The case for transforming the Geological Survey knowledge system: where digital geoscience spatial models meet the semantic grid[1]

T V Loudon

British Geological Survey, West Mains Road, Edinburgh EH9 3LA, UK.
Email: v.loudon@bgs.ac.uk

Preprint of

Loudon, T.V., 2006. The case for transforming the geological survey knowledge system: Where digital spatial models meet the semantic Grid, *in* Sinha, A.K. ed., Geoinformatics: Data to Knowledge: Geological Society of America Special Paper 397, p. 201-209, doi: 10.1130/2006.2397(15).

Abstract

As the Internet develops into a more powerful cyberinfrastructure (the Grid), Geological Survey knowledge will enter a framework where users assess its value by its relevance and evaluated reliability rather than its source, and artificial boundaries of place and discipline lose significance. Within Geological Surveys, much effort has been devoted to the development of the geological map through digital cartography. Consequently, a firm base already exists for more radical change to the geoscience knowledge system. The geoscience map is prepared from field records and structured data, and given meaning through the illustrated text narratives of map explanations. A more comprehensive view of this knowledge system is needed to gain the full benefits of the Grid. The flexibility of markup languages can link narrative to database, spatial modeling and semantic representations, and thereby combine the insights of human understanding with the power of computation. Rationalizing and extending the scope of existing ontologies (specifications of conceptualizations) to align their structure and vocabularies could provide a more comprehensive conceptual view. Thus, a major challenge for Geological Surveys will be to transform their knowledge system to conform to a new service-oriented infrastructure devised by others and its future evolution determined by users.

Keywords: Geoscience survey; digital cartography; spatial model; markup language; semantic grid; cyberinfrastructure.

[1] This paper is published by permission of the Director of the British Geological Survey (NERC). It reflects a personal, not a corporate, viewpoint.

1 THE OPPORTUNITY

Recent advances in geoinformatics suggest that it is timely to review the system for handling information and knowledge. Many of those involved in geoscience surveying may question the need for further change. This paper attempts an answer, directed at geoscientists, by indicating some possible directions and expected benefits.

The acceptance of the concept of the Digital Geoscience Spatial Model (DGSM) as an approach to geological survey (Smith, 2001), and the deployment of the Semantic Grid as infrastructure for e-science, combine to offer an unusual opportunity, namely, to transform part of the geoscience knowledge system. The work and proposals recorded at the GEON website on cyberinfrastructure for the geosciences (GEON, 2004a, b) indicate the radical changes envisaged for the broader field of geoscience. The ultimate aim for a Geological Survey knowledge system may be to meet user requirements and expectations, but the results will be more effective if they emerge from careful consideration of the subject matter. A successful outcome therefore depends on geological input at scientific, technical and management levels. Some potential benefits to geoscience survey are reviewed in section 2, some underlying concepts in section 3, and prospects for extending the vision of future geoscience survey in section 4, with conclusions in section 5.

The Grid

The revolutionary impact of information technology on society (like the railway, road, electricity grid, or telephone system before it) may occur when it ceases to comprise differentiated, proprietary technologies and becomes a ubiquitous infrastructure shared by all: a commodity, no longer offering a strategic advantage over competitors, but a cost to be managed like office supplies. This is seen as a likely consequence of the current evolution of the Internet into a new infrastructure of information and communications technology referred to as the cyberinfrastructure or Grid (Foster and Kesselman, 2003).

The Report of the [US] National Science Foundation Advisory Panel on Cyberinfrastructure states in its summary (Atkins et al., 2003, page 81): ‘The scope and scale of the ACP [Advanced Cyberinfrastructure Project] will require an annual budget of about $1 billion over and above the current PACI and infrastructure programs in NSF… Nevertheless, our recommendations should be considered as only a beginning.’ Geoscientists are involved, for example through the GEON project – a prototype national Geosciences Cyberinfrastructure Network. An aim of GEON is helping to weave the separate strands of the solid Earth sciences disciplines and data into a unified fabric. Initiatives elsewhere include European Commission funding, which supports the EGEE project (Enabling Grids for E-science in Europe), a two-year project of a four-year program, with links to government initiatives and funding (EGEE, 2004; e-Science Institute, 2004; UK Research Councils, 2004).

The Grid is bringing a large increase in the capacity and power of computer networks, thus allowing more ideas to be expressed, shared, and structured to extend human knowledge. It aims to provide facilities that, among other things, enable scientists to generate, analyze, share, discuss and communicate their investigations in a more effective manner. Hinke (2004) describes how Grid services ‘could revolutionize the ability of developers to easily construct extensible data handling and processing systems, by providing seamless access to various types of resources including computational resources, data archives, scientific sensors and the services to process data into some desired product, which itself could be formulated as a service.' The Grid might, for example, change the emphasis in a Geological Survey's objectives, from delivering published end products, to maintaining information resources from which end users could select services that would respond more flexibly to their specific needs.

