<<<Back to Table of Contents
On to Part L: Adjusting the emerging information system to new technology>>>

Geoscience after IT: Part K

Coping with changing ideas. Defining the user requirement for a future 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. A99-A108

Abstract - The information system must deal with the diversity of ideas in geoscience and their changes through time. To communicate information, ideas must be aligned and molded to fit a shared view of the world. Change can be traumatic and may be deferred until obvious benefits force old ideas to give way to new, and even then individuals only partly reconcile their ideas. The mechanical records of IT must reflect the flexibility, overlap, ambiguity, inconsistency, conflict and evolution of human interpretation. These, and other needs considered in earlier parts of *Geoscience after IT,* are brought together as a statement of what we want from the system, set out as a user requirement.

Key Words - Aligning ideas, paradigms, learning curve, user requirements.

1. Change

The geoscience record is in constant flux. New ideas and new data are continually being added, and old ideas and data revised. Conventional methods struggle with limited success to maintain a record which is readily accessible and up to date. If we are to find better ways, we need to form a view on how change works. IT must cope, not just with the changes it creates, but also with the diversity of ideas in geoscience and their changes through time.

1.1 Flexibility and sharing knowledge

The information system must be flexible in order to respond to change. For example, words can retain their place in a growing science only because their meaning depends on the context. Think for instance of the word *fault*, and the ideas it might bring to the mind of field geologists using the concept to explain the outcropping of sediments of unexpected age, and possibly searching for landscape features to mark the fault as a line on the map. Their views of its characteristics and connotations differ from those of, say, the seismic interpreter, the seismologist locating an earthquake epicenter, a prospector looking for fault-related minerals, or the structural geologist studying the movement of continental plates. The same word used by a geologist a century ago would carry subtly different implications, embedded in the knowledge and thinking of the time. The computer engineer, for whom the word has a totally different meaning, could be forgiven for failing to see even a metaphorical connection. A keyword search for documents about faults could be unhelpful. Nevertheless, the ambiguity associated with analogy (part J, section 2.2) gives room for growth and extension of ideas.

Information gains its meaning from its context. Geoscience information is gathered and made available to the information system from a variety of large and small projects (I 8.1). The projects are not devised within the information system, but are undertaken for reasons that stem from their **business setting**, which determines the objectives. The objective may simply be to satisfy curiosity. More likely the studies are directed to, for example: the search for oil and mineral wealth or help in its exploitation; collecting background information for protection of the environment; avoiding geological hazards or optimizing land use; or an attempt to understand more clearly the processes that formed the earth. The project objectives affect the sampling scheme, type of data, data collection method, and operational definitions. The data can be fully understood only through knowledge of the project and the approach used.

The development of a model of some sort precedes and is the subject of every investigation. Data from different projects may use the same terminology but different models, and thus be misleadingly similar but not fully compatible. This is one reason why the concept of a database as a pool of shared information (H 3) must be approached with care in geoscience. The important relationships among projects may be between models rather than between datasets. It is entirely possible that the models may be implicit rather than explicitly defined, which adds to the difficulties of data integration. Furthermore, a range of alternative models (multiple hypotheses) may be considered in parallel within a single investigation, as advocated by Chamberlin (1897). Yet from a multitude of independent projects there springs a coherent and integrated body of knowledge, as though coordinated by unseen hands. How does this happen, and how will it be affected by changing information technology?

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 integrating numerous concepts and results. Conformance with accepted procedures is encouraged by peer review, editors and referees, examination boards, textbooks, and standards organizations. Industry may encourage standards, for example to make more efficient use of information collected during hydrocarbon exploration. Government, with an interest in royalties and thus in the overall efficiency of the process, may reinforce this with legislation. On its own account, government may play a part by funding surveys of, say, topography, geology, soil science, hydrology, or oceanography. As long-term organizations, surveys tend to develop a uniform house style for investigation and presentation of their results.

