<<<Back to Table of Contents
On to Part K: Coping with changing ideas. Defining the user requirement for a future information system>>>

Geoscience after IT: Part J

Human requirements that shape the evolving geoscience information system

T. V. Loudon
British Geological Survey, West Mains Road, Edinburgh EH9 3LA, U.K.
e-mail: v.loudon@bgs.ac.uk

Postprint of article in Computers & Geosciences, 26 (3A) April 2000, pp. A87-A97

Abstract - The geoscience record is constrained by the limitations of human thought and of the technology for handling information. IT can lead us away from the tyranny of older technology, but to find the right path, we need to understand our own limitations. Language, images, data and mathematical models, are tools for expressing and recording our ideas. Backed by intuition, they enable us to think in various modes, to build knowledge from information and create models as artificial views of a real world. Markup languages may accommodate more flexible and better connected records, and the object-oriented approach may help to match IT more closely to our thought processes.

Key Words - Thought processes, human communication, geoscience markup language, object-oriented analysis.

1. Communication at the interface

1.1 Interwoven threads

A great deal is lost when we force our thoughts into the straitjacket of shared conventional records (part I, section 6). Imagine yourself leading a geological field excursion. You would certainly talk, producing strings of words - narrative descriptions, accounts of sequences of events, reasoning and explanations. You would weave the ideas together to tell a story, probably supplementing the narrative with gestures, pointing to features of interest and drawing diagrams, perhaps with a stick in the sand. You might look at detail with a hand lens, then stand back to see the wider picture. You might refer to recorded knowledge: “I will read you a brief account of the regional geology; you can see where we are on this map”. You might pass on tacit knowledge by demonstration: “look at the outcrop here and you will see what I mean”.

The spatial context of your observations and hypotheses gives them coherence, but the narrative is essential to tie the elements together. You might cope with leading a field excursion with a broken arm, but would have problems if you lost your voice.

Different parts of the brain focus on different types of information, but given the opportunity, they work together for a clearer view of the big picture (or the big story).

Our experiences may be single sequences of events, put in words as a narrative thread. Repetitive events, like seeing a similar sequence of beds again and again, merge into a single strand, with only exceptions (the fossiliferous bed) remembered separately. In our memories, the threads are woven together as a complex fabric. Our brains are constantly trying to make sense of our experience by drawing analogies, abstracting and summarizing. The result is ideas with form and structure. We have each built our own background understanding or world view that provides the frames (K 1.2) in semantic memory where we can accept and evaluate new ideas.

The ability to integrate information types, so important in the field, is hampered by the need to package recorded information in separate containers for narrative (books and reports), spatial information (maps), data (databases) and discourse (discussions). We therefore need to look more closely at what we are trying to do and how we might prefer to do it.

Three long-standing tools for sharing knowledge, as well as helping individuals to develop and organize their own thoughts, are language, image and demonstration. Two more recent tools are mathematical methods and computer software. These tools shape the way we think and what we think about, as well as the way we perceive the world. They correspond to different **information types** (text, spatial, tacit and structured information) and lead to different styles of thought, presentation and processing procedures. The earlier techniques evolved so long ago that we have lost sight of their origin, but in planning for new technology we must bear in mind their characteristics. The Encyclopedia Britannica (1973, *Language*) is as always helpful on such matters.

1.2 Language and narrative

Long ago and far away, our ancestors grunted and made noises, and then sequences of grunts, ascribed meaning to them, and evolved a living **language** - a sequence of tokens denoting objects, actions and agents. At a similarly primitive level, our ancestors looked around them, scratched maps on the sand, drew pictures on the cave walls. Spoken language could be used for discussion and command. It was a means of communicating to a group, for the speaker could broadcast the same message to all within earshot. Images were at first a secondary, more personal communication, directed to a select few, and requiring special skills even to draw crude sketches. But suppose early man could have broadcast not just to the ears but also to the eyes of his colleagues, with the ability to capture, record and communicate images in full detail to one and all. Methods of communication would surely have evolved differently. That power, denied to our forebears, is now available to us. Does it make a difference? Has an optimal system evolved naturally, or are we constrained by historical accidents?

Leatherdale (1974) pointed out that sixteenth-century philosophers assumed that, in experience, we always encounter well-defined or discrete “things”, and that they seemed to conceive of the business of language as the adroit matching of words to these separately given things of which they are the mark or sign. The more recent view is that language can be explained in terms of socially agreed understanding of words: not on any intrinsic connection between signs and the world, as that would imply absolute properties of language independent of human culture. Contexts need to be invented, and stories created, to make a character string meaningful (Laszlo, 1972).

In evolutionary terms, episodic memory (I 4) presumably developed as a means of learning from one's own experiences - a single thread of events winding through the continuum of space along the arrow of time. Language matches this pattern of thought with linear strings of words and sentences, referring to past, present and future. Narrative skills evolved to create **stories** as surrogates for experience, told, retold and remembered (inaccurately). The story that began as a surrogate experience may by constant reworking acquire the mythic quality of a folk tale. It **generalizes** by pulling out crucial, illuminating events, implying much more than it states by reacting with existing ideas in the listeners' minds and influencing their semantic memories.