The DGSM

Geoscience surveying operates at the interface of the solid Earth and geological knowledge, linking them through spatial location. It thus has important roles in collecting information and testing interpretations. The Digital Geoscience Spatial Model extends beyond conventional map-based survey. The model overcomes constraints of the map, such as its static two dimensions, fixed scale, inflexible information density, limited topics, rigid sheet boundaries, separation of spatial depiction from text explanation, and laborious and time-consuming drafting and revision procedures.

The model can take a flexible view of the geology as a set of interrelated things of interest (objects), which exhibit attributes such as properties, behavior and relationships. Its structure accepts classifications by distinguishing between object classes and instances (described in section 3) and recognizing class hierarchies (as in stratigraphy) within which attributes can be inherited. Potentially, it can relate observed objects to, say, the outcome of simulated processes, and track reasoning processes from observation to explanation or vice versa (Voisard, 1999). Being scale-free, the model can accommodate multi-scalar concepts like fractal representations. Objects described in mark-up languages, such as XML, can combine text, data and images readable by human beings with tags readable by computer, and thereby connect their respective strengths. The results of a DGSM can be stored and maintained in a hypermedia knowledge repository or object store, where the semantic web records relationships and reasoning.

The Semantic Web

Extensions to the Web aim to provide a language expressing both information and rules for reasoning about it (including rules in existing knowledge-representation systems), enabling people and computers to work in closer cooperation (CEOS, 2004). Sharing and processing information by automated tools as well as by people should lead to more effective discovery, integration and reuse of the information (Gahegan and Brodaric, 2002). The Extensible Markup Language (XML) allows users to extend their documents by embedding structures in them that computers can interpret (W3C, 2004; Cox, 2004; POSC, 2004). The Resource Description Framework (RDF) gives form to such a structure, as a triple of three linked items, like a sentence expressing the idea that an object has an attribute with a value. A Universal Resource Identifier (URI) can identify each of the three items, thus linking them to unique definitions available on the Web. The value can be another object, and the RDF statement can itself be regarded as an object or attribute.

There is thus scope to record, as part of an object's description or at the appropriate points in a document, inference rules that: define object behavior (its response to types of process), such as a class of surfaces being appropriately modeled in a specific way;

relationships, such as that the W, X and Y formations lie in a conformable sequence in the south but the W lies unconformably on the Y in the north; and reasoning, such as linking the form of a surface model to that of process simulations and thence drawing conclusions about its origin. A URI can point to anything, including a physical object, thus bridging the interface from the virtual to the real world.

Ontologies also play a part in the Semantic Web, each a published specification of a conceptual framework for a particular topic (NASA, 2004; W3C, 2004). They typically provide taxonomies that define classes of objects, the relationships between them, and sets of inference rules. They aim to be consistent and unambiguous and to integrate diverse viewpoints, facilitating sharing of content and services. 'Human endeavor is caught in an eternal tension between the effectiveness of small groups acting independently and the need to mesh with the wider community... An essential process is the joining together of subcultures when a wider common language is needed… The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web.' (Berners-Lee et al., 2001)

The Semantic Grid

The Semantic Grid brings together the ideas of the Semantic Web and the Grid. It supports the vision of e-science, helping scientists to work and collaborate more effectively by providing a language that expresses both data and rules for reasoning about the data. 'Our vision is of a generically useable e-science infrastructure, comprised of easily deployed components whose utility transcends their immediate application, providing a high degree of easy-to-use and seamless automation and in which there are flexible collaborations and computations on a global scale. The key to this is an infrastructure where all resources, including services, are adequately described in a form that is machine-processable, i.e. knowledge is explicit - in other words, the infrastructure provided by the technologies of the Semantic Web. The goal is semantic interoperability.' (De Roure, 2003).

The Semantic Grid adopts a service-oriented view (De Roure et al., 2001) 'based upon the notion of various entities providing services to one another under various forms of contract (or service level agreements).' Numerous software systems or agents that perform useful tasks on behalf of a user would interact in this setting. A service provider, such as a Geological Survey, would therefore have to ensure long-term digital preservation of trusted sources of evaluated, up-to-date knowledge (PADI, 2004). In due course, individuals and organizations must evolve to meet their changing roles within a more comprehensive and flexible system supporting the repository of scientific knowledge. For Geological Surveys, there is an opportunity to gain many benefits, mentioned next, provided essential features of the present system are not neglected.

