships that are not currently relevant to be treated as attributes (i.e. terminal nodes in the graph) thus moving their details into the background and facilitating comparative clarity in the foreground.
4. Reducing the noise (unimportant relationships) and focusing on important ones.

Note that this kind of functionality is not beyond the reach of current description logics. What is new here is the convenient packaging for easy deployment and visualization by the user. Currently, the derived perspective can only make explicit indirect relationships between concepts *that are linked directly through some single other concept*. For instance if A links to B and B links to C, the perspective can be used to create a temporary relationship between A and C. However, if A links to B, B to C, and C to D, the current implementation cannot create a relationship between A and D, as they lack a single common identifier.

Cyberinfrastructures are necessarily comprehensive in terms of data collections, tools, community of users and ontology. But there is a tension between these massive collections of resources, and the specific needs of individuals or groups that have interests in just a small subset. It is necessary to provide smart, adaptable browsing strategies that reflect current interests and hide irrelevant details in cognitively-sound ways. Furthermore, as humans we search for—and understand—scientific content by weighing together multiple factors that include, but are not limited to, ontology. Situating a representation of the user community within the cyberinfrastructure allows us to search across many other dimensions.

We have shown how a knowledge nexus can support multiple strands of meaning, and how perspective filters, defined against an OWL model, facilitate examination of a subset of connections within a complex concept space in a manner that suits thematic exploration. Moreover, by visualizing multiple perspectives we are able to quickly highlight connections or overlaps among the cyberinfrastructure’s resources, its researchers, their publications and ontologies. We believe these enhancements enable users to examine multiple facets of meaning and navigate between them with ease, and they help to address some of the shortcomings of how we think about and gather metadata (Fisher *et al.*, 2004).

Many questions remain, relating to the effectiveness of different kinds of concepts and relationships in conveying meaning. Or put another way, which aspects of the knowledge nexus add the most value in conveying meaning—taking into account the utility they add and the cost of computing or capturing them? A related open question concerns the degree to which emergent knowledge helps users to understand resources, and their knowledge community. There is strong evidence that it does help, for example the current interest in both citation graphs and social networks in science communities (e.g. Chen, 1999; Li *et al.*, 2002; Dumais *et al.*, 2003). Our work builds on this evidence and shows how the resulting knowledge can be integrated with ontological approaches to knowledge representation. We plan to study this

further via experimentation with CI user communities, to establish the relative utility of various kinds of knowledge and the effectiveness of different approaches to visualizing it.

*ConceptVista* is an open-source project developed by researchers at the GeoVISTA Center, Penn State University to support a number of collaborative projects involving the exchange of understanding between researchers in geography and the geosciences. The Java source code is available from *SourceForge* ([www.sourceforge.net](http://www.sourceforge.net)) and the application may be freely downloaded from <http://www.geovista.psu.edu/ConceptVISTA/index.jsp>.

## 6 Acknowledgements

This research was funded by the US National Science Foundation (NSF) via grants BCS–9978052 (HERO), ITR (BCS)–0219025, and ITR (EAR)–0225673 (GEON). The authors would like to thank the anonymous reviewers and editors for their invaluable help in improving this manuscript, researchers at the San Diego Supercomputing Center for their help in accessing the GEON catalogs, and the GEON research community itself for providing such a rich and interesting case study.

## 7 References

- Albrechtsen, H., Andersen, H., Bødker, S. and Pejtersen, A. (2001). *Affordances in activity theory and cognitive systems engineering*. Risø National Laboratory: Denmark, 37 p., ISBN 87-550-2928-0.
- Ansari, A., S. Essegai and Kohli, R., 2000. Internet recommendation systems. *Journal of Marketing Research* 7, 363-375.
- Baker, V., 1999. Geosemiosis. *Geological Society of America Bulletin*, 111 (5), 633-645.
- Berman, F. and Brady, H., 2005. Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences. URL: [www.sdsc.edu/sbe/](http://www.sdsc.edu/sbe/).
- Brennan, S., Mueller, K. Zelinsky, G., Ramakrishnan, I., Warren, D. And Kaufman, A., 2006. Toward a multi-analyst, collaborative framework for visual analytics. *IEEE Symposium on Visual Analytics Science and Technology 2006*, Baltimore, MD, IEEE.
- Brodaric B. and Gahegan, M., 2006. Representing geoscientific knowledge in cyberinfrastructure: challenges, approaches and implementations. In: Sinha, A. K. (Ed.) *GeoInformatics, Data to Knowledge*, Geological Society of America Special Paper 397, pp. 1-20.
- Brodaric, B. and Gahegan, M., 2007. Experiments to examine situated geoscientific concepts. *Spatial Cognition and Computation Journal* (Special Issue on Cognitive Semantics and Ontologies), *forthcoming*.
- Buckingham Shum, S., 2006. Sensemaking on the pragmatic web: A hypermedia discourse perspective. 1st International Conference on the Pragmatic Web, 21-22 Sept 2006, Stuttgart, Germany. URL: <http://oro.open.ac.uk/6442/01/KMI-TR-06-16.pdf>.
- 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 (3), 237-248.