From speech, written language evolved to reach a multitude of users separated in space and time. **External representation** of knowledge (outside the human mind), for example by writing and drawing, goes far beyond the here-and-now of story telling. Scientists, in building their knowledge base, are not limited to their own episodic memory. They can contribute to and access a vast repository of information. They can do so at a time and place of their choosing, examining a wide range of specialized sources, past and present, in summary and in detail.

Faced with the infinite complexity of the real world, a complete description of it is unthinkable. Instead, stories are told in innumerable ways, to illustrate a multitude of experiences from personal viewpoints. Yet, since they deal with the same reality, all refer in a sense to the same underlying story. Because the overall story is large and complex, scientists must specialize in subdisciplines. A major part of story telling, or scientific writing, is therefore devoted to linking to earlier accounts, establishing common ground, clarifying, reconciling inconsistencies, and noting new information, discrepancies and ways to resolve them. **Discourse**, that is, the expression and interchange of ideas, is the means of clarifying and reconciling the accounts, by discussion within a workgroup, or through the slow process of conventional publication.

1.3 Spatial concepts

We can describe spatial pattern and relationships, rather inadequately, in ordinary language such as: "filling broad channels from a northerly direction which were later buried to a great depth." However, evolution provided us with specific abilities to handle and memorize spatial data (McCrone, 1999). Spatial skills bring evolutionary advantages in catching things, or not bumping into them, as in swinging from tree to tree, where (with luck) the main sensory input is by eye. We can communicate exact and accurate spatial information as **images** (representations of the semblance or likeness of an object), such as diagrams, sketches, maps, photographs and video clips, which make use of these skills. Note, however, that losing ambiguity is not always helpful. If we depict the channels described at the start of this paragraph on a map, they either join up in one direction or the other, or both, or not at all. This might bring unwanted connotations of tributaries, deltas, braided streams or whatever. The ambiguous statement may reflect genuine ignorance. The graphical representation can force an appearance of certainty that does not reflect the real situation.

Unlike narrative that places events in sequences of single threads, spatial thinking lets us build extensive structures, such as geological maps and cross-sections, that give a comprehensive view over space and maybe geologic time. We can zoom in to see the detail, zoom out to see the spatial context, pan around to see the situation elsewhere, and compare spatial patterns arising from different topics. Narrative text cannot offer these abilities, but can be intimately linked to a spatial representation.

We can represent spatial forms by combinations of geometrical objects, such as points, lines, areas, surfaces, fields and volumes, with well-defined mathematical properties. They can therefore be handled precisely on the computer. Within this rigorous framework, we can assemble spatial objects drawn from a wide range of topics, say, topography, borehole records, formation boundaries, lithologies, engineering geology, planning zones and proposed construction sites, and process them together. We can visualize the location, spatial patterns and relationships of sets of objects using interactive computer displays that take advantage of our spatial skills and the accuracy of human short-term memory. Our visual systems have evolved to process moving images, helping gannets to catch fish and motorists not to collide. Computers can exploit this talent, helping us to visualize changing spatial patterns, like the development of a sedimentary basin.

1.4 Structured data

Much more recently than language and images, another type of knowledge representation was invented, namely, numerical measurement and quantitative modeling (F 1). An advantage of numbers is their ability to define relative magnitudes as precisely as required. The eye is adept at comparing two magnitudes, for instance the relative sizes of two fossil specimens seen side by side. It is much less skilled at comparing objects seen in different places or at different times. **Measurement** against a standard scale uses the numerical series, the most exact order we can form. It provides the portable yardstick that makes it possible to assemble any number of side-by-side comparisons as well as comparing the magnitudes as a group and examining subtle variations in space or time (also measured numerically).

Mathematical models (F 5) build on this ability of numbers to represent magnitude, and on the similarities between mathematical operations and physical processes. Sets of measurements and relationships between them can be summarized, properties can be sampled to represent the underlying situation, uncertainty can be measured, states can be compared and processes simulated. Practical applications in geoscience, including the quantitative description of spatial data and analysis of images, depend largely on computer support, and a computer program is a precise and convenient means of sharing the model.

Mathematical and spatial models (along with the data collected to investigate them) are both analogies (J 2.3) between the real world and the properties of numbers. Quantitative properties and relationships can be visualized as patterns in space. The map can also be seen as a means of visualization (MacEachern, 1998). We can thus bring quantitative and cartographic methods into a single numeric framework. Both analogies are at best fuzzy approximations to the real world, despite the exact mathematics of the internal reasoning.

The standards, rules and conventions of the more highly structured areas of the information repository add value by creating a coherent and easily accessible database, possibly derived from many independent sources. All are ultimately set within text narratives, which explain the objectives, the conventions, the reasoning and the conclusions. Detailed narrative threads may refer to quite specific spatial items or aspects of spatial and quantitative models. The reader should be able to view the relevant items highlighted against their spatial context, and be able to move freely between the narrative and the visualization.

1.5 Tacit knowledge

Perhaps the greatest amount of geological information is held, not in the written record, but in total in the minds of all geologists. The geologist who has surveyed an area develops a mental picture of the geology more complete than that shown on maps and described in papers. Much is **tacit knowledge** which is acquired through practice and cannot be articulated explicitly (Kuhn, 1962) - known but not expressed. For example, you might instantly recognize a specimen which you could not identify from the most exhaustive description, just as you learned to ride a bicycle by demonstration (transferring knowledge held in procedural memory), not by written instruction. In a discussion, or a field excursion, much can be learned that could never be written down. The importance of tacit knowledge means that education, training and learning throughout a scientist's career need demonstration, discussion and directed experience - to communicate what we cannot express.