2 THE BENEFITS

Worldwide, the work of Geological Surveys converges on geoscience maps and their accompanying explanations. These conventional products follow consistent standards and give extensive coverage at several fixed scales. The large format and high resolution of the maps enable users to focus on points of specific interest within a context supplied by peripheral vision. They generally correlate with topographic and other related maps. They are robust, highly portable, and can be used without supporting technology. They are assembled with care and consistent quality control. The maps are typically supplemented by text explanations that provide an illustrated background account of the geology of the area as a whole and of salient details, illustrated where appropriate by detailed maps and diagrams. Because they are formal publications, they provide a fixed reference point for geoscientists. They are widely available at reasonable cost, and copies are held indefinitely in several secure libraries. Revised versions are published when appropriate, generally after many years. The qualities of these conventional publications make it unlikely that they will be supplanted in the foreseeable future. But the end products constrain the approach and procedures of investigation.

Information technology (IT) has a growing role in the processes of geoscience surveying and in the preparation of maps and associated documents. Digital cartography can make the legacy of old documents available in new forms, which subject experts can enhance for inclusion in a system of digital spatial models. Field mapping can generate digital records directly. Modern information and communications technology introduces models that may lead to more effective methods of surveying and presentation. Adherence to standards makes it easier to find and exchange information (FGDC, 2004; USGS, 2004a, b). New systems can offer many benefits, but require careful design to retain existing valuable qualities, and to ensure that they tie in with related systems of wider scope (Brodaric, 2000).

An IT-based system can model spatial and other aspects of geological knowledge, including concepts and objects, their properties, relationships and processes. Anticipated long-term gains for users of such a geoscience knowledge system include the following:

Costs

The costs of preparing, storing, finding and using conventional documents are opaque, making cost comparisons difficult. However, at each stage the corresponding long-term costs of an IT system appear highly competitive, with more transparent costing leading to more efficient procedures.

Flexible structure

Maps have fixed scales, rigid content, and sheet boundaries unconnected to the significance and extent of geological objects. Digital spatial models can be seamless with no artificial boundaries of area, detail or subject matter. Unlike maps, models can handle complex systems and multi-scalar views of the geology, as in the self-similar dissipative-system models of sand bodies (van Wagoner et al., 2003).

Timeliness

Conventional publication is slow, and decades may elapse between data collection and publication. Much information is now collected digitally, and could be made instantly available, with access controlled by authorization procedures rather than by the mechanics and economics of printing and publishing.

Ease of access

Discovering and accessing relevant information from conventional sources is slow and laborious, particularly for users lacking the facilities of a major Survey. Web access, on the other hand, can be made available anywhere. Web search engines, metadata and browsers help the user to discover, filter and retrieve spatial and related information from many sources.

Connected information

Information from a single investigation is inconveniently segregated by form of presentation, such as: maps; memoirs; journal articles; textbooks; databases; archives of field records, samples and specimens; photographs; sketches; preliminary results; human knowledge. Multimedia and mark-up languages can link and combine information types at any level of detail.

Flexible presentation

Multimedia records can be filtered for relevance and assembled and presented to meet a wide range of purposes and contexts. For example, a map of a rock body could be shown on screen alongside text accounts of the processes of its formation and photographs of its outcrop, with options for, say, superimposing fitted diagrams of its origin and development, or displaying the consequences of cut-and-fill for road building.

Flexible visualization

The map is a single, rigid visualization. Spatial models provide information that can be visualized as maps, sections, fence diagrams, block diagrams, video sequences, correlation diagrams and so on, for user-selected areas, content, scales and projections. Thematic maps and multiple visualizations can overcome the information overload of the single map, clarifying the interpretation.

Formal structure

The object-oriented model separates metadata from detail, distinguishes between object classes and instances, and creates object-class hierarchies with multiple inheritance of attributes and behavior. It conforms well to patterns of geological thought and provides a clear framework for uniform standards on what to call things, what they mean, how they relate, where to put them and where to find them.

Reusing objects

By routinely identifying the properties and behavior of object classes, it should be possible to reuse object instances, such as the boundary of a rock body, for various purposes, such as thematic studies for land use, geological risk assessment, resource estimation, and other products custom-built for specific user needs. The Web and, in due course, the Grid will enable experts in any field to assemble objects and software from diverse sources, and record their methods in scripts that in turn may be reused in other contexts.

Reasoning

Object-oriented methods can link field evidence and conclusions to an ontology (for consistent terminology) and to objects describing aspects of theory and specific chains of reasoning. These could provide the evidence justifying the conclusions, help to evaluate the findings, and identify the wider consequences of new observations or ideas.

