Steps toward Grid-based geological survey: Suggestions for a systems framework of models, ontologies and workflows

T. V. Loudon
J. L. Laxton
British Geological Survey, West Mains Road, Edinburgh EH9 3LA, United Kingdom

Abstract

Consideration of an explicit systems framework for geological survey information is timely, to assist in developing and maintaining an integrated and coherent view of regional geoscience in a Grid-based context. A framework based on a solid Earth systems model is tentatively proposed in this paper. The developing advanced infrastructure of information and communications technology, the so-called Grid, points to more flexible global communication that will help to overcome artificial boundaries and divergence of concepts from separate places and scientific disciplines. Interoperability of information (the ability to amalgamate and work with concepts, terms or models from various sources, and thereby share and reuse information) will be a key to the Grid's success. Geological surveys can respond to the opportunity by changing their emphasis, away from publishing printed maps and related documents, towards maintaining a geoscience knowledge system from which scientific workflows can provide flexible services that match requirements specified by the user. The changing system should fit with, and build upon, existing patterns of human thought and the published record; include interpretation as an essential part of the conceptual building blocks that support geologists as they abstract, codify and reason, link observation to explanation, and predict what they have not yet observed; support improved representations of the geology; and encourage the use of generic concepts and ontologies, following international standards where appropriate.

Keywords: Survey organizations; information systems; expert systems; ontologies; workflows.

1 Introduction

Geologists generally take for granted the support of the conventional infrastructure that enables them to carry out an investigation and assemble and communicate their findings. The infrastructure ranges from field maps and notebooks, compasses and microscopes, to downhole loggers, seismic recorders, and the drafting, printing and publication industries. Reflecting the exponential growth and plummeting costs of computing power, storage capacity and communications bandwidth, the infrastructure is increasingly based on information and communications technology. So far, this cyberinfrastructure largely mimics familiar procedures, with little direct impact on the nature of the

geologists' work. Nevertheless, it may soon transform the environment for conducting geoscience. Geologists must consider the implications in order to guide the response of their science.

A plausible, wide-ranging vision of the developing cyber-environment is based on the so-called Grid (Foster and Kesselman, 2003), which would supply digital information as a commodity, similar to the way in which the electricity grid delivers power. It could lead to a global system of interconnected representations of knowledge where irrelevant boundaries of place and discipline lose their significance. Items of information from many sources can be extensively linked in a Grid-based system. The content can be structured to support reuse of items in different contexts, with computer-readable descriptions (ontologies) and paths of linked activities (workflows) that lead to more consistent, flexible and economical representations. These representations will potentially be accessible at the user's desktop or digital field notebook, where information from many sources could be assembled to work together as an integrated whole. For brevity, the term 'Grid' is used here in a broad sense to refer to an advanced cyberinfrastructure based on these concepts.

In this paper, the consequences are looked at from the viewpoint of a geological survey. They affect its business model (why the work is done), the surveying procedures (how it is done), and the framework (the structure that relates the components of geological information). The framework is the focus of this review. It is required in order to integrate, communicate and understand the multifaceted information, providing a shared logical structure for relating items through classification and organization. The framework should foster the collaborative nature of science by encouraging individuals and organizations to work together in building a shared system of knowledge, with standards to ease communication. If we can design one coherent, integrated framework for the geoscience component of a global knowledge system, then individual projects can be constructed as steps towards the collective goal of better understanding. It should support revised business models, changes in geological surveying procedures to match emerging opportunities, and services that provide users with products to meet their specific needs.

We suggest that a suitable framework can be based on a model of the systems of the solid Earth: a solid Earth systems model (sEsm). 'Model' is used in the sense of a conceptual construct representing a simplified view of some aspect of reality for a particular purpose. 'System' is used in the sense of a set of interacting parts operating as a whole, organized to perform a particular function. Just as metadata can be thought of as data describing more detailed data, so the framework might be thought of as a metamodel, or a model describing the sEsm. Within the overall systems model, there is of course an extensive hierarchy of component subsystems and component models that are concerned with more specific aspects and with their interactions.

The informal diagrams illustrating various aspects of the system of geoscience knowledge may help to clarify the issues involved, at the risk of oversimplifying a complex situation. They (Figs. 1-5) are intended to indicate the various components of the system and their linkages, and to suggest how they might relate to the global knowledge system potentially supported by the Grid. They aim to clarify how digital components can work alongside existing knowledge, build on existing procedures, make them more effective, and encourage the introduction of improved methods for surveying. The arrows indicate the flow of information through the system.

Geological survey agencies take a comprehensive view of the basic geology of a region and are therefore well placed to take a lead in defining the framework. Preparation for Grid-based geological survey is timely, because evolution of the cyberinfrastructure is proceeding at a pace that surveying agencies cannot match. However, the design must involve geologists and earth scientists generally, as well as users of geological information. This paper is directed primarily at geoscientists, aiming to clarify the requirement for a framework and provoke deeper consideration of possible responses. It considers how geologists extend their existing knowledge by the process of surveying in section 2, the mechanisms for sharing information in section 3, the opportunities

offered by the Grid and their consequences in section 4, and some steps towards implementation in section 5. Technical solutions are not considered in this paper, but in practical terms, the delivery mechanisms for geological survey products must attempt to keep pace with mainstream acceptance of cyberinfrastructure tools.

2 Gathering knowledge

2.1 Background knowledge

The system of geoscience knowledge has many interacting aspects that should work together as a whole to maintain a configuration that best describes, predicts and makes available salient aspects of observable geology. It should respond to the evolution of relevant aspects of knowledge, controlled by the self-adjusting mechanisms known as feedback. A Grid-based system implies wider sharing of knowledge, and significant changes to its representation, structure and handling. To see how wider sharing can be achieved, we need to develop a considered opinion on how the system works now, what should change, and what must be preserved. For present purposes, we need to identify aspects that must be kept in mind in designing the framework, starting with the knowledge already in the geologist's mind, followed by the objectives and procedures of surveying that extend that knowledge, using existing information records and creating new ones.

The greater part of the knowledge in the system is not recorded at all, but is held in the collective human memory – the background knowledge (on left of Fig. 1) of the human contributors and users of geoscience knowledge, acquired by training, education and experience. If we think of that knowledge as originating and existing in human brains, then information might be seen as a means of representing aspects of it, which can then be recorded, stored and communicated, ultimately interacting with knowledge in the recipient's mind. There is much philosophical debate on what knowledge is, the mechanisms by which representations of it are expressed and communicated as information, the terminology to describe it, and how much is unexpressed or even inexpressible. The main point here, however, is that the basic design aim in a Grid-based geoscience knowledge system is to improve the mechanisms that handle and supply information to assist human thought processes. The complicated interactions between background knowledge and recorded information are therefore relevant to the framework design.

Walsham (2005) provides an accessible review on knowledge communication. He builds on ideas from Polanyi and Giddens, and provides a range of related references. He argues that each of us has different tacit knowledge of the world in which we live, the outcome of active shaping of experience undertaken in the pursuit of knowledge. 'Tacit' knowledge (from the Latin tacitus, meaning silent) is understood or implied without being directly expressed. It cannot be captured, translated or converted but only displayed or manifested in what we do. Subsidiary aspects for communication, such as gestures, speech, text narrative, formulae, maps, or graphs, have no meaning unless rooted in this deeper knowledge. Walsham maintains that in communicating these subsidiary aspects (information), the results of action and reflection are 're-presented' in such a way that they can be 'read' by others (transplanting their sense into the cultivated seed-bed of the recipient's own tacit knowledge) and interpreted by them in terms of action and reflection. They are deeply involved in human processes of communication that cannot be divorced from their context. He argues (page 16) that the system must enable effective interaction between people with different tacit power and understanding, and should be concerned with disputed opinions as well as consensus views. In the present case, the framework design should take into consideration that not all geologists have the same background understanding and not all agree on every interpretation.

Many aspects of geological fieldwork depend on tacit, procedural knowledge that is learned from experience and demonstration rather than verbal instruction (Loudon, 2000, page A90). For example, a trained eye might readily correlate one outcrop with others by comparing it with hand specimens from nearby exposures, but be unable to do so from the most exhaustive description. Mechanisms for consultation with experts are thus part of the knowledge system, and are essential for communicating unexpressed knowledge.

Walsham (2005) provides references to the literature on knowledge communities, and communities of practice within organizations and between people in different organizations, coming together across boundaries to learn and 'share' knowledge on particular topics. He draws on Giddens' structuration theory, with its three inextricably linked aspects: how things are represented in communities (interpretive schemes), what is represented and for whom (norms), and who requires what information for what purposes (power relations).

This sociological view of knowledge sharing is relevant to the detailed design of future knowledge management systems, in particular to the mechanisms for monitoring the scientific validity and utility of the products in a global context, and for assessing the value that users and funding bodies place on each part and item of the geoscience knowledge system and on the system as a whole. These aspects are considered in section 3.1, in the context of the provenance and procedures for acquiring the knowledge.
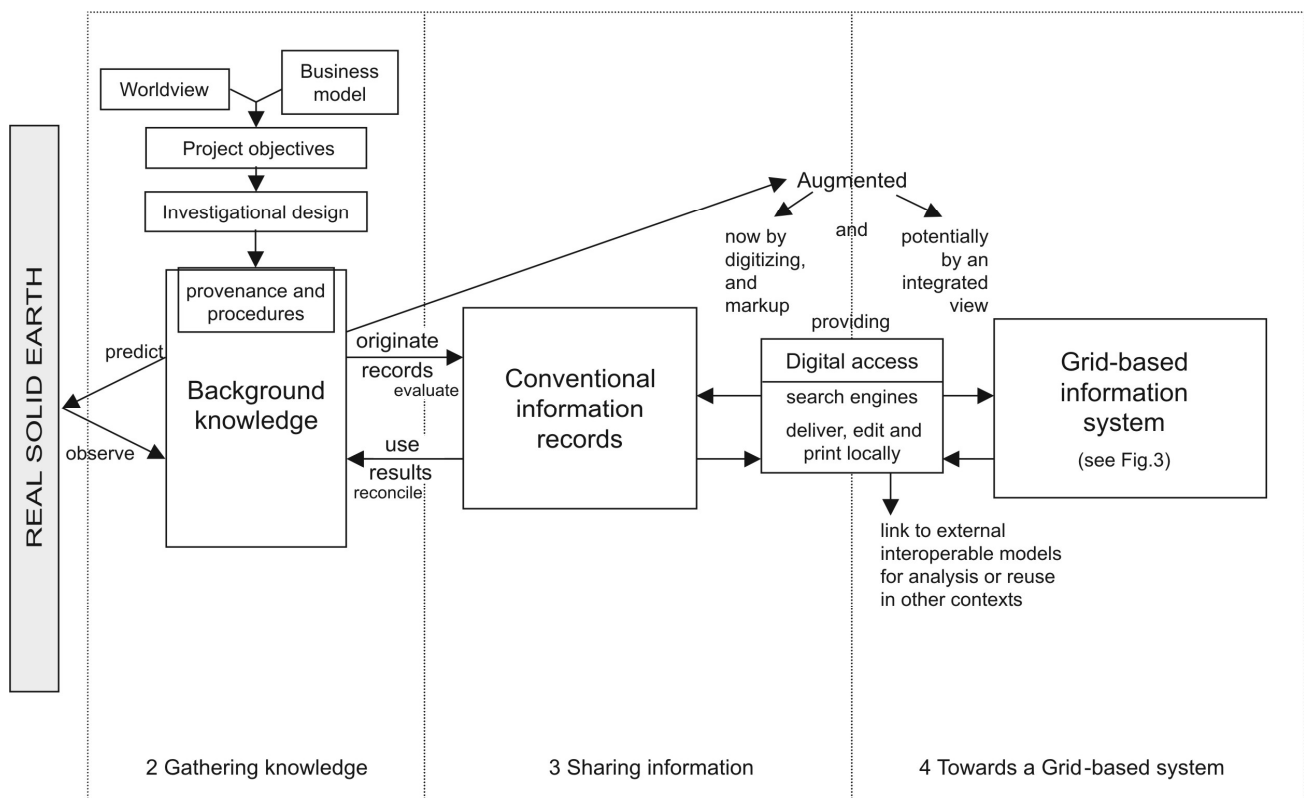

Figure 1: The geoscience knowledge system. Salient aspects of the real world, selected by prediction from background knowledge in accordance with a worldview and business model, are observed, and selectively described by conventional information records, for access by users. The process is augmented now by digital access, potentially by a Grid-based information system.