A reminder can recall forgotten memories. Hence the menus on a computer screen. *Recognition* of a command is easier than *remembering* the exact wording of a computer instruction. A valuable feature of an information system may therefore be to “jog the memory”, to present cues and analogies that can stimulate ideas in the pool of tacit knowledge. A second valuable feature can be the ability to access the tacit knowledge of others through discussion and inquiry. A third feature of the system could be the use of images and video demonstrations to illustrate, for example, the procedures used in collecting samples, or the precise points examined on an outcrop. These could clarify a narrative account, and would help others to repeat and test the observations.

1.6 Knowledge-based and rules-based investigation

Examples presented to students (I 3) are not typical of the procedures of experienced geologists surveying an unknown area. During an initial survey, a comprehensive set of observations is likely to be made and recorded, if only to avoid the cost of revisiting each outcrop. For a graduate research project, a local, self-contained problem might be sought, preferably with significance in a wider context. The abundance of such problems in geoscience makes it an attractive subject to study. On the other hand, the search for oil is more likely to employ techniques that are well established on a global scale, in tune with the uniform business objectives. The fact that the model is well defined before the investigation begins (being based on a clear user requirement) increases the scope for rules-based activity and so for automation.

This points to a distinction between exploratory investigation and systematic pre-planned investigation. The first is **knowledge-based**, starting from some preconception of the geological setting and developing and extending the explanation with each new observation. Evidence is constantly sought, by new observations or reworking data, to confirm or modify the current interpretation and choose the next step of the investigation in the context of growing background knowledge. A narrative account is built in episodic memory. The second is **rules-based**, where the pattern of the investigation is decided before work starts. In contrast to the exploratory search of the knowledge-based project, the rules-based project is analytical, studying known relationships by collecting and studying appropriate data, following a well-defined model. Standard procedures are specified, and instructions set out for following them. The resulting data are therefore consistent, and can be compared with one another and with data from other projects that followed similar rules. Short-term memories are recorded, and can be accurately recreated from the database.

Rules-based projects can be well suited to quantitative measurement and extensive instrumentation. The seismic investigations of the North Sea, and downhole logs from the subsequent drilling, provide examples of data collected by instruments according to pre-defined rules. Their consistency makes them particularly helpful in revealing regional pattern. A rigid, pre-determined structure can also extend the reach of the designers of an investigation by delegating data collection to instruments or methodical human data gatherers. The plan is inflexible and cannot readily be adjusted in the light of the initial findings.

Knowledge-based projects have fewer precedents to guide the activity. They explore the unknown, and procedures must be modified as more is learned. The students who arrive at the outcrop not knowing what they are going to see are in this position. In fact, projects are likely to involve both rules-based and knowledge-based procedures. For example, a seismic survey may collect data according to a predetermined scheme, but the interpretation of those data is knowledge-based, evolving as ideas are tested and knowledge of the geology of the area grows. The student, carrying out an exploratory investigation, may nevertheless follow predetermined conventions when measuring strike and dip, and might even randomize the sampling procedure to aid subsequent analysis. Every rules-based activity is ultimately set in the knowledge-based framework of the science as a whole.

The distinction between knowledge-based and rules-based activities is important in the present context, because (work on artificial intelligence notwithstanding) knowledge remains largely the preserve of the human being. The machine, on the other hand, can be adept at following rules. Bear in mind, however, that automation can *support* free thinking and trial and error. An IT response to knowledge-based activities is to use flexible multimedia, creating fully connected and searchable documents. An IT response to rules-based activities, on the other hand, is a rigorously defined model and database supporting standardized applications. It makes full use of instruments for data collection and analysis.

The computer is well suited to accurate long-term storage of complex images and detailed tabular data, such as lists of fossils or results of geochemical analyses; but the brain (where accurate detail is restricted to short-term memory) is not. The brain *is* well adapted to handling, within a frame of existing background knowledge, the loose structure of descriptions, analogies and explanatory reasoning typical of a narrative account; but the computer is not.

IT should aim to harmonize knowledge-based human skills with rules-based computer modeling. As the approaches overlap and combine, the information system should optimize the abilities of both. An important part of the solution is **interactive computing** - a conversation between the user and the machine, in which the screen display is modified rapidly in response to instructions or decisions entered usually by keyboard or mouse. The ability of the computer to follow rules quickly and accurately is complemented by the ability of the scientist to use background knowledge to control the progress of the activity.

1.7 Modes of thought

The various information types (text, spatial, structured, tacit) are handled differently by the brain. Each supports a different style of thought, communication and IT system. We use them, separately or together, to model knowledge and information in various modes of thought and investigation, such as the following:

Narrative - one can pass on information, or develop a point of view, by telling a story. As Francis Bacon pointed out in 1652 (see Leatherdale, 1974), one can revisit the reasons for reaching a conclusion by going over in one's mind the events that led to it. Each part of the narrative depends on the story so far. By telling the tale to others, they too may follow the line of reasoning.