Integration

The interpretation depends on many aspects of geological reasoning, such as depositional, structural, and metamorphic models, which may have to be considered together to explain even a small feature. The rich connections within and among hypertext records can help to clarify links between the separate lines of reasoning at any level of detail.

Reconciling viewpoints

Object oriented techniques should enable information from many sources to be combined. Multiple sources reduce uncertainty, and narrow the range of plausible hypotheses. But the original sources may have adopted viewpoints with different objectives and emphasis, leading to conflicting results. Reconciling them requires the judgment of human experts, and the computer can record their reasons and justification.

The motives for transforming the information systems of Geological Surveys are to support better science more efficiently. Survey records could be connected in more detail with the historical configurations of geological objects and processes and their underlying physics, chemistry and biology, while positioning geoscience appropriately in the mainstream of development of the larger, unbounded knowledge system. It is subject specialists (not computers) that have the background understanding of geoscience and of their own ways of thinking that enabled them to build the existing knowledge system. They are best placed to reorganize that system by formalizing a framework within which IT can help users to record, store, find, relate, select, assemble, communicate and visualize their knowledge. Deploying these mechanical aids effectively calls for some unaccustomed introversion by geoscientists about their working practices and the consequences of change made possible by the emerging methods of IT.

3 ASPECTS OF THE GEOSCIENCE KNOWLEDGE SYSTEM

Two streams of IT development contribute to the potential benefits listed in section 2. The first is knowledge representation (Sowa, 2000), including abstraction and modeling; the second is communication, and its infrastructure of Internet and Web-based technologies. The two streams converge in the hypermedia repositories and knowledge systems supported by the Semantic Web and Grid. It may be useful here to clarify some terms, and to highlight some approaches and suggest how they fit together. Their combined relevance in an integrated system for geological surveying remains largely unexplored.

Standards

International standards extend the ability to communicate within a comprehensive knowledge system. Various information communities (such as CGI, 2004; ISO, 2004; OGC, 2004; POSC, 2004; W3C, 2004) are concerned with procedures and standards in specific areas, including easier access to information (GILS, 2004; INSPIRE, 2004; National Geological Surveys Committee, 2004).

Systems

A system is a collection of elements or components that are organized for a common purpose. System boundaries are somewhat arbitrary, but the systems approach brings together widely used concepts, vocabulary and insights from many fields of study. The system of geological surveying observes and measures features of the real world. Its procedures represent, record, interpret, explain and communicate the results, extending the system of scientific knowledge with conclusions that can be tested against reality.

Abstraction

The system that makes scientific knowledge comprehensible to the human mind requires a process of abstraction; that is, reducing the volume of information while retaining salient features and relationships. Geoscience surveying progressively reduces the full detail of the real world to observations, then to field records, then (by a process of interpretation, reasoning and explanation) to general conclusions drawn from specific instances, and summarization and generalization. It aims (by feedback from higher levels to detailed levels of abstraction) to extend or modify a shared paradigm of a parsimonious, unified account of the observed phenomena.

Reconciliation

The paradigm is shared at a general level, but detail may be specific to the investigation. The objectives, methods and opinions of the investigators determine which features are regarded as salient, and influence how the abstraction proceeds. Using information from one study in another context therefore requires filtering and adjustment, usually under human control, to reconcile information from one investigation with that of another, for the purposes in hand (Kent, 1978, Sowa, 2000, chapter 6).

Models

A model is an approximate representation, for a specific purpose, of some limited aspects of the real or imagined world. It aims to reduce their inconceivable complexity to a manageable representation. Hierarchical models can represent phenomena at several levels of complexity or detail. Geoscience survey models interface (through spatial coordinates) with the real world, where observations are made and deductions tested.

Object-oriented analysis

The computer-based system should work in harmony with human thought patterns. Object-oriented analysis (Budd, 2000, Coad and Yourdon, 1991) aims to build on the methods commonly used to organize our observations of the real world. It differentiates the observed world into things of interest (objects) and their attributes and relationships; distinguishes between the whole object and its component parts; groups individual occurrences (instances) of objects into classes with shared properties and behavior; and provides for hierarchies of classes and inheritance of properties within them. The objects can potentially be reused in various applications. Object-oriented design and implementation (Brodaric and Hastings, 2002), as opposed to analysis, raise many issues not considered here.

Spatial models

Geoscience spatial models aim to represent the spatial properties and relationships of geoscience objects within a consistent mathematical framework (Mallet, 2002). The behavior of the objects can be defined, relating them to reasoning from process models, and tying the interpretation to database management systems and geographic information systems (to discover, filter and retrieve relevant information from the object store), interpolation systems (for filling gaps and estimating uncertainty), and visualization systems (to display the results).