2.2 Provenance

Understanding an item of information (Minsky, 1981), whether obtained by observation or from an existing record, requires it to be placed in a frame of reference (the context of relevant aspects of background knowledge). For perfect communication, the information supplier and user must share the same frame. Inevitably, however, they have separate viewpoints and sets of objectives. Kent (1978, pages 202-203) pointed out that no two people perceive reality in exactly the same way, and each can take conflicting views at different (or indeed the same) times. This can cause confusion, but also makes possible reconciliation, which he defines as a state in which the parties involved agree to a shared view relevant to the limited purpose at hand. He points out that reconciliation is an every-day tactic for narrow purposes, but as more parties interact for wider purposes, discrepancies in fundamental assumptions will become increasingly apparent.

Members of an information community, such as a geological survey, fine-tune the alignment of their frames of reference as a learning process while working together. Similarly, geologists gain insight into the background knowledge of the users of their work, and therefore of suitable forms for communicating information to them. The move to a Grid-based system will greatly increase the volume and diversity of information and frames of reference. For example, the background knowledge and priorities of geological investigations supporting civil engineering differ considerably from those directed at oil exploration. Only by understanding attributes of the information's provenance (where it came from), can the originator and user of information align their background knowledge by a temporary partial overlap of their viewpoints. This enables them to extract information from other sources that is relevant for their own purposes.

Geological surveyors predict from their existing knowledge where and how they can best make observations of the solid Earth. As they predict and observe (Fig. 1), they learn more, modifying and adding to salient (important and meaningful) aspects of their background knowledge. Typically, surveying proceeds as a set of projects (managed activities with objectives, resources and structure). Each project may have its own investigational design, including such aspects as instrumentation and sampling schemes. The design guides selective abstraction during observation, recording, description and explanation. The information may be acquired by mechanical data collection following predetermined procedures (as in some geophysical and geochemical studies), or from more flexible, holistic, surveying procedures (section 2.3), but in both cases the procedures are directed by existing knowledge.

Relevant data (Fig. 1) on the worldview (the context as seen by the investigators), business model (the objectives behind the investigation and the means of achieving them), project objectives, and investigational design describe the provenance (how the information came into being). The provenance should be recorded and made available as metadata to assist in defining valid applications and in reconciling information from various sources. Future digital field support systems could record such information as the fieldwork proceeds. Differences in provenance and procedures may be essential for efficiency and for the evolution of new techniques and interpretations. But a shared overall framework and standardized ontologies can help to avoid pointless and unnecessary variation, and the metadata can assist reconciliation and provide a context in which disputed issues might be defined and perhaps resolved.

2.3 Procedures and reasoning

As shown in Fig. 1, background knowledge is linked to observation of the solid Earth, to the investigational design, to conventional information records, and potentially to a Grid-based information system. Many diverse designs apply to geological investigations, but the business models of most geological survey agencies share a common purpose, with wide scope and range of

application. This purpose is to develop, record, maintain and communicate an authoritative, coherent account of the geology of a region. It is achieved by integration (through reconciliation) of knowledge from relevant available sources. The field geologist maps the structure and boundaries of the areas thought to be underlain by predetermined stratigraphic classes, in accordance with many lines of observation and reasoning (Harrison, 1963, page 227), such as systematic field observation and measurement of salient geoscience properties at sparse outcrops, establishing possible continuity with similar material nearby and correlation with prototypes (type sections), reconciliation with other sources of information, and analysis and interpretation of the properties in the context of the rock types they characterize. The emphasis is generally on understanding the nature, distribution, history and configuration of the rock types, required as background information for a wide range of commercial, regulatory and research activities. Conventionally, this has centered on the lithostratigraphic map and its explanation.

At each stage of investigation, geological surveyors take a top-down view, look carefully at the many aspects of what they know so far, and imagine how the situation might be in its entirety. Field geologists thus gain the necessary holistic understanding. It could be argued that they follow gestalt principles, including closure, similarity, proximity and continuity, which have much wider applications, for example in computer visualization (Skaalid, 1999), and that interpretation and visualization might be regarded as related aspects of a single process. It is this comprehensive view that gives the lithostratigraphic map its core significance in geoscience.

Actively seeking the most informative observations, the geologists refine their space-filling interpretation, for which they must rely on understanding beyond their limited observations and incomplete evidence. They are likely to avoid a predetermined, rigid sampling scheme, and instead follow an exploratory approach, modified as the survey progresses and more is learned. An item of information, regarded by the surveyor as an uncertain prediction from an evolving interpretation, should not be seen by the user, or incorporated in the user's model, as isolated factual data, but must connect to the context that gives it meaning.

Surveying reduces the unimaginable quantity of information inherent in the solid Earth to a manageable amount of information, which initially contributes to the knowledge in the surveyor's mind. The results may be jotted down, as ephemeral records for the surveyor's own use, or reworked in the geologist's mind to extract information that can be shared informally with colleagues. Each geologist is an originator and a user of shared information, perhaps both at the same time. The functions 'originate' and 'use' are shown separately in Fig. 1 to distinguish the roles (not the individuals) at the boundary between background knowledge and shared information. Information is evaluated as part of the recording process, and some reconciliation of viewpoints may be needed to use records that originated in another context – activities that require human control.

In field mapping, geologists are concerned with the configuration of objects and their properties in present-day space. An object can be regarded as the representation of an entity or thing of interest, an object instance refers to an actual occurrence of that object, and an object class is a category to which instances can be assigned within a defined, usually hierarchical, classification scheme. In classifying geological objects, such as the rock-types depicted on a lithostratigraphic map, the surveyors assess and take into account their interpretation of the genesis and historical development of the present-day geology.

The interpretations and explanations are rooted in concepts of object configurations in the geological past, and the processes and events that created and transformed them, leading eventually to the state and configuration of the present-day objects. Geology is a historical science. "The unchanging properties of matter and energy and the likewise unchanging processes and principles arising therefrom are *immanent* in the material universe. They are nonhistorical, even though they

occur and act in the course of history. The actual state of the universe or of any part of it at a given time, its configuration, is not immanent and is constantly changing . . . History may be defined as configurational change through time, i.e. a sequence of real, individual but interrelated events. These distinctions between the immanent and the configurational and between the nonhistorical and the historical are essential to clear analysis and comprehension of history and of science." (Simpson, 1963, page 24).

The idea of immanent processes and changing configurations fits well with the concept of emergent systems, which create patterns that appear to arise spontaneously through the interaction of adjacent parts according to simple rules, without central control (Van Wagoner et al., 2003, Nicolis and Prigogine, 1989). The rules may stay the same, but the behavior of a complex system of this kind is unpredictable in detail because of the influence of feedback effects. Nevertheless, it tends to evolve to a preferred pattern, often within a hierarchical structure of self-similar patterns. The behavior of the system as a whole cannot be explained by reducing the phenomena to simple parts controlled by mechanical processes governed by the deterministic laws of physical science. Instead we must, and do, view the system holistically as a coherent organized whole.

The unchanging nature (invariance) in time and place of the mode of operation of geological processes, and the conservation of certain properties (mass, energy) is the basis for geological explanation ("The present is the key to the past") and emphasizes the leading role of immanent processes in the knowledge system. Geological surveying therefore takes a model-based viewpoint, in which each object that the surveyor describes and records can be seen as an outcome of a likely scenario (a possible course of local operation of a geological-process model in its historical setting). The model, as a holistic interpretation of the observed properties, has more weight in the interpretation than the properties on their own. For example, siliceous grit directly overlying the granite from which it had been eroded would not be mapped as part of the granite because, in spite of their contiguity and similarity of appearance and physical and chemical properties, they are the results of quite different processes, far removed in time and environment – objects of similar character, but the outcome of quite different scenarios. Thus the interpretations central to geoscience surveying associate dynamic models of the geological past with their static present-day outcome. The resulting lithostratigraphic classification of rock types brings deeper understanding, and thus greater and more general predictive power, than a classification based solely on rock properties. The properties alone may, of course, be sufficient where the product can be taken to imply the process.

To reduce the discomfort of basing scientific pronouncements on subjective interpretation, the nomenclature may be based on factual observation of the object's properties, with the advantage that a change of opinion about the scenario need not require renaming the object. A line of reasoning may take objects as surrogates for the processes that formed them, and appear superficially to depend on the objects, not on the processes they are thought to imply. Thus, it might be argued that a formation was deposited in deep water during a cold period, because of its large extent, graded bedding and sole marks (thought to indicate turbidity currents), and because it merged in one direction with laterally extensive breccias and diamictites (thought to be slumped beds) and in the opposite direction with fine siltstones, containing dispersed boulders (thought to be ice-borne drop-stones). A full record of the reasoning, however, would include the links between objects, processes and interpretation as a hypertext sequence, or workflow of the thought processes. A shared ontology could identify and re-use workflows to reduce redundancy, and simplify the description. Each thread of reasoning could be identified and described in an index (Fig. 4), although the networks of relationships through the objects and process-models of the knowledge system cannot readily be shown in a simple diagram.

The so-called semantic Web (Berners-Lee at al., 2001) and proposed semantic Grid (De Roure, 2007) provide mechanisms to connect information to the underlying reasoning, and can relate models across disciplines through shared elements of the reasoning processes, the goal being semantic interoperability. To take an example within geoscience, data in maps from gravity survey and in maps from geochemical analyses of stream sediments for the same area, cannot be correlated directly. Each might reflect the effects of, say, a hidden granite pluton, but one reflects the density distribution at depth, while the other reflects mineralogical changes around the granite modified by later stream transportation. Only by considering the meaning of the datasets from a shared viewpoint, in this case through a geological interpretation, can the results be reconciled and integrated.

3 Sharing information

3.1 Evaluation

Surveyors informally evaluate as they observe (weigh the relevance and importance of observations), and commit selected information to memory, possibly supplemented by notes and sketches that may or may not eventually contribute to the permanent record. Processes such as abstraction, codification and reasoning (selectively reducing the volume of information, representing it in a standard, systematic form, and drawing conclusions on the basis of the evidence) then lead to progressively more general statements or conclusions. They may be informally recorded, and can provide feedback to improve prediction and assist further knowledge acquisition by observation. They may subsequently result in records that can be permanently stored and made generally available by publication. Further evaluation (Fig. 1) is then needed to ensure that the individual contributions are acceptable input to the shared record, and to keep track of their relevance and significance as the science evolves.

Within geological survey agencies, internal controls ensure that they publish valid, consistent products, which maintain the reputation of their 'brand name'. Editors and reviewers formally assess external papers. Later authors who quote, support or criticize previous publications implicitly evaluate them through their links and comments. Unevaluated and informal records may be held in archival collections to support the publications, but are normally accessed through a local expert who can point out their shortcomings and limitations.

Evaluation of shared information must ultimately reflect human judgment, but it is at present a laborious process that delays communication. As information technology develops, the systems will offer earlier availability. Publication takes on a new meaning as mechanisms develop for rapid exchange of smaller components of the information, and evaluation by current users is given greater weight. There may be scope for embedding evaluation criteria in algorithms that give flexible support to human judgment. Consideration of who requires what information for what purposes (the power relations mentioned in 2.1) would ideally lead to clear objectives defined within the business models. Explicit measures of success, linking objectives and results, might help editors, reviewers and managers to evaluate each contribution, and users to identify relevant information.

Surveying modifies, refines and extends what is already known. Its objectives could be defined in terms of prediction and generalization. For example, lithostratigraphic maps predict a wide range of rock properties in the context of their age and origin, reflecting the holistic approach. They therefore provide predictions that can be generalized and extended to all other aspects of the geology. The results of a survey might be regarded as a predictive representation – one that describes the real world in terms of predictions potentially verifiable by future (or temporarily withheld) observations. Thus the surveyor does not know that a particular geological formation will

underlie the entire area shown in the appropriate color on a lithostratigraphic map, but thinks this likely, implicitly predicting that further observations would confirm it. It is this test against the real world that ultimately determines the soundness of the interpretation.

One possibility for assessing the value of results is to consider surveying as an example of reinforcement learning (a learning system that encourages progress towards defined objectives). Sutton and Barto (1998) discuss reinforcement learning in the context of artificial intelligence, and Rafols et al. (2005) argue that predictive representations generalize well. "The *predictive representations* hypothesis holds that such representations are particularly good for generalization. A good representation is one that captures regularities of the environment in a form useful to the learning agent; and in a reinforcement-learning task, something is 'useful' if it increases the agent's ability to receive rewards. Thus, representations generalize well when the regularities they capture allow an agent to learn more efficiently how to increase its cumulative reward."