A benefit of a standard approach is that it simplifies the exchange and integration of information. Object classes and their relationships, which can be defined formally in data analysis, provide a context into which new information can be fitted. Hypotheses are erected for further investigation and linked to the current hierarchy of scientific laws. **Standards** are created by assertion and negotiation; enforced or encouraged by custom, education, agreement, peer pressure and sometimes legislation. They all contribute to a shared frame of reference in which ideas are more readily exchanged, part of the map for scientific research (K 1.2). Establishing the relationships between models and harmonizing the underlying concepts is an important theme in the geoscience literature.

There is, however, a **trade-off**, that is, some benefits are gained at the expense of others. Collecting data to be widely useful imposes an additional cost on a project, possibly unnecessary for the immediate objectives. A standard approach limits flexibility, and can lead to an unduly narrow view. **Diversity** arises from divergent objectives, fragmentation of disciplines, rival or competitive organizations seeking a new niche, research into new possibilities, availability of better or cheaper non-standard methods, and attitudes such as preferring ownership to communication of information. Diversity is particularly associated with the early experimental phase of a new development. As ideas mature, and a general paradigm gains wide acceptance, the emphasis of the science and the attitude of the scientists change from innovative to methodical. Standards are valued more highly. Exploratory investigations, which are knowledge-based and proceed by trial and error, may be supplemented by systematic, pre-planned rules-based studies.

Diversity also arises from ideas changing with time. Philosophers remind us that we can expect all scientific information ultimately to be wrong. Information repositories, such as the scientific literature, contain much that we accept, if only because the scientific community has so far failed to disprove it. Other information we might regard as no longer entirely valid because it conflicts with more recent ideas or new data. However, there are many strands in a complex explanation and in the observations that support it. They involve ideas from many sources at varying levels of generality, put together in different ways to explain the same phenomena. The POSC Epicentre Model (POSC, 1997) relates observations to "**activities**", thus bringing distinct versions of data, possibly collected at different times with other instruments or objectives, into the same setting.

A study that we regard as based on unacceptable reasoning may contain information that has residual value for unforeseen use in a new context. For example, a borehole description with unbelievable stratigraphy might yield useful data on lithology. The use of analogy and metaphor introduces an element of ambiguity and flexibility to scientific reasoning (J 2.2). By permitting interpretation in several contexts, analogy offers the prospect of reworking old material and finding residual value in otherwise obsolete information. Its inevitable imprecision helps cross-fertilization where data or ideas are placed in a new context. It is not surprising, in these circumstances, that a large part of any project is devoted to the difficult tasks of finding, assessing and reinterpreting earlier studies, driven by the need to accommodate change.

1.2 Paradigms

We tend to see what we look for, and more strikingly fail to see what we do not look for. Minsky (1981), in his well-known work on machine intelligence and the human-computer interface, used the concept of **frames** to describe the intricate context in which ideas are embedded by the human mind. An idea communicated from one individual to another can be fully understood only if the recipient (man or machine) has an appropriate frame in place to receive it. In other words, the ability to grasp an idea depends on what you already know. Data dictionaries (H 3) reflect this concept by defining and placing in context the terms used to record data. Laszlo (1972) made a broader statement: “There is no theory without an underlying world view which directs the attention of the scientist. There is no experiment without a hypothesis and no science without some expectation as to the nature of its subject matter. The underlying hypotheses guide theory formulation and experimentation, and they are in turn specified by the experiments designed to test the theories.”

Observations are set within a framework of current ideas. Thus the neptunists, believing all rocks to have been precipitated from a primitive ocean, could not have been expected to interpret correctly, or even to observe, the features which identify Salisbury Crags (I 3) as an igneous sill. During systematic examination of outcrops, however, unexpected features may be spotted which throw additional light on the nature of the rocks. Their significance may derive from analogies with observations elsewhere, or like Hutton's unconformity, with present-day processes. In many cases they would not be noticed except by a trained geologist aware of their possible significance, just as graded bedding, sedimentary structures or trace fossils must frequently have been visible to, but overlooked by, earlier generations of geologists.