The general geoscience spatial model

The observations and interpretations of geoscience are linked to a more general conceptual model. This refers at all levels of detail to the three-dimensional disposition and configuration (where things are and where they are arranged) of the present-day observable and conceptual objects of geoscience, to their observed and interpreted properties and composition, and also to their history throughout geological time, including the processes that created or transformed them and are crucial to their interpretation. It is too vast and inaccessible for full representation, and so remains an abstract concept. But it provides the essential conceptual framework for relating and reconciling various facets of its fragmentary representation.

The challenge

The multiplicity of views in geoscience might be reflected in the infrastructure by a variety of big and little ontologies, with working connections established through linkages, in a system that encourages their rationalization, and reconciliation for specific purposes, in response to real user needs. The interests and viewpoints of the end users of geoscience knowledge differ from (and may be unknown to) the collectors of the information, and the ultimate users of geological knowledge are not generally geologists. The future infrastructure will develop to offer links to all areas of general and scientific knowledge. There is a need for geoscientists and their information communities to adapt their knowledge system to a new infrastructure devised by others, and to keep pace with changing user needs.

This points to the need for planning ahead to meet the challenges of geoinformatics. Portals, browsers, selection, interpolation, and visualization systems will be subsumed in the infrastructure and accessed through agents (programs that act in a self-interested manner on behalf of their users, while interacting with many other agents in the computer network). Collecting, finding and using geological information will be undertaken in a multitude of different contexts, not just on the terms set by one cohesive geoscience information community. The challenge is therefore, not to prejudge the end result, but to influence and participate in the development and exploitation of an infrastructure that can offer the mechanisms for diversity, evaluation, selection and inheritance that are required for the beneficial evolution of knowledge.

The infrastructure offers no competitive advantage from selective ownership of new technology. On the contrary, as with purchasers of a camera-phone, the utility of each owner's investment increases as the standard technology is more widely adopted. Early adopters might benefit by subsidy from bodies anxious to maintain a country or region's long-term competitive position. But the main advantages of an organization's early involvement lie in gaining time to evaluate the scenarios, analyze the processes within the organization, determine priorities, and achieve a smooth adjustment to the developing knowledge system without unnecessary disruption to staff and customers.

For example, the British Geological Survey (2004) currently provides a portal and browser for access to its information. But, in the outside world, enquirers dealing with, say, epidemiology or insurance risks might overlook the possibility that the portal, or indeed geology, could be relevant. In the longer run, customers may wish to access geoscience information through a general search engine, where it could be integrated with information from other sources, rather than through an entry point specific to one organization. If so, a flexible plan for transition is required some years in advance. Time is essential to plan ahead, explore alternatives and gain experience in unfamiliar fields, such as establishing trusted, service-oriented, digital systems.

Long-term adjustments may need immediate attention, such as reviewing the structure and boundaries of organizations and information communities, establishing tentative pathfinder standards, defining areas of priority research, and reconsidering field methods, staff training and balance of staff expertise. Decisions may also be required on the assignment and fulfillment of long-term responsibilities, such as digital preservation and maintenance (PADI, 2004, Fedora, 2004) and procedures for rationalizing geoscience ontologies, maintained in the light of the current general geoscience model.

4 EXTENDING THE VISION

The vision of IT in geoscience survey is still being adjusted to the concepts of database, computer aided design, geographic information systems, spatial models, object-oriented analysis, the Internet and the Web. Projects based on these concepts have evolved step by step, and are now incorporated in the procedures of many geological surveys, although not yet bringing the full benefits listed in section 2. Maps can be scanned, stored and reproduced; geological features can be vectorized and described in a database; map sheets can be edge matched; GIS-based spatial indexes can record and retrieve a wide range of spatial features; metadata can be standardized for geoscience records; spatial models can record and visualize the three-dimensional form of geological features; object-oriented methods help to rationalize the design; digital field recording integrates the system with electronic surveying tools, the Global Positioning System and (potentially) augmented reality; the Web can select and deliver the results. Not only is a vision defined, but also many parts of the system are implemented and doing useful work. Many standards, such as those referred to in section 3, are already in place. They give a firm base for planning the next steps.

There is now a clear need for extending this vision to match the developments in the cyberinfrastructure. For example, a seamless geological map should allow the user to specify the required area, type of content, scale, projection, and format. But simply edge-matching existing sheets will not meet these needs. Map explanations and marginalia, such as stratigraphic tables, cross-sections, and generalized vertical sections, prove difficult to keep in step with depiction of the map face for an arbitrary area.