This hypothesis applies to human behavior, including the acts of selection and generalization involved in the process of geological surveying. The surveyor might, for example, improve the predictions implied by the map by adjusting the delineation of unexposed boundaries as a result of learning more about the behavior of the local rocks as the survey proceeds (capturing regularities useful in improving prediction). The surveyor thus gains the rewards of satisfaction and kudos from producing a better map and an enhanced reputation by explaining the more general geological consequences. But the hypothesis also applies to software agents searching for relevant geological information, by reinforcing their learning with 'rewards' for successful searches (in the form of adjustments to a parameter in the algorithm). Similarly, users of a digital field support system might align its reward system with their own objectives, with a view to helping them weigh the relevance of potential observations, and recognize and test more general concepts that extend their predictive power and usefulness.

The shared business models within a geological survey, and their similarity in different survey agencies, ensure that the products of the various projects, though differing in detail, contribute to a coherent body of geological knowledge. This adds value to their products, partly because the shared understanding brings greater generalization. In other words, what is learned about individual topics and areas can be shared and applied more widely, with less need to restrict the frame of reference for reconciliation. Standardization of concepts and their representation also adds value by wider sharing of processing tools, leading to more efficient handling and analysis of the information. The human and digital reward systems within the knowledge system must take account of the value that these components contribute. The framework must be designed to accommodate not only existing methods of evaluating information, but also partially automated and more complex procedures for evaluation as they develop.

3.2 Abstraction and context

The results of geological investigations can be selectively formalized, recorded, stored and communicated as conventional information records, possibly augmented by a system for digital access, as shown in Fig. 1. The greater part of shared representations (as opposed to tacit knowledge) is held and communicated in the conventional form of publications and archived collections. They include information at all levels of detail from field notes and data records, published summaries, and interpretations in the form of maps and accompanying reports, to less detailed maps at smaller scales, regional guides and overviews in scientific papers and books. All these records are the results of an abstraction process that reduces the volume of information while retaining salient points. There is feedback at all levels to maintain a coherent view, ensuring that detailed and summary records correspond, and that new and existing records are consistent.

The knowledge of each individual is inevitably specialized. But shared understanding, gained by education and experience, including study of one another's work, ensures that the overall understanding of geoscientists provides a coherent view of their own specialist field and of all aspects of their science at least at the level of a paradigm, where concepts are stable though not immutable. Kuhn (1962) describes the paradigm as comprising universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners. "The paradigm provides a map whose details are elucidated by mature scientific research, and since nature is too complex and varied to be explored at random, that map is as essential as observations and experiment to science's continuing development" (page 108). In communicating with a specific community, an author might assume familiarity with this background knowledge.

Beyond this background understanding, the information required to understand a document is either contained in it or in papers that it cites. The scientific paper, report or map is the typical unit in which geoscience information is recorded, referred to, and communicated. It may be concerned with recording observations, interpretations or both in a wider context that explains their significance. Or it may place them in a new context that extends their significance or makes the findings available to a new audience. As users differ in the extent and focus of their knowledge, objectives and outlook, the same information may be presented separately for different audiences. The users must consider the viewpoints of the originators when reconciling the information with their own viewpoints and needs.

The context of an individual document helps the user to understand it, following well-established conventions. For example, the map is a core product of most surveys. Geological maps are static and restricted to two dimensions, fixed scales, rigid sheet boundaries, inflexible visualizations and limited information density. The map representation is separated physically from the text reasoning and explanation; it contains hidden ambiguities and deals inadequately with uncertainty. Context, however, helps to overcome these deficiencies and explain the meaning of information on the map face. The colored areas that show the distribution of stratigraphic categories are overprinted on a topographic base map and geographic grid. The boundaries of each map are edge-matched with adjacent sheets. Map marginalia may include an explanatory key and stratigraphic table, cross-sections, generalized vertical sections and other diagrams and text, and a list of authors, organizations and dates that indicate the map's provenance.

The map may have a corresponding but separate narrative map explanation (text and illustrations) that cites relevant papers, justifies the interpretation of the geology and places it in the context of its geological and investigational history. The map and associated records provide information about the geological objects that are displayed on the map face, placing them in their spatial and stratigraphic context and describing a range of properties including their composition, spatial form and relationships. Their origin and geological history may be described in the text explanation, containing spatial references that can be located on the map, and thence on the ground.

The broader context of maps is given by summary information in the marginalia and by maps generalized to a smaller scale. Summary information for books and papers is provided by their titles, tables of contents, indexes, and abstracts. International efforts have achieved a high level of standardization in such areas as stratigraphic nomenclature (International Commission on Stratigraphy, 2007), an important aspect of the geoscience framework expressed in an explicit form. The formal structures of stratigraphic tables and map series have informal but widely understood extensions. Thus, the conventional framework has built-in aids to find, relate, understand and summarize relevant information.

Information technology now augments these conventional systems with, for example, databases, geographic information systems, and spatial models. It offers an alternative means of access through search engines with complex relevance criteria to find appropriate documents, and digital

delivery of papers, maps and images to examine, edit, manipulate and print at the desktop (Fig. 1). Conventional records can be located through bibliographical indexes, and digital text can be searched for relevant words or word combinations. References, indexes and databases also help to find more detailed archived information supporting the map interpretation. This information includes field notes, samples and specimens, photographs, and other related information such as geophysical or geochemical studies and borehole records. Potentially, the Grid-based information system will share and extend this entry point. The ease of linking items in a digital environment suggests that more powerful hypermedia models will connect the information from the various documents to overcome the distinction between map and context, and integrate all types of information at all levels of detail.

3.3 Information types

In the conventional framework just described, various types of information are combined in ways we take for granted. But a different way of looking at the information may clarify another aspect important to its organization. Information is represented in various ways, each corresponding to a memory type (Pinker, 1997, Loudon, 2000, pages A80-81) that is dealt with in a different area of the human brain (Fig. 2, box 1). Thus, short-term memory (at the bottom of box 1) holds accurate but brief memories of observed properties and comparisons, with results that can be recorded (promptly) as field notes or a database. Spatial memory deals with relative locations, sizes and shapes – information that can be captured as maps and sketches. Episodic memory is less reliable than short-term memory, but lasts longer, and allows us to recreate in our minds sequences of past experiences and events, extending to, say, a narrative account of historical geology and its underlying reasoning. Procedural memory remembers motor and cognitive skills, which are essential for such tasks as driving a car or surveying in the field. Semantic memory is concerned with background understanding of what one considers relevant, true and significant, such as aspects of an appropriate paradigm, which a computer system might support with a framework of ontologies.

Operational procedures, controlled by feedback, move information through the boxes of Figure 2, which represent various stages of organizing, expressing and sharing information. Each box shows the same five distinct information types, which are represented and manipulated differently, in parallel with patterns of human thought. The process of surveying initially extends the surveyor's background knowledge (box 1). Shareable information can subsequently be extracted (box 2), communicated to colleagues, and recorded as conventional records. In conventional publications (box 3), information types are either inextricably combined within a document (as in a book with map illustrations), or entirely split apart in separate documents (such as a map and report). A digital, potentially Grid-based, system (box 4) with hypermedia links and mark-up languages will have greater flexibility in relating information items, of any type, scale and extent, from any source. It must work alongside conventional records, but will contain information derived directly from surveying as well as that derived from conventional records.

Each information type must be identifiable, for example, in digital files by their suffix (as in HTML) or from metadata, because each information type is manipulated differently. Thus the user of a database management system can select data by specifying ranges of values of relevant properties, and can analyze them by statistical methods. The user of a geographic information system can overlay maps and images, identify and compare patterns, shapes and sizes of objects represented by color or ornament, pan around to see adjacent areas, zoom in or out for detail or overview – extended in spatial models by the ability to apply three-dimensional geometrical transformations, and to select a volume of interest and the most informative visualization. Text can be searched in episodic memory, with hypertext links that enable the user to follow threads of thought through the main story, and weave them together in the mind as a coherent narrative. Actual

demonstration, video clips, or illustrated manuals, taking the user step by step through procedures in the real world with appropriate commentary, can help in the development of procedural skills. Training, textbooks, education and experience develop the semantic memory, which in broad terms is related to the shared paradigm and ontologies, and in detail is specific to the background knowledge of each user.

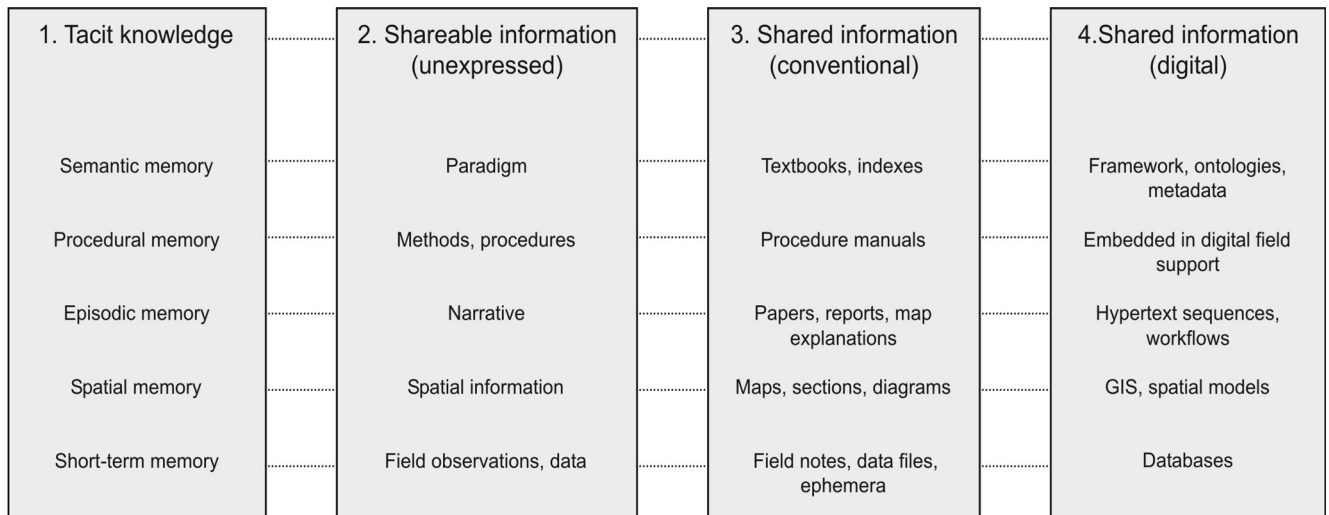| 1. Tacit knowledge	2. Shareable information (unexpressed)	3. Shared information (conventional)	4. Shared information (digital)
Semantic memory	Paradigm	Textbooks, indexes	Framework, ontologies, metadata
Procedural memory	Methods, procedures	Procedure manuals	Embedded in digital field support
Episodic memory	Narrative	Papers, reports, map explanations	Hypertext sequences, workflows
Spatial memory	Spatial information	Maps, sections, diagrams	GIS, spatial models
Short-term memory	Field observations, data	Field notes, data files, ephemera	Databases

Figure 2: Information types. The boxes show stages through which information passes, driven and controlled by operational and feedback processes. Each box refers to the same five information types, each of which is represented and manipulated differently.

The ease of linking items of information in a Grid-based system means that information types can be separated for processing purposes, and linked to provide information on various aspects of the same object or the same theme. For example, the system might select items of different information types referring to the same object, and represent them in linked frames side by side on the screen, where users can manipulate each in the appropriate mode, and reuse them in any appropriate context. Thus, selection of an object on a map, such as a set of boreholes, could call up relevant material held in databases, imagery, and narrative text (or vice versa). The human ability of weaving understanding from threads of thought involving different information types can thus be fully supported. As mentioned earlier in this paper, the geological content and forms of presentation vary for different users and applications, and the information records must be explicitly designed as components that are re-usable in different applications. The framework must enable all information types to work together at all scales.

3.4 Wider horizons

The scientific process strongly encourages a shared view of the world. Indeed, a primary purpose of science is to relate a myriad of observations to a few scientific laws. Explanation is the means of bringing together the consequences of numerous concepts and results, while standards contribute to a shared frame of reference in which ideas are more readily exchanged. Many geological processes of most concern to mankind involve interactions of the lithosphere with the atmosphere, hydrosphere and biosphere, and therefore refer to Earth systems as a whole.