Kuhn (1962), in his work on *The Structure of Scientific Revolutions*, distinguishes between “normal” science and revolutions in science. Normal science is based on a well-established view of a science in which the practitioners share the same exemplars or paradigms. Results are addressed only to professional colleagues, whose knowledge of a shared paradigm can be taken for granted, and who prove to be the only ones able to read the papers addressed to them. The **paradigm** comprises universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners. When the individual scientist can take a paradigm for granted, he need no longer in his major works attempt to build the field anew, starting from first principles and justifying the use of each concept introduced. That, as Kuhn dismissively remarks in his textbook, can be left to the writer of textbooks.

The need for experimental work, according to Kuhn, arises from the immense difficulties often encountered in developing points of contact between a theory and nature. Observation and experience can and must drastically restrict the range of admissible scientific belief, else there would be no science. But they cannot alone determine a body of such belief. “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.” (Kuhn, 1962, page 108) Three classes of problem - determination of significant fact, matching of facts with theory, and articulation of theory - constitute the literature of normal science. Research can be seen as a strenuous and devoted attempt to force nature into the conceptual boxes supplied by professional education. Once the reception of a common paradigm has freed the scientific community from the need constantly to re-examine its first principles, the members of that community can concentrate exclusively upon the subtlest and most esoteric of the phenomena that concern it.

One strong, but false, impression is likely to follow: that science has reached its present state by a series of individual discoveries and inventions that, when gathered together, constitute the modern body of technical knowledge, in a process often compared to the addition of bricks to a building. That, Kuhn claims, is *not* the way that science develops. Discovery commences with the awareness of anomaly - nature has somehow violated the paradigm-induced expectations that govern normal science.

Kuhn quoted an experiment by two psychologists, Bruner and Postina, who asked subjects to identify playing cards on the basis of a very brief glimpse. The experimenters introduced occasional cards of anomalous color, such as a black 4 of hearts. This was identified as the 4 of hearts or sometimes as the 4 of spades. Without any awareness of trouble, it was immediately fitted to one of the conceptual categories prepared by prior experience. When the brief glimpses were extended to a somewhat longer exposure, however, the subjects suffered acute distress and some broke down in confusion. Similarly, the emergence of new theories is generally preceded by a period of pronounced professional insecurity, generated by persistent failure of the puzzles of normal science to come out as they should.

When the profession can no longer evade the anomalies that subvert the existing tradition of scientific practice - then begin the extraordinary investigations that lead the profession at last to a new set of commitments, a new basis for the practice of science. A scientific theory is declared invalid only if an alternative candidate is available to take its place. A new theory is always announced together with applications to some concrete range of natural phenomena; without them it would not be seen as a candidate for acceptance. It is seldom or never just an increment to what is already known. "... schools guided by different paradigms are always slightly at cross-purposes. At times of revolution, the scientist's perception of his environment must be re-educated - in some familiar situations he must learn to see a new gestalt. Thereafter, the world of his research will seem, here and there, incommensurable with the one he had inhabited before" (Kuhn, 1962, page 111). This intrinsically revolutionary process is seldom completed by a single worker and never overnight. Kuhn quotes Max Planck: "a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."

Kuhn concerns himself largely with development of theory, but claims that the distinction between novelties of fact (discoveries) or novelties of theory (inventions) is artificial. He also points out that it may be the endurance of instrumental commitments that, as much as laws and theory, provide scientists with the rules of the game. Computer support, for example, makes possible a systems view of geoscience and some reformulation of traditional geoscientific reasoning on a more rigorous mathematical basis. Fundamental change to the information system can have far-reaching effects on the way science is conducted.