The underlying problem is that the map is a static two-dimensional projection of incomplete fragments of the general geoscience model mentioned in section 3. The surveying process organizes the fragmentary information by long-standing procedures designed for paper documents. They hold back the introduction of user map design and other concepts, such as georeasoning (Gahegan and Brodaric, 2002, Pshenichny, 2003), or a view of Earth processes sharing the common physics of complex dissipative systems (van Wagoner et al., 2003). The problem cannot be resolved by technical advance alone: ideas are shaped by the manner of their representation, and geoscientists must be involved.

Linking models of spatial objects to an account of the reasoning process is desirable for at least three reasons. First, the surveying model must include evidence and reasoning to fulfill its role as the interface connecting geoscience knowledge and the real solid Earth. Second, all scientists should be able to explain their procedures, so that colleagues can follow their reasoning, evaluate and test their results. Third, the links are needed to establish how the knowledge base must be amended to respond to the inevitable revision of interpretations and observations. Taken together, these issues suggest that the vision of future geoscience surveying may be too narrowly focused on extending the geoscience map as a spatial model, and that the underlying rationale needs reconsideration. The human brain appears to handle spatial and narrative information along different pathways (Loudon, 2000, parts J, K). The separation reappears in geoscience survey in the distinction between maps or spatial models on the one hand, and text-based narrative explanations on the other.

Space has special properties, reflected in methods (topology, geometry and spatial transformations) for describing spatial pattern and relationships. Spatial thinking takes a comprehensive three-dimensional view, perhaps considering geological time as analogous (in some respects) to a fourth spatial dimension. ‘We can zoom in to see the detail, zoom out to see the spatial context, pan around to see the situation elsewhere, and compare spatial patterns arising from different topics. Narrative text cannot offer these abilities, but can be intimately linked to a spatial representation’ (Loudon, 2000, page A89).

Information in narrative form may refer to sequences of events, reasoning and explanations. In a similar way, narrative may describe properties by sequences of words and sentences, perhaps enumerating the characteristics of a rock, and proceeding from general comments to detail. In our memories and in written text or the spoken word, the narrative follows threads of thought, linked and woven together to make sense of our experience by defining objects and processes, drawing analogies, establishing relationships, abstracting and summarizing.

Markup languages enable surveyors to organize their ideas as a set of narrative accounts (objects), trace the threads of reasoning through hypertext links between the objects, and link the objects with their counterparts in a digital spatial model. All are potentially tied to a shared ontology through interactive templates, and developed together in the field and office. Spatial representations, such as the geological map or spatial model, fill the space depicted, resorting to interpolation, approximation and generalization where direct knowledge is inadequate. The spatial coordinates provide additional GIS links to, for example, stratigraphic metadata, cross-sections, databases and explanatory text objects.

The knowledge base cannot record the full general geoscience spatial model, which remains only as a concept. However, the overall structure and features of geoscience and its general model are described in many textbooks, and implied in bibliographic databases such as Georef (AGI, 2004) and metadata such as the Epicentre model (see POSC, 2004, 'Specifications'). They provide the framework, classes and categories for the indexing and cataloguing of geoscience information.

The same framework guides the surveyors and the users of the completed survey. Within such a framework, the user should be able to search among the diverse incomplete fragments of the general model, select the most relevant, and visualize them as filtered extracts, possibly requiring reconciliation by the user. The user's agent might apply filters to select from such aspects as: the area of interest; the business setting; the range of topics and levels of detail in each object hierarchy; the user's objectives and the emphasis placed on particular properties; the sources and evaluation of the information; the levels of resolution and generalization that match the scale of visualization; the levels of confidence, ranging from confirmed observation to tentative explanation; the spatial relationships valid at the selected levels; the appropriate spatial properties, such as existence, location, slope, form, texture, arrangement, relationships; the place of the objects and properties in chains of reasoning and process-response models. Visualization of the filtered extracts would then involve geometrical transformations including projection.

The ontology for the digital Geological Survey model is a component of this more general framework. The model might be thought of as representing a self-consistent, approved and authorized interpretation and explanation of the disposition and configuration of the rock units and aspects of their properties, composition, relationships and history, within a defined area or volume. The internal consistency of the model reduces the need for users to reconcile fragments. The model should enable users to view surveying as a process, not just as a product. It must be flexible to accommodate the variety of routes that surveying procedures can follow, and the levels of detail to which they refer. For example, the framework must be able to accept as input, not just multimedia streams of marked-up field notes and spatial models, but also existing visualizations (such as the legacy of geoscience maps) and related material, reworked or otherwise. The knowledge system should thus be able to start with conventional information and incorporate more advanced methods as they prove their worth.