The National Research Council (US) (1993) set out proposals on 'Solid-Earth Sciences and Society' developed through wide consultation by 150 earth scientists over a five-year period. Their

influential conclusions took the view that the study of the whole-earth system provides an essential research framework for addressing global problems, interweaving many branches of pure and applied earth sciences. The study of Earth systems science became a driving concept for some key international scientific programs, and major universities began the long process of revising their curricula in the light of these proposals. There is therefore a move to study and teach geology as one specialized aspect of Earth systems science: "the seed of a new and revolutionary unification of the science of our planet, how it works, its past history and its likely future" (Cornell University, 2007). The next generation of users will have changed expectations of the role of surveys in Earth systems science. Graduates trained in this holistic approach are now potential staff in geological surveys and customers for survey information.

The diversity of users of geological survey information and their disciplines is apparent from listing just a few application areas, such as: resource estimation; mineral and energy extraction; civil engineering construction; land-use planning; agriculture; nuclear waste disposal; carbon sequestration; evaluation of threats from coastal erosion, landslips, earthquake and volcanic activity, and flooding; explaining past and present climate change; and studying environmental influences on evolution and extinction of life forms. In the same way as geological survey agencies are moving from a map-based approach to one based on spatial models and a systems view, applications are undergoing parallel developments within the diverse organizations and disciplines that use their output. Geological information must be considered in the context of a knowledge system with a unified design that matches the unification of the science.

A group of thirty-four leading scientists from the life, earth and computing sciences contributed to the report, 'Towards 2020 science' (Emmott, 2006). They concluded that computing would not merely help scientists with their work. Rather, the concepts, tools and theorems of computer science will become integrated into the fabric of science itself, providing an orderly, formal framework and exploratory apparatus for other sciences, thus helping to break down barriers between disciplines. The geoscience paradigm does not exist in isolation but is part of the wider paradigm for the Earth and life sciences. This suggests that a coordinated strategy is desirable to develop shared standards, at least at appropriate levels of framework, metadata and ontology, throughout the fields of life and Earth sciences and beyond. The alternative, of a piecemeal approach, could allow the concrete to set on needlessly diverse structures for individual disciplines. Geoscience participation is therefore timely. The global ambitions of future information systems will require extensions to an explicit framework for geoscience information and the use of more comprehensive and widely shared ontologies.

For some future external users of geological information, the geoscience paradigm (implied in conventional publication) may be unfamiliar or inappropriate. The framework should therefore provide an explicit representation of the structure of the underlying concepts and links to ontologies. Procedures of geological surveying must be reviewed in the context of multidisciplinary investigations, and develop interoperable models that enable workers in different disciplines to work together to integrate their knowledge. For example, the fixed-scale view of geological maps limits interoperability by constraining their interpretation in scale-space (Carey, 1962, Hay et al., 2002). The design of the framework should enable geologists to study processes, and record and share interpretations, at all scales. Such issues may require a review of the role of stratigraphic classifications in the wider context of Earth systems science.

The global study of Earth systems has tended to focus on the atmosphere and oceans rather than the solid Earth, and on geophysics rather than geology. Examples (a search engine should readily locate current web-sites) are: Earth System Modeling Framework, International Geosphere-Biosphere Programme, NASA's Earth Science Roadmaps, Semantic Web for Earth and Environmental Terminology, Program for Integrated Earth System Modelling, Solid Earth and Environment Grid,

Earth System Curator, Grid ENabled Integrated Earth system model, and the Electronic Geophysical Year. Geologists must build on existing initiatives in these fields.

4 Towards a Grid-based system

4.1 Objectives

By far the greater part of shared geoscience information, as described earlier, remains within long-established systems that developed before digital computers and networks played a significant role. The future framework must work with and build on this legacy. The records of geological surveying result from a human activity, and rely on the skills of the human brain. Mechanical tools, from pen, paper and printing press to geophysical instruments and computer systems, can assist but cannot displace the human element, which pervades the entire system. It follows that the mechanical components of the system should work in harmony with human goals and thought processes, and that the function of recorded information is to support and enhance the background knowledge of the human user.

The conventional system has significant deficiencies. A surprisingly large part of most scientific papers is a reworking of earlier published material, recast to explain the author's viewpoint. The result is high redundancy, that is, there is much repetition of the same information. The mechanics and economics of printing and publication result in rigid, self-contained representations and a slow-moving and expensive system. The Web has accelerated parts of the process, but has largely retained the conventional information structure.

A Grid-based system will subdivide geoscience information into smaller, re-usable elements, reducing redundancy. Rapid recording, editing and delivery methods can overcome delays. The ease of linking re-usable items of any information type for different purposes makes it highly flexible. Computer techniques, such as database, GIS (geographic information systems), and document processing, provide appropriate techniques for manipulating tabular, spatial and narrative information types, while hypertext linking can relate different information types referring to the same object. Object-oriented analysis provides a context in which individual things or entities of geological interest can be represented in the computer as object instances, and classified in hierarchies of object classes, inheriting properties as appropriate. Computer processing can handle routine procedures, such as selection, analysis, interpolation, simulation, and visualization of geological processes, replacing obscure rules of thumb with explicit definitions and justifications. All of these are seen as part of a system in harmony with human thought and conventional procedures.

The Grid can facilitate linkage of information from various sources and simplifies the users' view of the knowledge system by delegating decisions to middleware (software that connects between systems). De Roure (2007) states that: "Our vision is of a generically useable e-Research 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."

The user interface should not be restricted to the digital part of the knowledge system, but should improve access by users to all parts of the knowledge system, including indexes (and direct access in some cases) to the all-important conventional representations and unrecorded expert knowledge.

Ideally, providers of information services, including geological surveys, will ensure that diverse knowledge sources share a consistent framework, are visible to the same search engines, and are accessible through the same user gateways or Grid portals (such as GridSphere, 2007).

A systems framework within the context of a Grid-based system can offer a means of positioning and classifying items (thereby relating them to one another and to other work), defining them through ontologies, and finding them in distributed information stores by means of indexes. These developments support a change in the business models of geological survey agencies. The essential task remains that of maintaining information resources that provide a coherent and authoritative account of the geology of an area. But the emphasis changes – away from the geological map towards a solid Earth systems model; and away from publishing printed end products towards enabling users to select flexible services that respond to their specific needs.
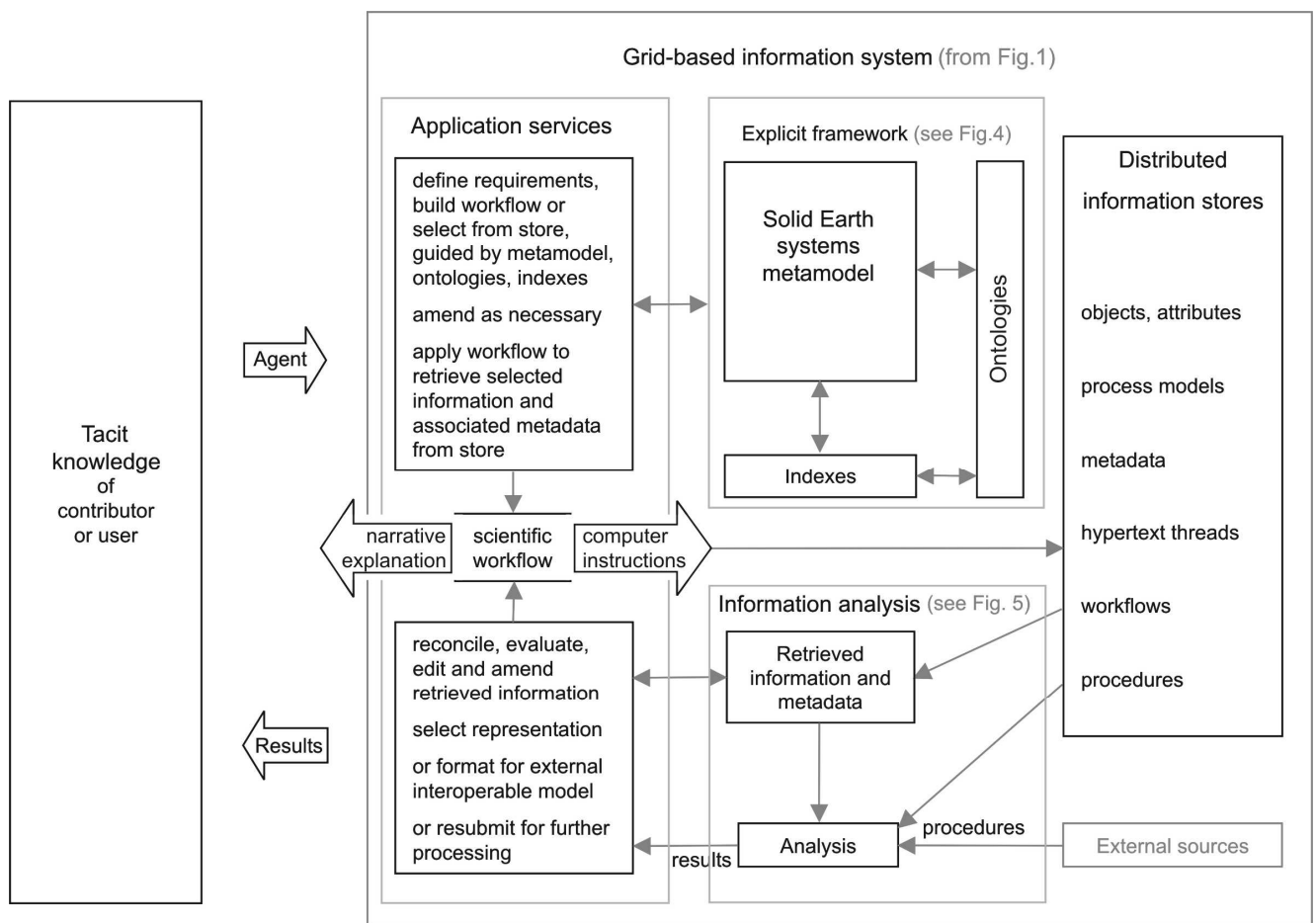

Figure 3: Grid-based information system (from Fig. 1), as explained in section 4.2.

4.2 Application services

The scenario illustrated in Fig. 1 suggests that access from a computer terminal can augment conventional procedures, and could be extended to a Grid-based system with an explicit framework. This is shown in more detail in Fig. 3. Users or contributors of information define their requirements from their basis of background tacit knowledge. An 'agent' (software acting on the user's behalf to perform a particular task) might assist users to select the services they require by

communicating with the application services in the Grid-based information system. It should enable users (those most likely to know their own requirements) to obtain appropriate products from existing services, or to find components to extend or build their own solutions.

A standard service-oriented architecture that is aligned with business processes is described in OMG (2007). The services might be represented as workflows, which bring together, like links in a chain, a sequence of computational components along with annotations that clarify the purpose and reasoning for the user's benefit. A formal workflow language, such as Kepler (2007), can model and describe the selection of information and control its flow through the procedures of computer analysis. Thus, given a request to provide, say, a lithostratigraphic map, the agent might consult with the user to determine the area, scale, detail and visualization technique required, assemble the appropriate information, and deliver it as a map representation to the printer selected by the user. The resulting 'scientific workflows' (Ludäscher et al., 2006) can potentially be built into a flexible archive for sharing and re-use or modification for specific needs.

The workflow archive can be seen as a Grid resource. It is a means of informing potential users of products obtainable from the geoscience knowledge system. However, the Grid will link to all areas of general and specific knowledge, and the longer-term objective is to participate in a global system where survey information can be fully integrated in its wider context, thus contributing to, and benefiting from, work in other fields. The geoscience system must therefore take into account the requirements of potential users outside geology. Consider, for example, a civil engineer assessing foundations for a building, or an epidemiologist studying the link between trace elements and regional variation in health patterns. Ideally, the search engines they consult, or the agents sent from their desktops to search the Grid, should discover any appropriate geoscience information, whether or not the enquirers are aware of its existence or relevance, and regardless of whether the form of its representation is spatial (map), narrative (scientific paper), or tabular (database). Search engines with this ability to examine all information types appear to be under development, but the primary search is likely to be in text form. Authors of general applications might therefore link the computer instructions of a workflow to a narrative explanation (in free text) designed to make its relevance visible to potential users and to standard search engines (Fig. 3). Establishing relevance by means of a search engine is based on analysis of information content, and can lead the user's agent to the survey's application services, which require a more rigorous structure.