- Cardoso, J. and Sheth, A., 2003. Semantic e-workflow composition. *Journal of Intelligent Information Systems* 21 (3) (November 2003), 191–225.
- Carroll, J., Bizer, C. and Hayes, P., 2005. Named graphs, provenance and trust. Proceedings of the 14th International Conference on World Wide Web, Chiba, Japan, pp. 613-622.
- Chen, C., 1999. Visualizing semantic spaces and author co-citation networks in digital libraries. *Information Processing and Management*, 35, 401-420.
- Chen, L., Yang, X. and Tao, F., 2006. A semantic web service based approach for augmented provenance. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06). URL: <http://ieeexplore.ieee.org/iel5/4061321/4061322/04061437.pdf?tp=&isnumber=&arnumber=4061437>.
- Chin, G., J. Myers and Hoyt, D., 2002. Social networks in the virtual science laboratory. *Communications of the ACM*, 45 (8), 87-92.
- Chin, G. and Lansing, C., 2004. Capturing and supporting contexts for scientific data sharing via the Biological Sciences Collaboratory. ACM Conference on Computer Supported Cooperative Work, 2004, Chicago, IL, ACM. URL: <http://portal.acm.org/citation.cfm?id=1031607.1031677>
- Clancey, W., 1994. Situated cognition: How representations are created and given meaning. *Lessons from Learning*. R Lewis and P Mendelsohn (Eds.) Amsterdam, North-Holland: pp. 231-242.
- Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R. and Robbins, D, 2003. Stuff I've Seen: A system for personal information retrieval and reuse. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, Toronto. URL: <http://portal.acm.org/citation.cfm?id=860451&dl=ACM&coll=portal>.
- Finholt, T. A., 2001. Collaboratories. *Annual Review of Information Science and Technology*. B. Cronin (ed.) 36, 73-108.
- Fisher, P., Comber, A., and Wadsworth, R., 2004. Land cover mapping and the need for expanded metadata. Proceedings of GIScience 2004, Third International Conference on Geographic Information Science, University of Maryland Conference Center, USA, October pp. 20-23.
- Fox, P., McGuinness, D., Middleton, D. Cinquini, L., Darnell, J. Garcia, J., West, P. Benedict, J. and Solomon, S., 2006. Semantically-enabled large-scale science data repositories. ISCW-2006, 5<sup>th</sup> International Conference on the Semantic Web. URL: [http://iswc2006.semanticweb.org/items/in\\_use\\_9.php](http://iswc2006.semanticweb.org/items/in_use_9.php).
- Frodeman, R., 2003. *Geo-Logic: Breaking ground between philosophy and the earth sciences*. Albany, SUNY Press.
- Gahegan, M., Agrawal, R., Banchuen, T. and DiBiase, D., 2007. Building rich, semantic descriptions of learning activities to facilitate reuse in digital libraries. *International Journal of Digital Libraries*. URL:

<http://www.springerlink.com/content/q102m641460h77v6/?p=312504c3b4d0489385ba014bf9ce795f&pi=3>.