Temporal - geological explanations trace the course of past events, relating observations to a conceptual time sequence of past conditions (states) and processes.

Spatial - geoscience maps and images provide the means to locate observations and link them to spatial pattern and spatial relationships. To understand the pattern of ice flow, or sequence and extent of lava flows, the students (I 3) were led naturally to a map. The meaning of the observations depends on their spatial context.

Demonstration - narrative description is more powerful if augmented by actually demonstrating what happened, in the field or laboratory (the **ostensive** approach). For a full picture, it may be desirable to retrace the work, and so share the experience, of an earlier investigator.

Quantitative - the benefits of precise measurement have wide application and are immediately obvious in, for example, hydrocarbons exploration, where detailed prediction of the form and properties of the strata is required, leading to estimates of the location and magnitude of the oil and gas reserves.

Statistical - statistical theory provides a rigorous basis for marshaling complex evidence for testing hypotheses, and estimating probabilities and uncertainty, by computation from appropriately sampled observations and measurements.

Process-response - the concept that physical, chemical and biological processes operated in the past as they do now, is the basis for much geoscientific thinking. A coherent picture of past processes should be internally consistent and should predict responses (consequences) comparable to those of present-day systems.

Experimental - some processes and responses can be explored by experiment, that is, under circumstances that the scientist can control, leading to a more exact understanding of the relationships.

Trial and error - where the course that an investigation will take is not clear, a **heuristic** approach may be adopted, trying out a range of possibilities, following those which seem most successful, and modifying them as more is learned.

The information type and mode of thought play a large part in determining whether IT methods are relevant and which methods are appropriate. Conversely, technology influences the ways we think and the combinations of modes of investigation.

1.8 The need for a Geoscience Markup Language

This chapter is concerned with where we want to go, not how we get there. However, the requirement may be clearer if we have a mechanism in mind. Conventional methods of recording information have deficiencies. The poor connectivity enforced by earlier technology results in high redundancy and inflexibility. Processes to manage, manipulate and explain information are frozen along with their representation. Change is cumbersome, because minor corrections in the literature are easily overlooked, and when ideas change, the full knock-on effects on other work are seldom obvious.

We are looking for a mechanism that can offer better facilities for new investigations, while incorporating legacy material. A markup language could be one approach to improving communication. It can represent conventional narrative text, but can also include tags. Unlike HTML, where the tags control presentation and links, XML (E 6) can also tag words or sections by topic (this is a fossil name, this is the section on structural geology). Presentation is handled separately through a style sheet and can be controlled by the reader.

As its name indicates, XML is extensible. The ability of users to define their own tags could result in a large and unwieldy language. Therefore, specific dialects of XML have grown up and been partly standardized within subject areas. Thus Chemical Markup Language (CML) can tag chemical formulae, and display them, or models of the molecular structure, with a choice of conventions and notations. Other dialects have been developed for fields such as mathematics and music, to handle their specialist notations and offer flexibility in presentation.

A Geoscience Markup Language (GML) could also be a specialist subset of XML, and thus have the ability to tag words or sections (modules) of the text to reflect their topic. Such modules could be linked to others, within the document or elsewhere. We think of the conventional literature as subdivided into encyclopedias, books, serials, reports, notes, maps, and so on. The subdivisions are based on physical form and process of delivery. Instead, we could visualize a GML document as a collection of modules brought together for a particular purpose, with many of them reusable in other contexts. Authorship takes on a different meaning where modules are assembled from many sources, possibly offering alternative explanations of the same phenomena.

A markup language, however, also offers the possibility of linking text closely and selectively to modules from, say, metadata, data, software, models, demonstrations on video or sources of expert advice. Many of these would be accessed through a database management system or a GIS (E 5), although this need not be apparent to the casual user. A module could be displayed in different ways to meet individual needs (a table of data or a graph, a contour map or a perspective view). It would be an integral part of the document. But it would also offer the possibility of moving to and from a different environment, such as a GIS to explore the spatial context, or a database for comparison with analogous datasets from other investigations. Another user option could be to select the level of generalization, while retaining the ability to drill down to additional detail when required. Change to one module would be seen by all modules connected to it, and knock-on effects could be traced through the links.

The technology is largely in place, and starting to operate in some other subjects. Such languages immediately offer greater flexibility and expressive power to the author. Geoscientists have special needs for interworking with spatial, historic and stratigraphic information, and have their own vocabulary and procedures. They could therefore justify a separate dialect of XML. But there is a long and painful learning curve, and geoscientists will be involved in much trial and much error before a robust solution for everyday use can emerge.

With an appropriate markup language, linkage to the underlying context of assumptions, laws and hypotheses could be recorded and therefore the effects of changes in the underlying ideas could be clarified. To take this further, however, we need to consider the process of building knowledge from information, and the object-oriented approach that stems from this.

2. Processes and the repository

2.1 Explanation

Having looked at communication (J 1), the next question is how scientists explain their observations and make sense of streams of observational data. How do we build knowledge from information? The Encyclopedia Britannica (1973) is again helpful with its entry on *Scientific Method*, where it defines the pursuit of science as “the search for knowledge and understanding through formulation of the laws of nature.” The **theoretical function** of science is that of providing explanations of natural phenomena by discovering relationships between these phenomena and other events. These relationships fall under general laws that enable us to make predictions as to what events to expect in particular circumstances, and sometimes, by controlling the circumstances, to ensure that these events will occur. The **practical function** of science, that of enabling us to adjust our lives to nature and, sometimes, nature to our lives, thus derives from its intellectual function - that of explaining phenomena by means of scientific laws.