The agent that selects and edits a workflow is guided by interaction with the user on the one hand, and the explicit framework (Fig. 3) on the other. The role of the framework is to develop standards for geologists to agree on the definitions and structure of their knowledge, and make that structure explicit and machine-processable. Its role within the Grid-based information system is somewhat analogous to that of the geoscience paradigm within a geologist's brain. Its three components, now considered in turn, are the solid Earth systems metamodel, the associated ontologies, and indexes.

4.3 The solid Earth systems metamodel

The solid Earth systems metamodel is intended to represent a coherent overall structure outlining how geoscientists relate, organize, store and locate their shared ideas. As described in section 5, actual computer implementations refer to specific aspects of geoscience and require a much more detailed and formalized approach. The structure might be compared to the geographical map referencing system of latitude, longitude and elevation that defines conventions for referencing points in geographical space, enabling the user to indicate a route across a map by a sequence of point coordinates, or to define an irregular area, or a volume in three-dimensional space. The metamodel obviously refers to many more dimensions than a conventional map, including the countless dimensions of state-space, as well as stratigraphic time, and scale-space. As with any

other referencing system, the metamodel should provide a means of defining any valid point, but does not imply the existence of information at that point.

The solid Earth systems model refers to the three-dimensional disposition and configuration (where things are and how they are arranged) of the present-day observable objects of the solid Earth, and their observed and interpreted properties, composition and relationships. This is illustrated in Fig. 4 as a set of object instances, referring to object classes with specific ontologies and classifications that may or may not be part of a wider system of generic ontologies. The overall arrangement of the Earth components is depicted as a spatial configuration of the object instances, based on spatial relationships and inferred time relationships. Geological processes, which can also be classified and described and may involve a generic ontology, act on the configuration of objects. The objects, processes and relationships are shown in the inner box of Fig. 4.
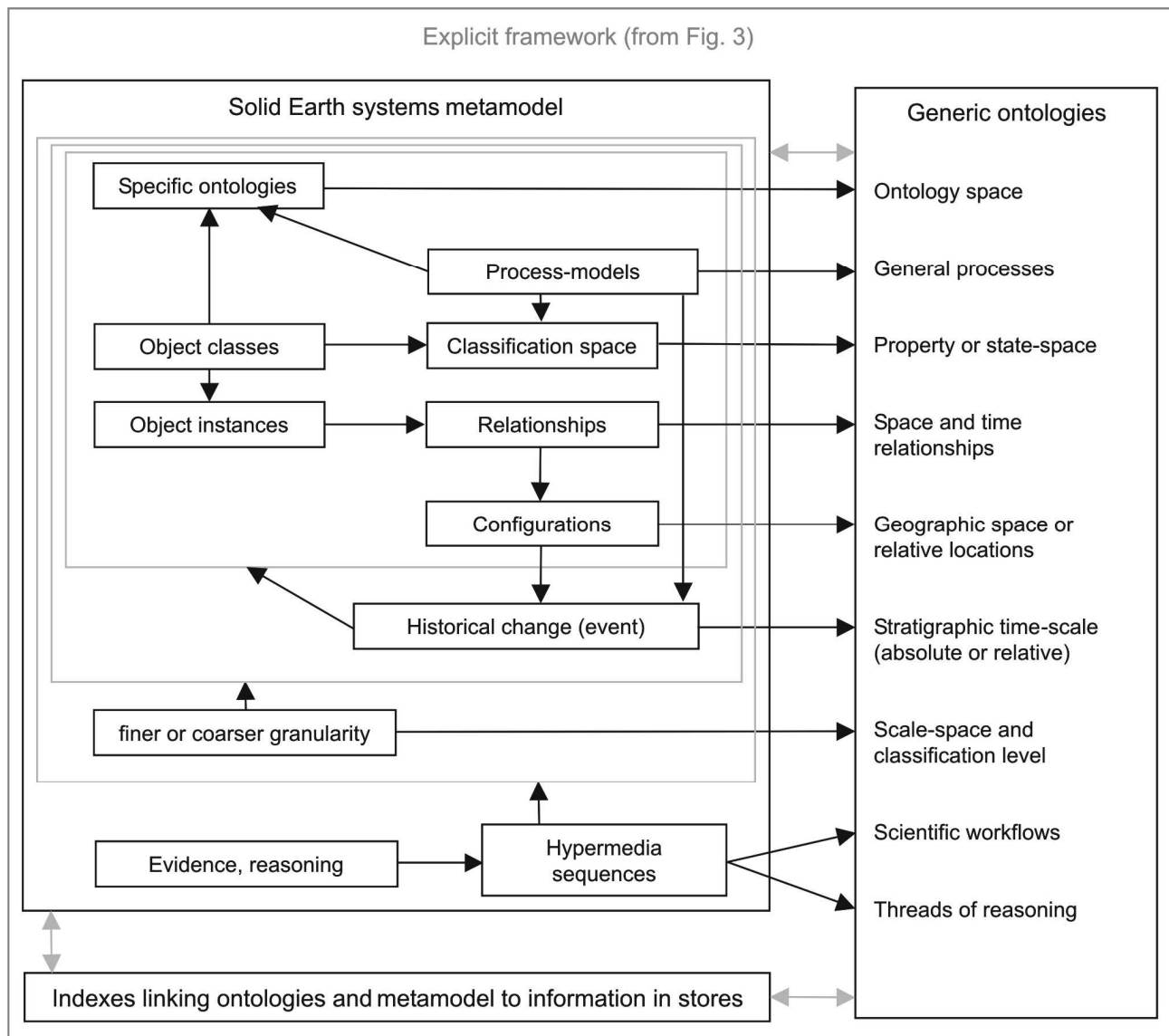

Figure 4: An explicit framework for the information system (from Fig. 3). The metamodel describes the contents of the solid Earth systems model, providing a means of referencing them, and bringing the components into a shared framework, which relates them to appropriate ontologies, indexes and hypermedia sequences.

The model also refers, however, to the events and historical changes throughout geological time, including the past properties and configurations of conceptual objects and the processes and events that created and altered them. These historical changes are arranged in terms of absolute or relative time scales and are seen as essential aspects of the interpretation. Conceptually, the inner (present-day) box has equivalents describing the historical configuration of objects and processes at every moment in the geological time-scale. It is therefore shown as enclosed in a second (history) box, referring in another dimension (not shown) to the entire sequence of configurations and the historical changes between them. To make matters more complicated, any of these aspects could be considered at any level of scale, detail or granularity. The present-day and history boxes are therefore embedded in a third (granularity) box, in yet another dimension (not shown) representing scale-space (3.4), where finer or coarser granularity may indicate change in resolution (the shortest distance apart of two points that can be discriminated on an image) or detail (as in changing from a narrower to a broader term or category).

Of course, the model in its totality has no concrete existence, for its size is vast beyond representation, and its full detail is beyond investigation. For good practical reasons, even the broad aims of a geological survey are limited to investigation of various facets and fragments of this general model. Conceptually, geological time and space are continuous, and potentially their quantitative representation could be precisely located at any point. In reality, the actual representation space is practically empty, with sparse information located at only a few loosely defined points. However, the top level of this model, the metamodel, could be part of a conceptual framework for the knowledge system specifying its contents, their organization, and the relationships among them.

This model could guide information searches, in which the user could select from the displayed parameters and ontologies. Indexes could relate that selection to relevant items in distributed information stores. Geological evidence and reasoning could be tracked through geographical space, scale-space and geological time as a hypermedia sequence during survey investigations, thus linking evidence and reasoning through this multidimensional structure. Users searching for information could likewise specify their areas of interest as multidimensional tracts within a defined distance of a particular path through the framework. Their specification can be represented as a workflow to retrieve appropriate information (Fig. 3).

The reference structure is a means of relating observations, interpretations and reasoning derived from various fragments and facets of incomplete knowledge of local aspects of the geology. The structure should bring them into a shared context, and identify their relationships within this wider view. It should support interoperability, linking a framework specific to geology to more generic external ontologies. It should identify basic aspects, such as spatial, stratigraphic and lithologic properties that can be widely shared, and encourage consistent description of them during abstraction, codification and reasoning (3.1). It should provide a hospitable and extensible setting for bringing general concepts, such as scale-space and complex systems, into surveying. The framework must adapt as the science evolves, but the top levels should be relatively stable.

The metamodel must obviously be developed in collaboration with other Earth systems modeling initiatives (3.4), since many geological processes relate directly to the atmosphere and hydrosphere. It differs from the metadata describing the framework for data from the oceans and atmosphere, as in, for example, the Grid ENabled Integrated Earth system model (GENIE, 2007), and related activities. The solid Earth model places more emphasis (2.3) on a holistic, historical, and interpretative view. It relies on subjective interpretation based on the tacit knowledge of its originators and users, and therefore requires subjective reconciliation (2.2) to share fragments of knowledge gained from differing points of view.

An initial strategy for handling the complexity is to simplify by classification, defining discrete categories that refer to specified zones in the multidimensional space of the metamodel. On the basis of observations and background knowledge, surveyors in the field then assign object instances and relationships to these object classes or categories. Within the metamodel, object classes can be located within classification space, based on their property and state-space ontologies. The ontologies are not restricted to this metamodel but may be maintained externally in a wider context.

4.4 Ontologies

Ideas from ontology have entered geological surveys, and will contribute to the design of future geoscience knowledge systems – a design that must be directed by the needs, working practices, and input of geoscientists. The word 'ontology' refers in philosophy to the systematic study of existence. Within a computer system, what 'exists' is what is represented. The representation, for some specific purpose, of an abstract, simplified view of the world, is referred to as a conceptualization. In computer science, an ontology is seen as a specification of a conceptualization (Gruber, 1993). As a bridge linking human knowledge and computer representation, it provides a controlled vocabulary to identify the things of interest (entities or objects), the processes or events that transform them, and their characteristics and relationships in a particular field of interest (knowledge domain).

According to Raskin (2006) "Ontologies are a form of controlled terminology that differ fundamentally from taxonomies, thesauri, and other controlled hierarchical or linear lists of domain terms commonly adopted by organizations. Ontologies enable child terms to inherit all the properties of their parents, rather than being subcategories of their parents. This fundamental assumption enables knowledge to be re-used and supports scalable knowledge construction, as it is not necessary to redefine higher-level concepts previously defined. Ontologies provide the mechanism for articulating how a child concept differs from its parent, using the ontology concepts themselves. Furthermore, ontologies support multiple inheritance, so that compound concepts are easily generated."

Ontologies can help to provide a consistent context for storing, discovering, selecting, retrieving, analyzing and sharing information as it moves from conventional to more flexible computer representations, providing a coherent framework for the bottom-up automation and extension of preexisting systems. Ideally, geological surveys would be committed to a shared ontology; in fact, as pointed out in 5.3, ontologies differ for historical and other reasons. Ontology mapping (Ludäscher et al., 2003) can help to automate the combination of data from several sources within and outside a survey to share information, integrate data and guide reconciliation where sources diverge.

Proposals for the Semantic Grid and interoperable models will help related disciplines to work together. According to Alper et al. (2006): "The Semantic Grid is a recent initiative to systematically expose semantically rich information associated with Grid resources to build more intelligent Grid services. The idea is to make structured semantic descriptions real and visible … with an associated identity and behaviour. We can then define mechanisms for their creation and management and protocols for their processing, exchange and customisation… The background knowledge and vocabulary of a domain can be captured in ontologies – machine processable models of concepts, their interrelationships and their constraints… Metadata labels Grid resources and entities with concepts… Rules and classification-based automatic inference mechanisms generate new metadata based on logical reasoning".

Ontologies and workflows set in a suitable framework should promote re-usability of representations of: objects, their characteristics (properties and composition) and relationships; geological process models; and surveying procedures. The framework should encourage regional

and global consistency through evolving standards. There are huge potential benefits in sharing concepts, models and computer representations among all appropriate disciplines as interoperable elements, that is, items defined and recorded in such a way that they can be interpreted and processed together. However, items of information can be interoperable between disciplines only if the larger context of viewpoints, frames of reference and methods of investigation can be reconciled, and even within a single discipline, problems may arise.

Based on philosophical analysis, Brodaric and Gahegan (2006, p.2) identify "several challenges to geoscientific information interoperability, present an approach that addresses some of the challenges, and… demonstrate the approach." They propose technical solutions. They point to some possibly unique characteristics of geoscientific knowledge, although many apparent differences in specialized topics may simply reflect their divergent evolution. In the same volume, Richard (2006) proposes models for Earth material, geologic units and geologic structure as a starting point for a framework for developing interoperable systems. He places these in a setting of more general, top-level concepts (SUO WG, 2003).