1.3 Dynamics of change

A plausible picture of the development of new technology is the so-called learning curve or S-curve like that of Fig. 1 (compare Coad and Yourdon, 1991). The vertical axis represents some measure of the appropriateness, success or value of a new development, say, the automobile, the telephone, or the computer. The first stage of invention and experimentation proceeds slowly until initial successes create interest, investment and consequent rapid development. Subsequently, as the technology matures, progress is slowed again by the law of diminishing returns. By then, the main framework of the system is fixed, and change is limited to slow, marginal improvement.

Consider, for example, the procedures of geological mapping which took shape in the early nineteenth century. An initial phase of discovery and invention was followed by experimental diversity and rapid progress as new methods of representing knowledge of spatial characteristics were developed (Rudwick, 1976). This in turn gave way to slow, systematic consolidation, building resistance to further change. Mapping methods were refined to the point where enhancements became marginal, and consistency was valued above innovation. Thereafter, newcomers to the craft were trained to follow conscientiously a set of well-established procedures. A blinkered view is a positive asset when knowledge-based work gives way to a rules-based approach. Fig. 1 suggests desirable qualities in the practitioners at different stages of a maturing technology.

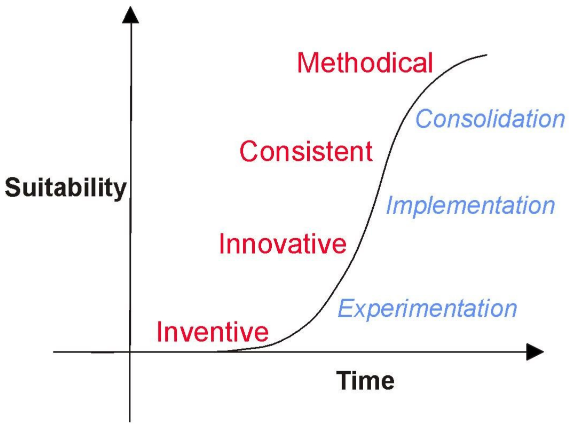

Fig. 1. Learning curve - the development of new technology. Techniques improve slowly at first, but initial success leads to investment and rapid growth, curbed eventually as innovation to a mature system shows diminishing returns.

Largely unheeded by the traditionalists, new supporting technology is set to sweep aside assumptions on which their hard-won skills were based. Technology tends to displace existing procedures rather than support an entirely new departure. There are at least two distinct S-curves: one showing the development of the older technology, the other the new (see Fig. 2). During its initial development, the newer technology is unlikely to be competitive with the old. Not until the new technology has reached the stage of rapid growth do many workers in the field see benefit in adopting a new approach. Those able to accept new ideas may move ahead, supported by the new technology. Earlier information may need to be reworked or lost. But skilled workers who were selected for their conscientious dedication to repetitive routine may be psychologically unwilling to adjust. There is discomfort and risk in moving from a mature technology to a fast-developing one.

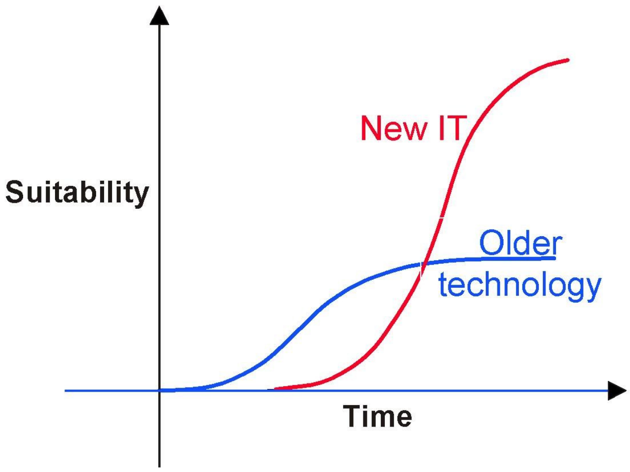

Fig. 2. Crossing of two learning curves. New technology displaces the old when clear benefits appear.