A starting point for scientific explanation is **classification** (systematically assigning objects to categories based on their properties). Recognizing an object as a sediment or an intrusion has implications about its geological behavior. The words used to name objects are nouns. Adjectives, like red, angular or hard, describe their attributes but are less useful for classification, saying little about how the objects behave. The students' visit to Salisbury Crags (I 3) began by distinguishing, identifying and naming the sedimentary and intrusive rocks, relating them to their behavior in the geological past. The extension of this activity to formal data analysis is described in H 3.

Scientific discovery involves finding **hypotheses** (suppositions made as a starting point for further investigation) that could be refuted by further experience (see Popper, 1996), but which nevertheless survive testing by observation or if possible by **experiment** (observations made in circumstances over which the scientist has control). A hypothesis in I 3 was that the rocks of the cliff face are part of an intrusive sill. This was tested by thinking about what additional observational evidence might be found and then looking for it. Although the original events are beyond the reach of experiment, a detailed model of aspects of the processes, say, the baking of the sandstone, might be tested experimentally given appropriate facilities to replicate the high pressures and temperatures involved. The purpose may be to find more laws about the behavior of the things we can observe or to incorporate the results in a broader explanatory theory.

Scientific laws enable us to organize our thinking into coherent systems as well as to make predictions. The laws are at many different levels of generality, arranged in a hierarchical system in which laws at a low level are logical consequences of sets of laws at a higher level, and so on. The lowest-level laws are general propositions whose instances are directly observable facts, but higher-level laws may be theoretical concepts in a wider system explaining new phenomena. Explanations in geoscience may present the observed situation as a logical consequence of preceding events and of more fundamental regularities, such as the laws of physics, operating on initial conditions of a geographical or historical kind. If we view the process of successive explanation as the erection of a hierarchy of laws of increasing generality, there is no reason to prevent different hierarchies from being constructed in different ways to explain the same phenomena (Kent, 1978).

2.2 Analogy

“To explain the origin of hypotheses I have a hypothesis to present. It is that hypotheses are always suggested through analogy. Consequential relations of nature are infinite in variety and he who is acquainted with the largest number has the broadest basis for the analogic suggestion of hypotheses” (Gilbert, 1896, quoted by Leatherdale, 1974).

Analogy is the resemblance in some particulars between things otherwise unlike. Analogy can be a resemblance in an ensemble of qualities, or of properties or attributes. In metaphor, the mind sees and expresses an analogy. The metaphorical use of language in science arises when familiar vocabulary is extended to describe novel insights and interpretations. Thus in coining the term electric current, ideas were carried across from the familiar current in a river. **Metaphor**, according to the Oxford

English Dictionary, is “the figure of speech in which a name or descriptive term is transferred to some object different from, but analogous to, that to which it is properly applicable.” Analogy in logic, according to the same source, is the process of reasoning from parallel cases, based on the assumption that if things have some similar attributes, other attributes will be similar. Most of the truly fruitful facts about nature, Leatherdale (1974) suggests, have been discovered by reasoning from analogy.

According to Leatherdale, explanation involves an inescapable use of analogy. This is partly because the unobserved part of the description in an explanation, being unobserved, cannot be directly described. It must be verbalized and conceptualized in terms of other experiences. Explanation works by analogy of content as well as of structure. When the analogy is well marked in terms of content, or observable characteristics, we speak of a **model**. The model is an essential tool, in that it enables us to think about the unfamiliar in terms of the familiar.

Models enable us to construct and meaningfully describe the concepts of theories in the same way as metaphors enable us to think about or describe things or concepts not normally describable in a literal vocabulary. Because they function in this way, they give meaning to, and thus an explanation of, theories. This in turn enables them to connect theory with observation and experiment. Thus, in the belief that past processes obeyed the same physical and chemical laws as today, analogies are drawn in geology with present-day processes, as in comparing an unconformity to a present-day erosion surface, in seeing finer crystals as indicating more rapid cooling, or in explaining changes in sandstone petrography as baking against an igneous intrusion.

2.3 Model and reality

Geoscience investigations are usually concerned, not with creating a new model, but with refining an existing one. They build on earlier work and must be closely linked to past records. The model is concerned not only with what is there, but also how it came about - how the operation of physical, chemical and biological processes, in a sequence of events in geological time, brought about the observed consequences. The model influences the classification of objects. For instance, the geologist sees an important distinction between a granite and an overlying pebble conglomerate, despite their similar composition and appearance, because they formed in quite different circumstances.

As pointed out earlier (B 4), the neat and tidy classification of rocks shown on a geological map or reported in the literature is unrealistic. The overlap, ambiguity and uncertainty so painfully apparent in the field have been banished. Crisply bounded areas of uniformity have somehow replaced the fuzzy boundaries and mish-mash of intercalated variation. Goodchild (1992) suggested that: “We need better methods for dealing with the world as a set of overlapping continua, instead of forcing the world into the mould of rigidly bounded objects.” Quantitative techniques (F) are a possible candidate for representing the geological “continuum”, although Mandelbrot pointed out that continuity is conspicuous by its absence in natural phenomena (G 6). We now need to consider whether the categories are necessary, or an artifact imposed by inadequate technology.