The global aspect implies that ontologies must be expressed in terms that conform to Grid standards and can be widely understood. This implies that an explicit shared framework for geoscience knowledge must be linked as far as possible to a more general framework of ontologies (Fig. 4).

4.5 Generic ontologies and indexes

The multiplicity of viewpoints in local studies of geoscience might be reflected in the infrastructure by a variety of large and small ontologies, specific to the problem at hand. The ontologies provide an overview of the detailed information content. Working connections among them can be established through linkages, in a system that encourages their rationalization, and reconciliation for specific purposes (with inevitable loss of some information), in response to real user needs (Ludäscher et al., 2003).

Each instance of an object, relationship, configuration, process or event, refers to concepts that identify appropriate ontologies for its description (Fig. 4). The ontologies discriminate and classify the concepts; and the solid Earth systems model places the ontologies within the context in which they are viewed in geological survey. The extent of shared ontologies and compatible frameworks will determine the degree of interoperability in the information system.

The list of generic ontologies in Fig. 4 requires further explanation. The individual external ontologies should ideally be part of a comprehensive hierarchy of ontologies (*ontology space*) where the relationships between them were apparent at a higher level in the hierarchy (SUO WG, 2003). The solid Earth systems model is part of a broader framework and is related to, and interacts with, other models referring, for example, to the hydrosphere or atmosphere. Thus, geological definitions that refer to physical, chemical and biological terms and processes might be defined externally, within their primary field. Many involve processes that can be described in terms of the *general processes* of physics, chemistry or biology. Components of the models are likely to be re-usable in other subjects, and again could be identified and shared through an external ontology.

Geological classifications are likely to refer to a range of values of various properties within which an object class is expected to lie. It therefore refers to an area of *state-space*, a general concept representing the set of all possible states (defined by combinations of values of properties). State-space can be thought of in geometrical terms as a multi-dimensional space where each dimension represents a property, and each point represents a specific state.

Space relationships and *time relationships* may place items on a numerical scale. But in referring to events in the stratigraphical past where geographical coordinates are lacking, knowledge of space relationships may be limited to such terms as: adjoining, truncating, overlapping, above, below, in

front of, behind, between, and beyond. The relationships might be illustrated on a sketch with no fixed direction or scale, but could not be correctly positioned on a map with a geographical grid. Time relationships, particularly in detailed survey, may not be able to place events on a geochronological scale and be limited to concepts such as before, after, and during, probably deduced from spatial relationships. Mark-up languages supported by databases are a possible tool for recording and reasoning with the relationships, and are supported by software for searching documents, to identify references to time and space relationships in the text (Boguraev and Ando, 2005) and build appropriate indexes.

An appropriate mathematical framework for representing such relationships is a directed graph, as used in critical path analysis, which can position events according to either time relationships or absolute numerical values. Stratigraphic and space indexes derived from information in marked-up text and maps could be valuable geological tools. The need for viewing and recording geological processes and objects within *scale-space*, and the desirability of considering scale in linking them to models from other disciplines, was mentioned in section 3.4.

The *scientific workflows* described in section 4.2 can provide digital access (Fig. 3) to meet the user's processing requirements, by retrieving and processing re-usable components from information stores, for appropriate presentation and visualization at the user's desktop or to incorporate in other models (Ludäscher et al., 2006). Survey experts might prepare standard, off-the-shelf workflows, for example to print geological maps, which ensure that the results are a valid interpretation of the data. Surveyors can extend the options by recording data in three dimensions and reworking earlier maps with the help of remote imagery. The end user could thus select, combine and adjust re-usable workflows to derive various products from the same information base, customized for the appropriate class of user requirement. A longer-term objective might be for the user's agent to create workflows to meet specific requirements as defined and requisitioned by the customer.

Scientific workflows have the important potential to record the course of the surveying process (Alvarado et al., 2005) as it proceeds through the framework, tracking the process of observation, reasoning, and interpretation. This process is thus made explicit, with full attribution and provenance to identify the originator's specific viewpoint, linking the findings to other overlapping work, and embedding the activity of surveying within its wider geoscience context. The hypertext sequences of scientific workflows have potential value both for setting out procedures for providing users with information from shared information stores, and for recording the procedures of geological surveying to relate observations to interpretations. They are a convenient means of linking sequences of computer operations, providing explanatory commentary (which can make them visible to standard search engines) and interacting with the user where choices have to be made.

Libraries of re-usable workflow components need not be limited to geological applications. The *threads of reasoning* that are a means of recording the reasoning that underlies interpretations and explanations are also hypertext sequences. Their relevance could well extend beyond geological information, and they might therefore be viewed as a generic tool.

4.6 Constraints on object behavior

Because the Grid aims to hide complexity from the user, workflows should be able to connect to rules and properties constraining the behavior of object classes and instances, preferably expressed in Grid-wide standard form. Workflows might take the form of metadata attached to the individual item, or might be implied by class characteristics defined in the ontology. They should enable the middleware to link objects to appropriate procedures (Fig. 5) for tasks such as filtering, transforming, generalizing, analyzing, interpolating, visualizing geological objects, and simulating
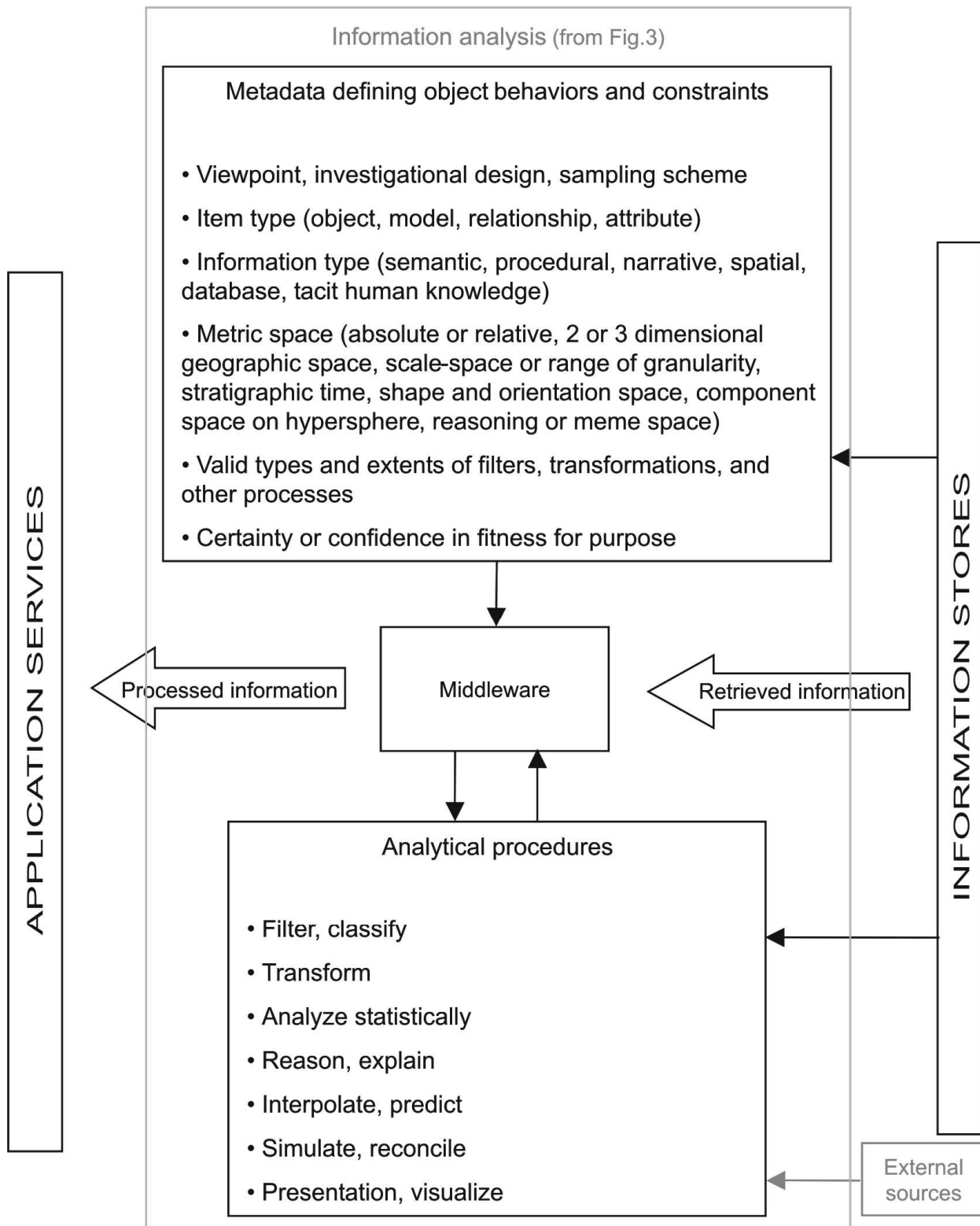
Figure 5: Applying constraints on object behavior. The retrieved information (Fig. 3) and the analytical procedures are obtained from internal information stores or external sources. The metadata define the acceptable modes of analysis of the information, enabling the middleware to match it to appropriate models or processes.

geological processes. Metadata should be available to indicate the behavior of the objects and their appropriateness for likely modes of analysis, represented in general terms that support concept and model interoperability across disciplines (Fig. 5).

The potentially wide range of users for the geoscience component of the knowledge system implies that it must develop in line with future, general-purpose, knowledge management systems. In the context of the Semantic Grid, the sources of information should be adequately described by metadata in a form that enables middleware to match the information to appropriate and valid procedures. For example, an agent searching for the deepest point reached by a formation in a particular basin might locate a set of well data. The metadata might indicate that for commercial reasons the wells were deliberately drilled on anticlines, and the middleware might reasonably conclude that an appropriate approach might be to simulate possible fold patterns based on statistical data for similar environments, fitted to the data but weighted for the known bias.

The middleware (Fig. 5) requires access to metadata clarifying the behavior of each object, and the resulting constraints on the appropriate models for its analysis. The metadata might be available at the level of the ontology in which the object is defined, or inherited from a higher-level object, or associated directly with an object class or instance. There is of course a hierarchical structure of objects, so that for example a configuration describing individual objects and their relationships might itself be regarded as a higher-level object. The provenance of the objects, including the objectives and investigational design of the project in which they were described (Fig. 1), is also relevant to their analysis. The corresponding metadata should be linked with the objects through the project workflow.

The object metadata, the models and the processes must be categorized and described to specify how the object behaves when a particular category of process model is applied. Thus they require ontologies, which, like the middleware, will inevitably have a major local component but should evolve to match general Grid standards, and should be described as far as possible in general multidisciplinary terms. This area is at an early stage of development, and in the short term, decisions about behaviors and constraints must continue to rely on the intervention of human experts.

5 Steps to implementation

5.1 Digital cartography

Traditionally geological maps have been the main means of communicating the results of geological survey. Geological map series produced by a particular geological survey conform to its standard symbology, with relevant items explained in the map keys. The key, as well as providing a description of the meaning of individual symbols, also commonly provides additional information, such as a stratigraphic hierarchy for lithostratigraphic symbols. These keys represent early geoscience ontologies. However, there are problems with using geological maps as the primary means of communicating geoscience information. Although geological maps from a single series aim at a standard symbology, in practice it can vary in detail from map to map, in part because completion of a map series can take many years. Each map therefore has its own ontology (Brodaric and Hastings, 2002).

A bigger communication problem arises between map series, particularly if produced by different geological surveys. Although geological maps follow similar conventions, there can be significant differences between the underlying ontologies, a fact that may be disguised by the apparent similarity of the symbolic conventions. This leads to a key property of geological maps – much

information on them is implicit rather than explicit. It is assumed that users of geological maps have in common a geological training and background that enables them to interpret the implicit information, but even with such a shared background there is much room for misinterpretation and misunderstanding.

Geological maps were digitized using computer-aided design (CAD) software from the 1980's onwards. The motivation was to speed up the production process for paper maps. The software was used within cartographic departments with little or no involvement of geologists, for whom the traditional geological map production workflow remained unchanged. The maps continued to use the symbology developed for the paper maps, albeit now realized through color tables and symbol libraries. The digital files produced by CAD software were not seen as an end product, but merely a step towards the production of the traditional paper map – and the output of such systems was largely judged by how closely they could mimic a map produced using traditional methods. However, these developments coincided with a requirement to derive various thematic maps from the standard geological map, and the digital files produced by CAD software could speed their production through the selection and merging of map features.