- Gahegan, M. and Pike, W. (2006). A situated representation of geographical information. *Transactions in GIS* 10 (5), 727-749.
- Gould, S. J., 2000. De-constructing the science wars by reconstructing an old mold. *Science*, 287, 253-261.
- Guarino, N., 1998. Formal ontology in information systems. In: Guarino, N. (ed.) *Formal Ontology in Information Systems*. Proc. FOIS'98, Trento, Italy, June 6-8 1998. IOS Press, Amsterdam, pp. 3-15.
- Haase, P., Hotho, A., Schmidt-Thieme, L. and Sure, Y., 2005. Collaborative and usage-driven evolution of personal ontologies. *Semantic Web: Research and Applications, Proceedings*. 3532: 486-499.
- Harvey, F. and Chrisman, N., 1998. Boundary objects and the social construction of GIS technology, *Environment & Planning A* 30 (9), 1683-1694.
- Hey, T. and Trefethen, A., 2003. e-Science and its implications. *Philosophical Transactions of the Royal Society of London Series A - Mathematical, Physical and Engineering Sciences* 361, 1809-1825.
- Kim, S. and Fox, E. A., 2004. Interest-based user grouping model for collaborative filtering in digital libraries. 7th International Conference of Asian Digital Libraries, 13-17 Dec 2004, Shanghai, China. URL: [http://pubs.dlib.vt.edu:9090/16/01/ICADL04\\_shk.pdf](http://pubs.dlib.vt.edu:9090/16/01/ICADL04_shk.pdf).
- Latour, B., 1987. *Science in Action*, Cambridge, MA: Harvard University Press.
- Lemke, J., 1997. Cognition, context, and learning: A social semiotic perspective. *Situated Cognition: Social, Semiotic, and Psychological Perspectives*. Kirshner, D. and Whitson, J. (Eds.) Mahwah, NJ, Erlbaum, pp. 37-55.
- Li, G., Uren, V., Motta, E., Buckingham-Shum, S. and Domingue, J., 2002. ClaiMaker: Weaving a semantic web of research papers. *The Semantic Web - ISWC 2002: First International Semantic Web Conference*. Horrocks, I. and Hendler, J. (Eds.), *Lecture Notes in Computer Science*, Vol. 2342, Springer: Berlin, pp. 436-441.
- Lin, K. and Ludäscher, B., 2003. A system for semantic integration of geological maps via ontologies. *Proc. of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data (SCISW)*, 2003.
- MacEachren, A., Pike, W., Yu, C., Brewer, I., Gahegan, M., Weaver, S and Yarnal, B., 2006. Building a geocollaboratory: Supporting Human-Environment Regional Observatory (HERO) collaborative science activities. *Computers, Environment, and Urban Systems* 30, 201-225.
- McGuinness, D. and van Harmelen, F. (Eds.), 2003. *OWL Web Ontology Language Overview*. <http://www.w3.org/TR/owl-features/#s1>.
- McGuinness, D. and da Silva, P., 2003. Infrastructure for Web Explanations. *International Semantic Web Conference 2003*, Published in: *Lecture Notes in Computer Science*, Vol. 2870, Springer: Berlin, pp. 113-129.