The distinction between model and reality is an important one. The model must be tuned to human thought while reflecting something useful about the real world. A continuous model, such as a contour map, may be an appropriate representation of discontinuous reality, as long as the discrepancy is not important in the context in which the model is being used. The danger comes when the limitations are forgotten and questions that cannot be properly answered by the model, such as the length of a coastline (G 6), are addressed within it. Separate models relying on different assumptions are required for different purposes. For example, a statistical model might regard a process as deterministic and predictable, together with a superimposed random element for which only the statistical properties (as opposed to individual instances) could be predicted. On the other hand, a dynamic nonlinear model might regard the process as deterministic (in the sense of following natural laws) but unpredictable because small variations in the initial conditions could lead to a large change in the outcome (Baker and Gollub, 1996).

There is another issue. The model must be one that the available data and technology can support. New IT solutions extend the range of models that are realistically available. Looking at the computer display illustrated in Fig. 1, for example, there is a clear possibility of modeling three-dimensional rock bodies in new ways. The image can show discontinuous areas. Zooming in to part of the image could cause the areas to fragment and reveal more detail, as discontinuous as before. The scope of the model in Fig. 1 is limited. It is a stunning image when seen in full motion on the screen, and no doubt serves its purpose well. But it is tied to just one set of properties, related to acoustic impedance within a body of rock.

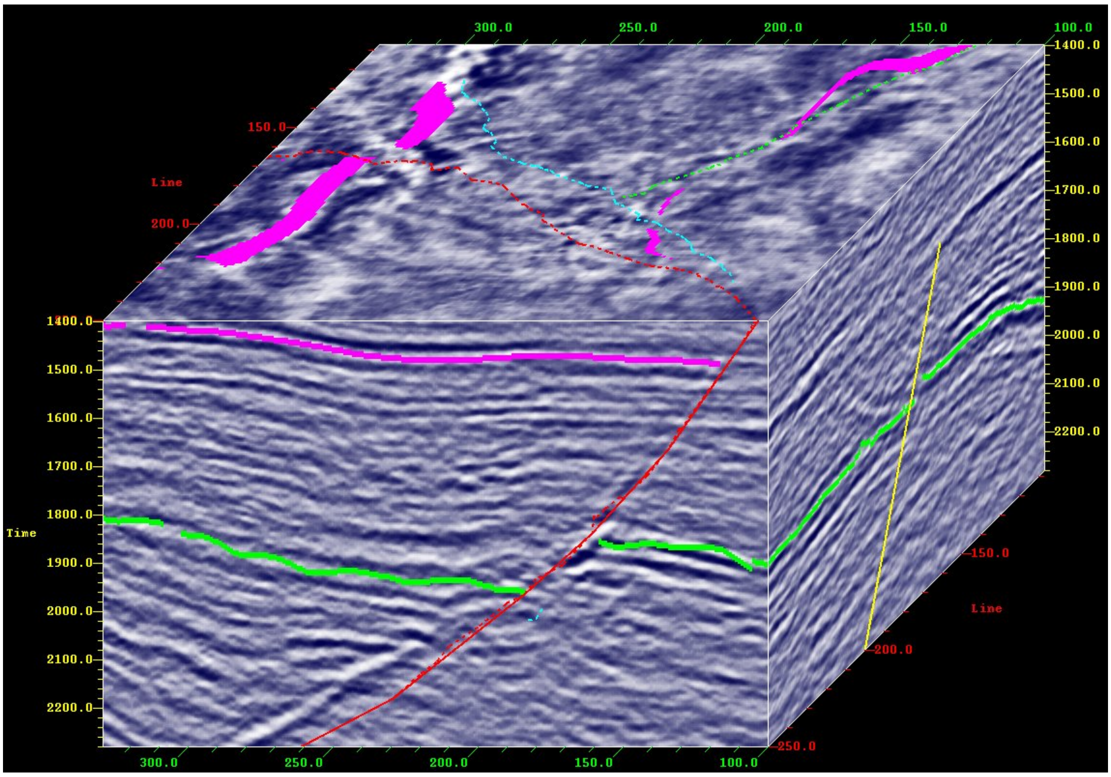

Fig. 1. Display of 3d seismic data. Animation (not available here) enables you to move through the data volume to follow structural and stratigraphic trends. Reproduced by permission of Landmark Graphics Corporation. More at http://www.lgc.com/

Take another example. Satellite imagery records a number of related properties, namely separate bands of the electromagnetic spectrum measured simultaneously. The properties can be analyzed quantitatively and, for example, classes based on discriminant analysis (F 5) can be displayed (Fig. 2). They again serve their purpose well, and are a useful reminder of the variation hidden in conventional cartography. But they complement and extend earlier methods, rather than displacing them. One reason is that three-dimensional seismic surveys and satellite imagery are unusual in their dense and regular sampling patterns that make detailed analysis possible. Relating them to other variables sampled on other patterns by other means calls for background knowledge and human interpretation. Limitations of access and measuring procedures mean that the data for most variables are inevitably inaccurate and incomplete.