5.2 Spatial information systems

It soon became clear that selection of map components using CAD software had significant limitations, because it was dependent on map symbology. The ability to replace this limited capability with selection on the basis of geoscientific attributes was one of the main early drivers for the implementation of geological maps within geographical information systems (GIS). In carrying out such an implementation a decision had to be made as to whether the geological map should be modeled as a map, or as that subset of geological reality portrayed on the map. The principal distinction between the two approaches is that in the first case only those properties reflected in map symbology are implemented whereas in the second approach all available properties of the mapped features are implemented.

The first approach, apart from replacing CAD symbology with scientific terms familiar to geologists, also adds information by allowing the use of geoscientifically structured ontologies such as a rock classification system or stratigraphic lexicon. These allow more complex queries such as 'all igneous rocks' or 'pre-Permian rocks' to be made. Map symbols that reflect multiple properties, such as a line style representing an inferred normal fault, can also be broken down into their component properties, in this case 'fault type' and 'positional confidence'.

The second approach, implemented in GIS, shares these benefits but in addition allows the encoding of information implicit on the map, thus making it explicit. This approach can be seen as the first step to the creation of a more general geoscientific data model, independent of any particular medium for communication.

The implementation of geological maps within GIS systems, with an underpinning geoscientific data model, led to the creation of spatial databases in which map data could be integrated with point information, such as boreholes and samples, held in relational database tables. This encouraged the development of more comprehensive corporate data models embracing a wider range of geoscientific information. Such integration was further enabled with software developments that facilitate the handling of spatial data within relational databases, rather than in vendor-specific formats within GIS systems. These developments were used to underpin an increasing move to delivering geoscientific information in digital form, often tailored to the requirements of particular customers or end-users.

Increasingly geological surveys are creating computer representations (spatial models) that extend mappable data into three dimensions (Smith, 2005). These models may depict the same geological

objects as those shown on geological maps, extending them below the mapped surface, or they may represent some other geological property such as water in an aquifer. Because both the spatial models and maps portray the same real-world objects they can be described using the same data schema.

5.3 Data exchange

Providers of geoscientific information have developed, independently, their own data models. Although these data models all describe the same real world objects, such as faults or boreholes, the means of description differ. This is partly due to a different emphasis in the business models of different organizations. For example, organizations concerned with groundwater contamination might carry out a wide range of chemical tests on samples. It is also due to the fact that any mapping from the real world to a data model is necessarily arbitrary, because there is no single 'correct' answer. For example, rock fabric can be described either as an independent structure or as a property of a rock body.

The resulting differences in data models have meant that it has not been possible for customers to easily integrate geoscientific data provided by different suppliers. This problem was particularly acute in those countries with both federal and state or provincial geological surveys such as Australia, Canada, Germany, and the United States. In North America this led to the development of the North American Geologic Map Data Model (North American Geologic Map Data Model Steering Committee, 2004). The need for data exchange extended internationally, and was prompted by increasing customer demand for more standardized geoscientific information that would allow the development of standard software to process it. This led to an international web-based collaboration (CGI, 2007b) under the auspices of the International Union of Geological Sciences (IUGS) Commission for the Management and Application of Geoscience Information (CGI, 2007a) "to develop a conceptual model of geoscientific information drawing on existing data models" and to "implement an XML/GML encoding of the model subset" for data exchange. The scope of the model was set in the first instance as being geological maps and boreholes – the two types of data generally most in demand from geological surveys – but the aim is to extend this subsequently to more types of geoscientific information. The common data model is being developed in Unified Modeling Language (UML) and the current version can be seen at the CGI Data Model 'Twiki' web site (CGI, 2007c).

It is not envisaged that geoscience data providers should transform their internal data models to the agreed common data model. The cost of such a transformation, which would include rewriting an organization's software applications, would be prohibitive, but more significantly the common data model would be unlikely to meet each organization's business requirements. The objective is to enable data models of individual data providers to be mapped to the common data model for delivery. There is likely to be some loss of information in this process, but, over time, data providers can develop their own data models to conform more closely to the international model.

To exchange data derived from the common data model between organizations the model needs to be mapped onto a mark-up language. Mark-up languages retain the structure of the data model and are both machine readable and human readable. Upon receipt the marked-up file can be transformed back into a database implementation reflecting the common data model. The mark-up language being developed for geoscience is GeoSciML (Sen, M. & Duffy, T., 2005) and is based on GML (Geography Markup Language). GML is an XML (Extensible Markup Language) grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. One of the characteristics of GML is that it separates geometry from the description of features, which means that GeoSciML can be used equally well to transfer data from maps or from three-dimensional spatial models.

The simplest means of data exchange is to transfer the GeoSciML file directly. However, because GeoSciML is based on GML, it can also be delivered using the developing OGC Web Mapping Service (WMS) and Web Feature Service (WFS) standards. These services allow the data to be viewed and interrogated within a web browser, and importantly, also to be integrated with data from other sources. A test-bed is being developed to test the delivery and exchange of GeoSciML using WMS and WFS.

5.4 Ontology development

GeoSciML will enable interoperability at a technical level, so that different data providers will be describing the same geoscientific objects using the same properties, but there will not necessarily be scientific interoperability. For example, although the data model may specify 'Geologic Age' as a property of a 'Geologic Unit', if different stratigraphic classification systems are used by different data providers to define 'Geologic Age' then the resulting data will be incompatible even though it will still conform to GeoSciML. To achieve scientific interoperability the terms used to describe a particular property need to be the same, and this requires agreement on geoscientific vocabularies. Achieving this will be a major task due to the large volumes of legacy data described using data provider's own vocabularies.

The most realistic approach in the medium term is to compile and structure the concepts in existing vocabularies and use this structuring as the basis for developing a mapping between concepts in different vocabularies (Brodaric and Gahegan, 2006; Ludäscher et al. 2006). This mapping will not always be one-to-one, because different organizations may define broadly similar concepts in slightly different ways, but it will define the extent of overlap of concepts. There is more likelihood of concept mismatch, and more scope for misunderstanding, where concepts are defined in different languages. This process of concept mapping will provide a basis for multilingual systems that go beyond translation to a genuine exchange of meaning. Identifying areas of partial concept overlap is essential to achieving genuine interoperability between geoscience organizations. It will be even more important when using geoscience data in a future Grid environment in conjunction with other discipline domains between which there is likely to be a lower degree of concept overlap.

5.5 Drafting a roadmap

Considerable progress has been made in bringing geological maps (as a core product of geological surveys), and many associated documents and databases, into a systematic digital representation based on international standards. However, survey agencies are faced with larger tasks in adapting to changes in the cyberinfrastructure. These changes will affect their business plans and surveying methods, as well as the issues involving the systems framework discussed in this paper. Individual surveys will, of course, progress at different rates, and their priorities will differ. It may be desirable in each case to plan ahead for these changes, determining local priorities, the sequence in which tasks should be addressed, and the resources required. With some of these issues, the geological profession as a whole must be involved. As a possible starting point, Table 1 sets out some suggestions about aspects of the development related to the systems framework. Some are already well developed; others are potential extensions to the system. The various aspects can develop in parallel, providing increasingly detailed and well-described items of information. Ontologies should lead to greater standardization of the descriptions, and improve the connection of information of all types (data, map, model, and text) at all levels of detail, for the user to select, assemble, and edit as hypermedia information for analysis, presentation, visualization or incorporation in an external model. Future standards will increasingly be multidisciplinary. We hope that the list will provoke discussion on a road map for planning future developments in geological surveying, and help to identify gaps where more work is required.

Aspect	Actions	Outcome	Benefits
Database	Record quantitative and qualitative data	Digital datasets	Reuse dataset for any applicable analysis
	Describe content in metadata, preferably in standard form	Self-describing data	Usable by others, or at a later date
	Data modeling and standard (relational) structuring of data and metadata	Coherent database structure	Widely usable interconnected database, with user selection on basis of quantified and qualitative values
Digital maps	Scan maps	Digital representation	Overlay, pan, zoom; view locally, print on suitable printer
	Vector-digitize and label items	Itemized content	Items can be edited, updated
	Geographic information system and attribution of properties to objects	Spatial database	Spatial operations and complex linkages
	Data modeling, object-oriented analysis and markup	Item classification	Selectable content and representation, with standard markup for data exchange
Digital spatial model	Link to digital terrain model	Three dimensional representation	Geometrical transformations, visualization of three-dimensional configurations
	Link to spatial items from other sources	Access to additional data	Wider choice of content
Text processing	Scan documents	Digital representation	Locate and retrieve with search engine, view and print locally
	Capture word-processed text, Optical Character Recognition of scanned items	Itemized content	Edit, update items or sections within document; search within document for keywords; select format and layout
	Object-oriented analysis and markup	Item classification	Link content to other sources, re-use in various contexts, follow hypertext threads
Geological models	Simulate geological processes	Relate process and response	Explain observed data in terms of processes represented mathematically
	Simulate aspects of prediction	Superimposed scenario and prediction	Ability to, for example, visualize features of probable shape or finer granularity superimposed on an interpolated surface
Digital field support	Field recording	Digital field records	Records prepared and edited in the field or office in standard form
	Links to global positioning system, digital surveying tools, existing records	Networked field support	Wide range of contextual information available during the surveying process, with record of surveying workflow
Workflows	Prepare and store workflows to prepare routine products	Range of products from stores of re-usable objects	User can select and edit workflows to produce relevant products
	User prepares own workflow to meet specific needs	User-defined product	User's agent creates relevant user-specified product from internal and external sources
	Prepare documents describing workflow methods and results	Workflow documentation	Explains products to users; makes workflows visible to standard search engines
	Prepare workflow library for fieldwork support	Support to match surveyor's needs	Clearer field notes and images with less repetition, editable for archiving

Aspect	Actions	Outcome	Benefits
Metadata and ontologies	Describe and analyze information contents	Specific ontologies	Clarify structure and content of information to guide discovery, analysis and use
	Ontology mapping	Linked ontologies	Connect information from various sources, levels of detail and information types
Ontological framework	Insert semantic links	Links among items of all types to record underlying reasoning	Record of reasoning process; follow threads of original reasoning; explore alternative interpretations; identify knock-on effects of new data or interpretations
	Create metadata for objects and models, defining their behaviors and constraints	Generic metadata available for use by middleware	Assists middleware to match information with appropriate methods of analysis and display
	Separate observational records from interpolation and interpretation	Identify and relate evidence and reasoning	Spatial data can be evaluated, manipulated, or generalized, with specific needs in mind
	Codify the solid Earth systems metamodel (Fig. 4)	Representation of geoscience paradigm structure	Individual projects can link and relate to the same ontological framework
	Develop scale-space framework	Record of scale of processes and their products	Improve modeling; support complex and emergent systems, multiresolution deformable models, flexible generalization
	Develop time and space framework	Define geographic and stratigraphic relationships	Searchable absolute and relative stratigraphic ages and locations, relationships and configurations
	Develop state-space framework	Define properties of objects	Define and clarify the record of properties of objects and processes
	Develop classification-space framework	Define classes by their extent in state-space	Method of selecting object boundaries can be defined, including zero-crossings in state-space and scale-space (for interoperability)
	Agree standards for interoperability	Define standards for sharing information	Simplifies re-use of objects and attributes (characteristics and relationships), models, methods and concepts
Virtual communities	Establish long-term and transient communities	Ability to call on and use other opinions	Better decisions through sharing background knowledge, better communication through agreed standards
Evaluation	Estimate predictive power, relevance and consistency	Probability estimates available where appropriate	Relating judgments of individuals and groups and clarifying the implications in terms of Bayesian statistics

Table 1: Aspects of digital information relevant to a geological survey information base.

6 Conclusions

Scientific investigation and communication rely increasingly on their information technology infrastructure, taking advantage of its exponential growth in computing power, communication bandwidth and storage capacity. The Internet is seen as evolving to 'the Grid', supplying digital

information (representing fragments of global knowledge) as a commodity, just as the electricity grid delivers power. Concepts, tools and theorems of computer science are being woven into the fabric of science, providing it with an orderly, formal framework and exploratory apparatus that helps to break down barriers between disciplines. Many studies of geological events that most concern mankind involve interactions among the lithosphere, biosphere, hydrosphere and atmosphere, and can be unified within Earth systems science – the science of our planet, how it works, its history and its likely future.