- Nambiar, U., Ludaescher, B. Lin, K. and Baru, C., 2006. The GEON portal: accelerating knowledge discovery in the geosciences. Workshop On Web Information And Data Management: Proc. Eighth ACM International Workshop on Web Information and Data Management, Arlington, Virginia, USA.
- National Academies of Sciences (2004). *Mapping Knowledge Domains. Proceedings of the National Academies of Sciences (PNAS)* 101, Supp. 1.
- National Science Foundation (NSF), 2003. Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure. URL: [http://www.communitytechnology.org/nsf\\_ci\\_report/ExecSum.pdf](http://www.communitytechnology.org/nsf_ci_report/ExecSum.pdf).
- Novak, J. D., 1998. Learning, Creating, and Using Knowledge: Concept Maps™ as Facilitative Tools in Schools and Corporations, Lawrence Erlbaum Associates, (Mahwah).
- Novak, J., Wurst, M., Fleischmann, M. And Strauss, W. 2003. Discovering, visualizing and sharing knowledge through personalized knowledge maps. Proceedings of the AAAI Spring Symposium on Agent-Mediated Knowledge Management, Lecture Notes in Artificial Intelligence 2926, Springer: Berlin, pp. 213-228.
- Novak, J. D. and Wurst, M. 2004. Supporting knowledge creation and sharing in communities based on mapping implicit knowledge. *Journal of Universal Computer Science* 10 (3), 235-251.
- Noy, N. F. and Hafner C. D., 2000. Ontological foundations for experimental science knowledge bases. *Applied Artificial Intelligence* 14 (6), 565-618.
- Ogden, C. K., and Richards, I. A., 1927/1963. The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism, New York: Harcourt, Brace & World, 363p.
- Peirce, C. S., 1931. The Collected Papers of Charles Sanders Peirce. Harvard University Press: Cambridge, MA.
- Pickles, J., 1995. Representations in an electronic age: Geography, GIS, and democracy. Ground Truth: The Social Implications of Geographic Information Systems. J. Pickles (Ed.) New York, Guilford Press, pp. 1-30.
- Pike, W. A., Ahlqvist, O., Gahegan, M. and Oswal, S., 2003. Capturing context in collaborative science: Supporting collaborative science through a knowledge and data management portal, workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data, at Second International Semantic Web Conference, Sanibel Island, FL, October 2003.
- Pike, W. and Gahegan, M., 2007. Beyond ontologies: towards situated representations of scientific knowledge. *International Journal of Human-Computer Studies*, 65 (7), 674-688.
- Raskin, R., and Pan., M., 2007. Semantic web for earth and environmental terminology (SWEET), *Computers and Geosciences*, *forthcoming*.
- Reeve, L., Han, H., and Chen, C., 2005. Information visualization and the semantic web. In V. Geroimenko and C. Chen (Eds.), *Visualizing the Semantic Web* (2nd edition). London: Springer: Berlin.

- Rescher, N., 1978. *Peirce's Philosophy of Science*. London: Notre Dame.
- Reynolds, R., Sinatra, G., and Jetton, T., 1996. Views of knowledge acquisition and representation: a continuum from experience centered to mind centered. *Educational Psychologist* 31 (2), 93-104.
- Rodriguez, M. A. and Egenhofer, M. J., 2003. Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering* 15 (2), 442-456.
- Sacerdoti, F., Chandra, S. and Bhatia, K. (undated). Grid systems deployment & management using Rocks. URL: <http://www.geongrid.org/pubs/local/SacerdotiChandra.pdf> (accessed Jan 22nd, 2008).
- Sheth, A. P., 1999. Changing focus on interoperability in information systems: from system, syntax, structure to semantics. In M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman (Eds.), *Interoperating Geographic Information Systems*. Kluwer: Dordrecht, pp. 5-29.
- Solomon, K. O., Medin, D. L., and Lynch, E. 1999. Concepts do more than categorize. *Trends in Cognitive Sciences* 3 (3): 99-104.
- Sowa, J., 1999. *Knowledge Representation: Logical, Philosophical, and computational Foundations*, Brooks Cole: Belmont, CA.
- Sowa J., 2000. Ontology, metadata, and semiotics. Conceptual structures: Logical, linguistic, and computational issues. Ganter, B. and Mineau, G. (Eds.). LNCS 1867, Springer: Berlin, pp. 55-81.
- Sowa, J., 2002. Signs, processes and language games: foundations for ontology: URL <http://www.jfsowa.com/pubs/signproc.htm>.
- Stevens, R., Robinson, A. and Goble, C., 2003. myGrid: Personalised bioinformatics on the information grid. *Bioinformatics* 19 (Suppl. 1), i302-i304.
- Sugumaran, V. and Storey, V., 2002. Ontologies for conceptual modeling: their creation, use, and management. *Data & Knowledge Engineering* 42 (3), 251-271.
- Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., and Hübner, S., 2001. Ontology-based integration of information: a survey of existing approaches. In: Stuckenschmidt, H. (ed.), *IJCAI-01 Workshop: Ontologies and Information Sharing*, pp. 108-117.
- Whitehead, A. N., 1929. *Process and Reality: An Essay in Cosmology*. New York, Social Science Book Store. 546p.
- Whitehead, A. N., 1933. *Adventures of Ideas*, Macmillan: New York.
- Whitehead, A. N., 1938. *Modes of Thought*, Macmillan: New York.
- Whitehead, A. N. and Russell, B., 1997. *Principia Mathematica to 56*, Cambridge: UK, 456p.
- Yapa, L., 1996. What causes poverty? A postmodern view. *Annals of the Association of American Geographers* 86, 707-728.
- 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.