The framework must support evolution by modification and replacement of items in the intertwined sets of sub-models, objects and methods for processing them. It must be amended and extended as new topics enlarge its scope. The ontology that specifies this conceptualization must accommodate more specialized ontologies referring to sub-models. Likewise, geoscience ontologies must take their place within the much broader framework of the cyberinfrastructure, for geoscience explanations reach out to other branches of science, and users of geoscience information approach it from many different business settings. In the long run, specific markup languages, ontologies and methods may be subsumed in infrastructure products of greater generality, to the advantage of all concerned.

No committee can assemble the wisdom to rationalize the diversity of the knowledge system. The Semantic Grid may therefore be designed to simplify and rationalize the plethora of methods and standpoints by selection through evolutionary processes, perhaps guided by the rituals of scientific discourse and of the market place, supported by the shared efforts of consortiums, standards committees, Geological Surveys, and other information communities dedicated to the common good.

5 CONCLUSIONS

The Grid extends the means of representing, processing and communicating human knowledge, by greatly increasing the capacity and power of computer networks and infrastructure services. Geological survey has long focused on cartography, and more recently on its digital equivalents. The large volumes of electronic data from digital spatial models and the larger volumes from exploration seismic data can be more readily transmitted through the increased bandwidth of the Grid. But its ability to accommodate detailed relationships may prove more significant. Spatial models and narrative explanations are built side by side, and the Grid provides the opportunity to maintain their close links at all levels of detail, linking data and observational evidence with reasoning, interpretation and explanation.

Markup languages have the potential to link text narrative at all levels of detail with spatial models, metadata, Earth processes, experiments, databases, images, material collections, and human experts. Collaborative development of standard frameworks and ontologies brings unified structures, classifications and vocabularies to this knowledge system, and encourage their rationalization, thus bringing together the results of many individual efforts. The Grid will enable Geological Surveys and similar organizations to provide services from which users (who best understand their own requirements) can obtain appropriate products or create their own.

Basic geology is at the core of the geoscience knowledge system, and geologists now have an opportunity to embrace a technology free of the constraints of pen, paper and printing press. Geoscientists can move their legacy of existing knowledge into the mainstream of a more comprehensive knowledge system, and ensure its appropriate digital preservation. Geological Surveys are well positioned to advance the transformation of this part of the knowledge system, while academic research, such as the GEON Program, brings new perspectives to geoscience surveying.

Acknowledgments

I am grateful to the British Geological Survey for making this work possible, in particular the help of Bill Hatton, Ian Smith and John Laxton, and to Boyan Brodaric of the Geological Survey of Canada for his encouragement in improving this paper.

6 REFERENCES CITED

(Web references as at September 2005)

AGI, 2004, American Geological Institute: GeoRef Information Services http://www.agiweb.org/georef/

Atkins, D.E., et al.., 2003, Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, 84 p. http://www.communitytechnology.org/nsf_ci_report/ExecSum.pdf

Berners-Lee, T., Hendler, J., and Lassila, O., 2001, The Semantic Web: ScientificAmerican.com, May 17, 2001. http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21

British Geological Survey, 2004, BGS GeoPortal: http://www.bgs.ac.uk/scripts/GeoPortal/home.cfm, BGS GeoIndex: http://www.bgs.ac.uk/geoindex/home.html

Brodaric, B., 2000, Digital geological knowledge: from the field to the map to the Internet: Digital Mapping Techniques '00 – Workshop Proceedings. U.S. Geological Survey Open-File Report 00-325. http://pubs.usgs.gov/of/of00-325/brodaric.html

Brodaric, B., and Hastings, J., 2002, An object model for geologic map information (14 p.): ISPRS Symposium on Geospatial Theory, Processing and Applications. Ottawa 2002. http://www.isprs.org/commission4/proceedings/pdfpapers/131.pdf

Budd, T., 2000, Understanding object-oriented programming with Java: Reading, Addison-Wesley, 420 p.

CEOS, 2004, Committee on Earth Observation Satellites – Semantic Web Link Collection: http://www.op.dlr.de/ceos/dstt/semantic_web.html

CGI, 2004, Commission for the Management and Application of Geoscience Information: http://www.bgs.ac.uk/cgi_web/welcome.html

Coad, P., and Yourdon, E., 1991, Object-oriented analysis: Englewood Cliffs, New Jersey, Yourdon Press, 233 p.