In response to this changing environment, geological surveys are adapting their legacy of knowledge, including the geological maps and explanations that are its core expression worldwide. Many have recognized the value and technical feasibility of: generating sets of thematic maps by recombining map elements within a GIS-based cartographic system; supplementing lines on a map with three-dimensional digital representations; visualizing map data alongside other geospatial information; describing paths of linked activities with scientific workflows and thereby re-using items in various contexts; developing internationally standardized schemas; and linking diverse studies through ontologies (GEON, 2007). Object-oriented analysis has provided a structure for representing the digital information as discrete items that can be re-used in various contexts for a range of defined purposes. Some geological organizations are already weaving these strands together (Natural Resources Canada, 2007), edging away from publication of maps and memoirs as their mainstream product and moving towards contributing to a whole-earth knowledge system that relates processes and products, and enables users to obtain flexible responses to specific needs.

The change of approach does not alter the primary objective of a geological survey agency – to develop, record, maintain and communicate a reliable, authoritative, coherent, and up-to-date account of the geology of a region. The objective is achieved by integration of knowledge from relevant available sources, systematic field observation and measurement of salient geoscience properties, and their analysis and interpretation in the context of the rock types they characterize. The emphasis is generally on understanding the nature, distribution, history and configuration of the rock types, required as background information for a wide range of commercial, regulatory, and research activities, such as mitigation of natural hazards or assessment of natural resources.

However, global sharing of survey information across geographic, institutional and disciplinary boundaries involves much more than digitizing existing material. The transition calls for revision of the systems framework, in step with changing business models and reconsideration of the concepts and methods of surveying, to meet the expectations of a wider range of users and requirements. The framework must support and utilize reasoning networks and interpretations that take into account historical processes and object configurations. The originators and users of information must reconcile their views of the underlying concepts, classifications, procedures and understanding of processes, so that these may work together as interoperable components of a more inclusive system. Interconnected, re-usable items of various hypermedia information types can be assembled to meet a particular requirement, and processed side-by-side by the appropriate tools to weave a composite understanding in the tacit knowledge of the user's brain.

The system as a whole must build on what already exists, but it should enable geoscientists to augment their conventional representations and unexpressed knowledge with new approaches to surveying geology and sharing the results. It should take into account the facts that not all geologists have the same background understanding and not all agree on every interpretation. Mechanisms for consultation with experts are essential for communicating unexpressed knowledge, and must remain part of the knowledge system. Relevant data on the worldview, business model, project objectives, and investigational design and procedures describe the provenance and context in which information was collected, and may be essential to understand it. These data should be recorded and made available as metadata to assist in defining valid applications and reconciling

information from various sources. Constraints on object behavior could also be recorded as metadata to guide their analysis. Obscure rules of thumb that handle routine procedures, such as selection, visualization, analysis, interpolation, and simulation of geological processes, could be replaced with explicit definitions and justifications. Evaluation pervades the observation and recording of information, and artificial intelligence techniques can augment the essential human judgment.

The ability of geoscientists to understand one another depends on overlapping knowledge of parts of a general paradigm, of which the solid Earth systems model is a part. An explicit framework for a Grid-based system could therefore include the metamodel, or top-level description, of the solid Earth systems model. It could provide a shared multidimensional framework to which individual items of information can be referenced, a higher dimensional equivalent of referencing a point on a map with geographical coordinates. The framework could thus provide a means of relating objects whose similarity can be defined by parameters of geological significance, not just spatially but also in terms of stratigraphic age, configuration, environment and processes of formation, properties, and scale or granularity. The dimensions relate to ontologies, many of which are seen to be of more general interest, and should therefore be standardized over a wider field than geoscience. Indexes can relate items in distributed information stores with points or zones in the framework.

The metamodel and associated ontologies provide a means of describing a route through the model, such as that followed by the sequence of hypermedia operations in a scientific workflow. The route might refer to the path followed by a project investigation, or by a thread of reasoning, possibly essential for shared understanding. It may refer to the application of a processing procedure, or to retrieving information from topics of interest. The framework of metamodel, ontologies and indexes could be used by application services to provide an informative user interface, in which the user indicates topics of interest, and the system provides a graphical display of the relevant available information, enabling users to progressively refine their requirements as they learn more. Such a framework might also guide reworking of legacy information for added value and benefit, as computer support and interdisciplinary cooperation lead to a deeper understanding of the solid Earth as a whole.

Surveys will continue to extend their proven, open standards, within a model-driven service-oriented architecture, aiming for interoperability and adoption of generic methods for describing object behaviors and constraints, and building compatibility with related disciplines and new methods. Consideration of an explicit systems framework for geoscience is timely, initially implemented as local experiments, potentially in a standard form to aid global communication.

Acknowledgments

References cited

(Web references checked July 2007)

Alper, P., Corcho, O., Kotsiopoulos, I., Missier, P., Bechhofer, S., Kuo, D., and Goble, C., 2006, S-OGSA as a reference architecture for OntoGrid and for the Semantic Grid: Paper presented at the Global Grid Forum (GGF16) in Athens, Greece, February 2006.
http://www.semanticgrid.org/GGF/ggf16/papers/OntoGrid-GGF16-SemGrid-Wrkshp.pdf

Alvarado, M., Bañares-Alcántara, R., and Trujillo, A., 2005, Improving the organisational memory by recording decision making, rationale and team configuration: Journal of Petroleum Science and Engineering v. 47, p.71-88.

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

Boguraev, B., and Ando, R.K., 2005, TimeML-compliant text analysis for temporal reasoning. http://www.ijcai.org/papers/1202.pdf

Brodaric, B., and Hastings, J. 2002, An object model for geologic map information, *in* Richardson, D., and van Oosterom, P. eds., Advances in spatial data handling: Tenth International Symposium on Spatial Data Handling: New York, Springer-Verlag, p. 55-68.

Brodaric, B., and Gahegan, M., 2006, Representing geoscientific knowledge in infrastructure: Some challenges, approaches, and implementations: *in* Sinha, A.K., ed., Geoinformatics: Data to knowledge: Geological Society of America Special Paper 397, p. 1-20.

Carey, S.W., 1962, Scale of geotectonic phenomena: Journal of the Geological Society of India, v. 3, p. 97-105.

CGI, 2007a, Commission for the management and application of geoscientific information: http://www.bgs.ac.uk/cgi_web/

CGI, 2007b, CGI data model collaboration: https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/WebHome

CGI, 2007c, GeoSciML V2 candidate release schema: https://www.seegrid.csiro.au/twiki/bin/viewfile/CGIModel/HarmonizedModel?rev=1;filename=V2SchemaAll051907.pdf

Cornell University, 2007, The science of Earth systems: The evolution of a new scientific approach: http://www.geo.cornell.edu/eas/education/new_ses/newses.html

De Roure, D., 2007, The Semantic Grid: vision: http://www.semanticgrid.org/vision.html

Emmott, S., ed., 2006, Towards 2020 science: Cambridge, UK, Microsoft Research, 82 p, http://research.microsoft.com/towards2020science/downloads.htm and http://www.nature.com/nature/focus/futurecomputing/index.html

Foster, L., and Kesselman, C., eds., 2003, The Grid. Blueprint for a new computing infrastructure, 2nd edition: San Francisco, Morgan Kaufmann, 748 p.

GENIE, 2007, Grid enabled integrated Earth systems model: http://www.genie.ac.uk/index.htm

GEON, 2007, The Geosciences Network: http://www.geongrid.org/about.html

GridSphere, 2007, GridSphere Portal Framework: http://www.gridsphere.org/gridsphere/gridsphere/guest/home/r/

Gruber, T.R., 1993, A translation approach to portable ontologies: Knowledge Acquisition, v. 5, no. 2, p.199-220, doi: 10.1006/knac.1993.1008.

Harrison, J.M., 1963, Nature and significance of geological maps, *in* Albritton, C.C., ed., 1963, The fabric of geology: Stanford, Freeman, Cooper & Company, p. 225-232.

Hay, G.J., Dubé, P., Bouchard, A., and Marceau, D.J., 2002, A scale-space primer for exploring and quantifying complex landscapes: Ecological Modelling, v. 153, no. 1-2, p. 27-49, http://www.geog.umontreal.ca/gc/PDFs/New_PDFs/Hay_2002_EcoMod_Full.pdf

International Commission on Stratigraphy, 2007: http://www.stratigraphy.org/

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

Kepler, 2007, Kepler Project: http://kepler-project.org

Kuhn, T.S., 1962, The structure of scientific revolutions: Chicago, The University of Chicago Press, 172 p.

Loudon, T.V., 2000, Geoscience after IT: A view of the present and future impact of information technology on geoscience: Oxford, Elsevier, 140 p. (Also available as 2000, Computers & Geosciences, v. 26, p. A1-A142.)

Ludäscher, B., Liu, K., Brodaric, B., and Baru, C., 2003, GEON: Toward a cyberinfrastructure for the geosciences – A prototype for geologic map integration via domain ontologies: U.S. Geological Survey Open-File Report 03-471, http://pubs.usgs.gov/of/2003/of03-471/ludascher/

Ludäscher, B., Liu, K., Bowers, S., Jaeger-Frank, E., Brodaric, B., and Baru, C., 2006, Managing scientific data: From data integration to scientific workflows, *in* Sinha, A.K. ed., Geoinformatics: Data to knowledge: Geological Society of America Special Paper 397, p. 109-129, http://users.sdsc.edu/~ludaesch/Paper/gsa-sms.pdf

Minsky, M., 1981, A framework for representing knowledge. Reprinted, pp. 95-128, *in* Haugeland, J., ed., Mind Design: Cambridge, Massachusetts Institute of Technology Press, 368 p.

Natural Resources Canada, 2007, Consolidating Canada's geoscience knowledge: http://ess.nrcan.gc.ca/2002_2006/ccgk/index_e.php

National Research Council (U.S.), 1993, Solid-Earth Sciences and Society: Washington, DC, National Academy of Sciences, 346 p.

Nicolis, G., Prigogine, I., 1989, Exploring uncertainty: New York, W.H. Freeman, 313 p.

North American Geologic Map Data Model Steering Committee, 2004, NADM Conceptual Model 1.0, A conceptual model for geologic map information: Geological Survey of Canada, Open File 4737, 1 CD-ROM; U.S. Geological Survey Open-File Report 2004-1334, 58 p., http://pubs.usgs.gov/of/2004/1334 , http://nadm-geo.org/ , http://geopub.nrcan.gc.ca/moreinfo_e.php?id=215903

OMG, 2007, The Object Management Group and Service Oriented Architecture: http://www.omg.org/attachments/pdf/OMG-and-the-SOA.pdf

Pinker, S., 1997, How the mind works: New York, Norton, 660 p.

Rafols, E. J., Ring, M. B., Sutton, R. S., and Tanner, B., 2005, Using predictive representations to improve generalization in reinforcement learning: Proceedings of the 2005 International Joint Conference on Artificial Intelligence, p. 835-840, http://www.cs.ualberta.ca/%7Esutton/papers/RRST-ijcai05.pdf

Raskin, R., 2006, Development of ontologies for earth system science, *in* Sinha, A.K., ed., Geoinformatics: Data to knowledge: Geological Society of America Special Paper 397, p. 195-199.

Richard, S. M., 2006, Geoscience concept models, *in* Sinha, A.K., ed., Geoinformatics: Data to knowledge: Geological Society of America Special Paper 397, p. 81-107.

Sen, M., and Duffy, T., 2005, GeoSciML: Development of a generic geoscience markup language. Computers and Geosciences, v. 31, no. 9, p 1095-1103, doi: 10.1016/j.cageo.2004.12.003.

Simpson, G.G., 1963, Historical science, *in* Albritton, C.C., ed., The fabric of geology: Stanford, Freeman, Cooper & Company, p. 24- 48.

Skaalid, B., 1999, Gestalt principles of perception. College of Education, University of Saskatchewan: http://www.usask.ca/education/coursework/skaalid/theory/gestalt/gestalt.htm

Smith, I.F., ed., 2005, Digital geoscience spatial model project, final report: British Geological Survey Occasional Publication, no.9, 56 pp.

SUO WG, 2003. Standard Upper Ontology Working Group (SUO WG) Home Page. IEEE P1600.1: http://suo.ieee.org/

Sutton, R. S., and Barto, A.G., 1998, Reinforcement Learning: An Introduction: Cambridge, Massachusetts Institute of Technology Press, http://www.cs.ualberta.ca/~sutton/book/ebook/the-book.html

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. Search and Discovery Article #40080, http://www.searchanddiscovery.com/

Walsham, G., 2005, Knowledge management systems: Representation and communication in context. Systems, signs & actions. v. 1 no. 1, p. 6-18, http://www.sysiac.org/uploads/1-1-Walsham.pdf