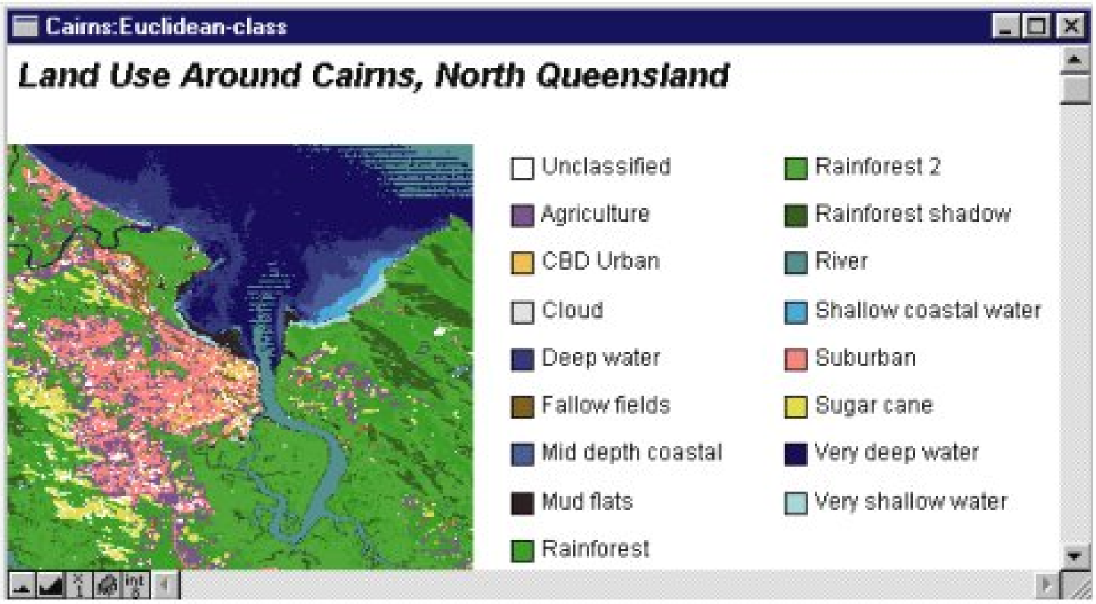

Fig. 2. Classification of land use from a satellite image. Example of satellite imagery classified by an iterative technique. The user indicates typical areas for each class, the computer extrapolates by quantitative analysis of the spectral bands and displays the color image, the user corrects and extends the classification, and so on. Published by permission of Rockware. More on http://www.rockware.com/catalog/pages/dimple.html

Our aims cannot be solely descriptive, for geoscience is concerned with recreating a story from incomplete evidence. It is a story about objects, identified by nouns, given meaning by the models in which they participate, and their properties, described by adjectives of qualitative assessment or quantitative measurement. The creation of the main object classes, such as stratigraphic units, and the assignment of instances to these classes, seldom depend on subtle quantitative comparisons. They depend on drawing analogies and spotting crucial features. They depend on building up a pattern of behavior of objects within models and relating them to a place in the overall scheme of things (the current paradigm), more likely tied to fuzzy concepts than to measurable properties.

Even where an example or instance of an object class is described, such as the lithology of a core or well sample, a descriptive term (say, biohermal dolomite) may place it in a category that reflects an impression of many characteristics. We thus benefit from the ability of the human brain to recognize complex patterns specific to the context. This lends itself to narrative description, not to point-by-point quantitative comparison. From an initial broad appreciation of the situation, we observe and describe to extend our model. It is not difficult to think of examples where several conceptual process models are invoked. In the example at I 3, they dealt with sedimentary deposition, igneous activity, regional tilting and glacial erosion. The processes they refer to are worldwide. But we were looking at their consequences within a small area. Without conscious deliberation, we selected objects (the rock bodies) that took part, with the same definition, in each of the models. We naturally placed the objects and the processes at appropriate positions within the same framework of space and geological time. In M 2.3, we look at spatial and stratigraphic models which make that framework explicit for computer processing. Meantime, we note that objects seem to be chosen by a subtle process that depends on the intuition and background knowledge of the human mind.

At a general level, object classification and identification (H 5) are well suited to our thought processes, and may be assisted but not greatly changed by new technology. Quantitative methods, on the other hand, are well suited to computer processing. They are appropriate where there is a clear physical model that can be represented mathematically, as in seismic processing, gravity corrections, and so on. Their crucial contribution may lie, not in interpreting the geology, but in clarifying the geological significance of the records by removing extraneous effects (F). Where wider conclusions follow quantitative analysis of the raw data, as for example in seismic stratigraphy, they may result from non-quantitative reasoning. The new insights nevertheless depended on technology extending the reach of human thought.

Subtle variations in properties or composition (M 2.3) may be detected by quantitative analysis (F 5), as, for example, the identification of distinct lava flows from petrographic studies in the example of I 3. Statistical reasoning, based on randomly sampled measurements, can be a surprisingly powerful approach, even when based on the apparently weak concept of testing whether observed patterns were likely to have arisen by chance.