The information system as a whole is based on technologies that are being systematically superseded by computer-based techniques. Measured by cost-effectiveness, the S-curves may already have crossed. But the curves are a gross over-simplification. There is no single paradigm shift. Change occurs at all levels of detail, and may have knock-on effects and implications for other levels. There are many strands of information technology applicable to various branches of geoscience. No single, smooth curve can be examined to see where we stand.

A better analogy might be a mountain which is surrounded by many foothills and shrouded in impenetrable fog. The obvious strategy in aiming for the summit is to go upwards. But when you apparently reach the top (because every direction leads downhill) you may merely be on one of the lower foothills. A research environment can be much like this, but with many workers throughout the world starting from different points and following different routes up the mountain. Despite the fog, research workers can get an idea of their relative success by shouting to one another, or, more conventionally, communicating through conferences and the scientific literature. The individual who has reached a sub-optimal peak and hears shouts from above, can take a bearing and proceed in the direction of the sound. Pushing the analogy a little further, the researcher who is comfortably atop a low foothill might consider carefully before discarding cherished ideas to make a long and dangerous traverse, aiming for a higher foothill which, on arrival, might turn out to have been abandoned in its turn. This imposes a degree of stability in the system. For most research workers, only major benefits justify a change of direction. In these circumstances it is invaluable to gain some idea of the lie of the land.

Inventions and new techniques, as well as new discoveries, can displace earlier commitments despite inevitable resistance to change. Blackmore (1999) used the term **memes** to describe ideas, skills, habits, stories, songs or inventions that are passed from person to person by imitation. Taking a meme's eye view, she suggested how meme's evolved, meeting the prerequisites of evolution - variation, selection and heredity. If the process of science is seen as developing and testing models, then it is to be expected that, although the main body of geoscience knowledge is unlikely to be overturned in the foreseeable future, it will undergo continual amendment. Hypotheses will be disproved and new ideas emerge. The fittest, as selected by the scientific community, will survive.

Within their own areas of specialization, scientists may actively seek inconsistencies in the paradigm, attempt to disprove proposed explanations and align the model to their own concepts, thus encouraging diversity and evolution of ideas. Outside their specialism, they are more likely to accept a consensus view. The impact of replacing a model and the knock-on effects on the remainder of the knowledge base are determined by the model's scope and relationships. In geoscience, ideas and methods change at all levels of detail, sometimes with knock-on effects creating minor incremental shifts and partial inconsistencies that ripple gradually through the information system - paradigm drift rather than paradigm shift.

1.4 Reconciling ideas

All individuals presumably have their own unique view of reality, based on their personality, training and experience. By arranging and classifying the stream of sensory experience that impacts on short-term memory, they develop their episodic memory of events and their relationships, and the semantic memory defining their current world view (I 4). The ambiguity and inconsistency of human thought make it possible for one individual to hold incompletely defined opinions and a selection of alternative, possibly incompatible views. This has the advantage that a group of individuals can align their ideas with each other and reconcile their views when required for the purpose in hand. Their reconciliation probably does not extend far beyond the requirements of that purpose, and may not conform in every detail. But it enables you to read and understand this without necessarily believing a word of it.

Communication requires a common **frame of reference**, that is a shared viewpoint or set of presuppositions, which can only be developed through education and training. The diversity of ideas and the difficulty of understanding subjects outside one's own chosen field suggest that alignment of ideas is incomplete, partial and specific. Examples can readily be imagined among the group of students examining the outcrop (I 3). **Knowledge** (justified true belief) can be seen as the product of a **social system**, that is, how people perceive each other and their shared activities.