Cox, S., 2004, XMML - eXploration and Mining Markup Language: https://www.seegrid.csiro.au/twiki/bin/view/Xmml/WebHome

De Roure, D., 2003, The Semantic Grid: vision. http://www.semanticgrid.org/vision.html

De Roure, D., Jennings, N., and Shadbolt, N., 2001, Research agenda for the semantic grid: a future e-science infrastructure: Technical report UKeS-2002-02, UK e-Science Technical Report Series, National e-Science Centre, Edinburgh, UK, 37 p. 'Documents' at http://www.semanticgrid.org/ or http://www.semanticgrid.org/documents/semgrid-journal/semgrid-journal.pdf

EGEE, 2004, Enabling Grids for E-science in Europe: http://public.eu-egee.org/

e-Science Institute, 2004, UK National e-Science Centre: http://www.nesc.ac.uk/esi/intro.html

Fedora, 2004, The Fedora™ Project: An open-source digital repository management system: http://www.fedora.info/

FGDC, 2004, US Federal Geographic Data Committee – Geospatial metadata standards: http://www.fgdc.gov/metadata/metadata.html

Foster, I., and Kesselman, C., editors, 2nd edition, 2003, The Grid: blueprint for a new computing infrastructure: Amsterdam, Morgan Kaufmann, 748 p.

Gahegan, M., and Brodaric, B., 2002, Computational and visual support for geographic knowledge construction: filling in the gaps between exploration and explanation (19 p.): Symposium on Geospatial Theory, Processing and Applications. Ottawa 2002. http://www.isprs.org/commission4/proceedings/pdfpapers/154.pdf

GEON, 2004a, Cyberinfrastructure for the Geosciences: http://www.geongrid.org/

GEON, 2004b, Ontologies: http://geongrid.org/resources/ontologies.html

GILS, 2004, Global Information Locator Service: http://www.gils.net/

Hinke, T.H., 2004, Grid technology as a cyberinfrastructure for Earth Science applications: NASA Earth Science Technology Conference, June 2004, 8 p. http://esto.nasa.gov/conferences/estc2004/papers/a3p1.pdf

INSPIRE, 2004, Infrastructure for Spatial Information in Europe – European Geo-Portal: http://eu-geoportal.jrc.it/gos

ISO, 2004, International Organization for Standardization: http://www.iso.org/iso/en/ISOOnline.frontpage

Kent, W., 1978, Data and reality: Amsterdam, North-Holland Publishing Company, 211 p.

Loudon, T.V., 2000, Geoscience after IT: Computers & Geosciences, v 26, no. 3A, 142 p.

Mallet, J.-L., 2002, Geomodeling: Oxford, Oxford University Press, 599 p.

NASA, 2004, Semantic web for earth and environmental terminology: SWEET ontologies: http://sweet.jpl.nasa.gov/ontology/

National Geological Surveys Committee, 2004, Canadian Geoscience Knowledge Network: http://cgkn.net/

OGC, 2004, Open Geospatial Consortium: http://www.opengeospatial.org/

PADI, 2004, Preserving access to digital information: http://www.nla.gov.au/padi/index.html

POSC, 2004, Petrotechnical Open Standards Consortium: http://www.posc.org

Pshenichny, C.A., 2003, Conference Reports, Georeasoning Workshop: IAMG Newsletter, no. 67, p.15, 19. http://www.iamg.org/newsletter/newsletter.html

Research Councils UK, 2004, Links to e-Science on the web: http://www.rcuk.ac.uk/escience/links

Smith, I.F., 2001, The Digital Geoscience Spatial Model – The shape of the BGS of the future: Earthwise, issue 15, p.8. http://www.bgs.ac.uk/magazine/magazines/docs/15/23.pdf

Sowa, J.F., 2000, Knowledge representation: logical philosophical, and computational foundations: Pacific Grove, California, Brooks-Cole, 594 p.

USGS, 2004a, International Clearinghouse for Geologic Map Databases and Standards: http://ngmdb.usgs.gov/intdb/

USGS, 2004b, The North American Geologic Map Data Model: http://www.nadm-geo.org/

Van Wagoner, J.C., Hoyal, D.C.J.D., Adair, N.L., Sun, T., Beaubouef, R.T., Deffenbaugh, M., Dunn, P.A., Huh, C., and Li, D, 2003, Energy dissipation and the fundamental shape of siliciclastic sedimentary bodies: AAPG Search and Discovery, Article #40080, 9 p. http://www.searchanddiscovery.com/documents/vanwagoner01/images/vanwagoner.PDF

Voisard, A., 1999, Abduction and deduction in geologic hypermaps: Lecture Notes in Computer Science, no. 1651, p. 311-329. http://citeseer.ist.psu.edu/voisard99abduction.html

W3C, 2004, World Wide Web Consortium: http://www.w3.org/

Extensible Markup Language (XML): http://www.w3.org/XML/

Resource Description Framework (RDF): http://www.w3.org/RDF/

Semantic Web: http://www.w3.org/2001/sw/

Web Ontology Language Overview (OWL): http://www.w3.org/TR/owl-features/