IT offers the opportunity to build models that span the modes of human thought (J 1.7), combining their individual strengths. Quantitative reasoning, such as a computer process that simulates states and events, can be embedded in a narrative that explains its significance, limitations and context. Quantitative reasoning can be linked to cartographic and spatial thinking by computer visualization. Individual measurements gain meaning from the context of spatial pattern. Human insight, intuition and modes of thought remain supreme, but can be expressed in new ways. The wide range of models which IT supports can lead to better understanding, provided their properties and the limitations of their analogies are clearly appreciated.

2.4 The object-oriented approach

Our thought processes, constrained by technology, ultimately determine how we record and handle data. The object-oriented approach attempts to match those processes with IT procedures. According to Coad and Yourdon (1991), three methods of organization pervade our thinking about the real world:

- differentiation of experience into particular objects and their attributes
- distinction between whole objects and their component parts

- distinction between different classes of objects.

It is not difficult to think of examples from geoscience in terms of rock types, fossils, stratigraphic units, geological processes and so on. For instance, here is an outcrop (object), somewhat overgrown and deeply weathered (attributes), beside the river and under the bridge (spatial relationships). The outcrop (whole object) consists of beds of sandstone (component parts), containing grains of quartz and mica (components). It is interbedded (spatial relationship) with shales (different object class), and contains (spatial relationship) fossils (different object class).

Reality is a seamless web of infinite complexity, but the human mind can cope with only a limited amount of information at one time. Abstraction reduces the complexity by separating out a **model** dealing with a small number of things that are important to the purpose in hand. All words, language and data are abstractions and incomplete descriptions of the real world. There is thus no correct model of a situation, only adequate or inadequate ones. An **object model** describes the structure of objects in a system, their identity, relationships with other objects, attributes and operations. Common relationships are *being* (as in sandstone *is a* sedimentary rock), *having* (as in this sandstone *has* graded bedding) and *doing*. A **dynamic model** describes those aspects of a system concerned with time and the sequencing of operations - events that mark changes, states and organization, whereas a **functional model** captures what a system does, without regard for how or when it is done.

Language, images, quantitative modeling and demonstration all share the tendency to see the world in terms of objects, attributes and processes, from which may spring the noun, adjective and verb structure of our language (Leatherdale, 1974). Thus, communication in geoscience, by whatever means, concerns **processes** (which cause things to change) and **objects** (the things of interest), the object classes, and their **attributes** (properties, composition, relationships and behavior). An object should not be constrained by information type. One object, such as a borehole description, might comprise a text description and a geographic reference. It thus includes both text and spatial information types, which might be stored separately and accessed by different software. Many of the objects invoked in a narrative have second homes in other, possibly more structured environments. For example, a paper describing a fossil locality might include a list of species that could also appear in a paleontological register, and could be plotted on a map of fossil distribution, and linked to a stratigraphic table. The user must therefore interface with distributed objects, related to various topics, and represented by a mixture of information types.

Object classes, by definition, belong in a hierarchical sequence (H 5), **inheriting** attributes from classes farther up the hierarchy. Thus *sandstone* may inherit properties from its superclass *sedimentary rocks*. A data model (I 2.2) can assemble object classes into topics such as (examples in brackets): stratigraphic (formation), bibliographic (document), petrographic (thin section), paleontological (specimen of fossil). The topics are not mutually exclusive, so that a fossil description could be both a paleontological object and a bibliographic object. Within each topic, rules and standards can ensure that information is consistent and comparable. The fossil, as a paleontological object, is named according to the rules of fossil nomenclature, described according to paleontological conventions. The fossil description, as a bibliographic object, is cataloged according to international rules. A single object may thus be firmly embedded in at least two topic areas. We return later (L 6.1) to the application of object-oriented methods in analysis, design, programming and database work.

3. References

Baker, G.L., Gollub, J.P., 1996. Chaotic Dynamics: an Introduction. Cambridge University Press, Cambridge, 256pp.

Coad, P., Yourdon, E, 1991. Object-oriented Design. Yourdon Press, Englewood Cliffs, N.J., 197pp.

Encyclopedia Britannica, 1973 edn. William Benton, Chicago.

Gilbert, G.K., 1896. The origin of hypotheses, illustrated by the discussion of a topographic problem. Science, N.S., 3, 1-13.

Goodchild, M.F., 1992. Geographical Data Modeling. Computers & Geosciences, 18 (4), 401-408.

Kent, W., 1978. Data and Reality. North-Holland Publishing Company, Amsterdam, 211pp.

Kuhn, T.S., 1962. The Structure of Scientific Revolutions. The University of Chicago Press, Chicago, 172pp.

Laszlo, E., 1972. The Systems View of the World. Braziller, New York, 131pp.

Leatherdale, W.H., 1974. The Role of Analogy, Model and Metaphor in Science. North-Holland, Elsevier, Amsterdam, 276pp.

Popper, K.R., 1996. Conjectures and Refutations: the Growth of Scientific Knowledge. Routledge, London, 431pp.

3.1 Internet references

McCrone, J., 1999. Going inside - the neuronaut's guide to the science of consciousness. http://www.btinternet.com/~neuronaut/index.html

MacEachren, A.M., 1998. Visualization - cartography for the 21st century. International Cartographic Association Commission on Visualization conference, May 1998, Warsaw, Poland. http://www.geog.psu.edu/ica/icavis/poland1.html

<<<Back to Table of Contents

On to Part K: Coping with changing ideas. Defining the user requirement for a future information system>>>