"No two people have a perception of reality which is identical in every detail. In fact, a given person has different views at different times. . . But there is considerable overlap in all of these views. Views can be reconciled with different degrees of success to serve different purposes. By **reconciliation**, I mean a state in which the parties involved have negligible differences in that portion of their world views which is relevant to the purpose at hand. . . . . For the purposes of survival and the conduct of our daily lives (relatively narrow purposes), chances of reconciliation are necessarily high. . . But the chances of achieving such a shared view become poorer when we try to encompass broader purposes, and to involve more people. This is precisely why the question is becoming more relevant today: the thrust of technology is to foster interaction among greater numbers of people, and to integrate processes into monoliths serving wider and wider purposes. It is in this environment that discrepancies in fundamental assumptions will become increasingly evident." (Kent, 1978, p 202-3)

It seems that the information system must cope with overlapping information from different sources, and with many, possibly contradictory, versions of the same ideas. It must provide mechanisms for retaining ideas in their historical context, and for individual users to sift out their own reasonably consistent working views, without losing sight of the alternatives. It must allow the geoscience community to evaluate, select appropriate models, and build evolving views into their current paradigm.

2. Themes and problems

The potential benefits of IT determine future directions, and a number of themes emerged from earlier chapters. One theme was integration. In conventional systems, the scientist must cope with the separate interface and different content of a book, a map, a discussion group, a seminar, or a field study. Windows on a computer screen can offer a coherent view of the various information types, without undue delays and without librarians or booksellers. Material can be filtered for relevance to the specific user and displayed appropriately. With matching procedures, the user can edit the result, and add new information as text, images or data, again with few delays in making the information available and with a reduced need for human intermediaries. The screen on the desktop has many advantages. The map user can select the topics for display, pan to the areas of interest and zoom in or out to the level of detail required. The reader of text can follow references on the spot, search for keywords, and highlight passages for future reference.

If the benefits are clear, however, it is also clear why they are for the most part potential rather than immediate. The coherence of a well-planned document can be lost in a maze of hyperlinks. The accuracy of short-term memory cannot be brought into play when there are long delays in access over the Internet. Information well-printed on paper is more attractive, more convenient to handle, has sharper resolution and is easier to read than anything on a screen. For detailed study, therefore, a printed version is desirable. It can of course be prepared on a desktop printer, even copying the original page layout if desired. Full-size maps are more difficult, as specialized printers of large size and high resolution are required to produce a good copy. Having selected the appropriate area, scale and topics on the screen, either small extracts can be prepared on a standard printer, or a full-size copy can be prepared by an in-house print shop or by a cartographic bureau. However, if conventional products are to hand, they are likely to be more convenient and of better quality than their new-fangled equivalents.

There are also more crucial problems. In geoscience, digital information for remote access scarcely exists outside the petroleum industry and some large organizations. Much information on the World Wide Web is too ephemeral for bibliographical reference, and, for commercial reasons, digital maps are of limited availability. Globalization, in the sense of worldwide exchange of information regardless of discipline boundaries, promises efficiency gains by reducing redundancy in information holdings and offering rapid access to comprehensive information resources. It depends on widely accepted global standards, but comprehensive standards for geoscience are not in place. There is considerable inertia in the system.

Another theme was that of finding more flexible and rigorous expressions of the scientists' conceptual models. Quantitative, statistical and three-dimensional spatial models were mentioned as more complete and precise representations of scientists' ideas. Annotated photographs and video clips, keyed to the model, can help to connect the interpretation with observations. As methods of communication, they all fail if users lack the equipment or skills to receive the message. The full benefits depend on the system as a whole being centered on IT, and therefore are also affected by the inertia just mentioned.

The theme of metadata was seen as important for a number of reasons. To understand information, you must know how it was collected, and users therefore need access to project metadata. There is also a tendency for different authors to attach slightly different meanings to terms used in papers and maps. It could be helpful to users to have metadata with precise definitions of the objects, and indications of where authors deviate from the standard definition. This can be achieved by hypertext links from the information *to* the metadata. Hypertext links *from* well-structured metadata can also help in searching for relevant material. Starting from the list of topics and relationships in the metadata, it should be possible to trace paths to treatments of these ideas in the literature. Furthermore, if documents indicate their dependence on earlier ideas and background theory through hypertext links, then they are positioned on a map of concepts, which could guide readers to relevant papers within their current understanding. As new ideas are introduced, or old ideas questioned, the links could show the knock-on effects and the ripples of change. Again, however, there is a snag. Links with HTML on the World Wide Web are one-way links to locations, not the necessary two-way and multiple links joining persistent objects.

Yet another theme was the evaluation of contributions to the knowledge base. In some areas, such as the oil industry or in some geological surveys, the quality of data is assessed through rigorous procedures of documentation, checking and evaluation. The user may have more confidence in information coming from a 'brand name' of this kind. The editing and refereeing procedures of scientific journals should serve the same function in a broader setting. The relationship of 'quality' to the intellectual foundations of the science is obscure, however. If there really is a set of ideas and studies generally seen as shared exemplars of how things are done in geoscience, there is no obvious mechanism for identifying that paradigm. Presumably individuals develop their own unique world views from rather fuzzy, overlapping and contradictory ideas. The paradigm appears to be a rather subtle concept, where the practitioners feel they know what is what and pass on the knowledge by nods, winks and tone of voice. A more explicit means of evaluating ideas would help them to evolve efficiently. It should involve the users as well as the providers of information. Again, it can only be part of the wider paradigm shift.

A final theme is the business context, the issue of why geoscientists behave as they do, and what forces drive them. In that sense, business decisions will drive change. Inertia comes from the huge investment in earlier systems, and the commitment of scientists and organizations to existing methods. Many have much to lose and little to gain from change. But on the high slopes, vast commercial enterprises are shifting their ground. Already, IT has brought about local changes throughout geoscience. Technical problems are being overcome. The potential to make money by reducing costs and improving efficiency may have the effect of gravity on a snowfield. When the avalanche finally starts to accelerate, it is only geoscientists who can determine whether or not the fallout benefits themselves and their science. That seems a good reason to list the potential benefits, and consider where we want to go before we arrive.

3. User requirements

A reference list of desirable IT features in the overall geoscience information system can focus ideas and even serve as an idealized check-list for new systems. It must, however, be used with caution. Some features are not economic at current prices and some may not be available at all. Features that require long-term availability may conflict with rapidly evolving technology (L 6, L 6.3). Others imply a change of attitudes which may take many years or may never happen. Most can be provided by other means. For example, user training or specialist support can reduce the need for user-friendly systems. Such a decision can have knock-on effects on other aims, however. For example, if you decide that computer specialists should run the programs, this may rule out interaction between the user and the program. It follows that a broad appreciation of overall developments and their interdependence is needed to make good decisions about even a small subsystem.

If you draw up a **wish list**, that is, a list of features you would like to see in your own system, ask yourself what is practicable and how it can be achieved. If you draw up a **user requirement -** the basis for a contract with IT specialists to supply specific facilities or services - compare the estimated costs of the system over its lifetime with those of alternative solutions. It is unwise to lead the field where no-one will follow. It is unwise to follow others into a dead end. With these provisos, here, for future reference, is an annotated summary of some desirable features in an IT-based information system.

Aide-memoire for a user requirement

The information system includes recorded information and the processes that assemble information and build knowledge. Most geoscience knowledge is held in the minds of scientists, who communicate through the ***user interface*** *with the rest of the system.* ***Repositories*** *store and manage recorded information for access by the originators and others.* ***Processes*** *manage, manipulate and present the information, helping the user to understand its significance and make decisions. The* ***business*** *context determines the objectives of geoscience investigations, and the deployment and management of resources to achieve them.*

3.1 User interface

Methods of accessing and supplying information should suit the users' ways of working and be easy to use, accepting and delivering appropriate information as, when and where required.

1. *User-friendliness*. A consistent, simple user interface that matches the scientist's way of thinking, including memory levels and modes of thought, should be used throughout. The system should support the specific needs of individual scientists; their joint needs within a workgroup; and communication

with the world at large. The information system should be structured to reflect the processes of building knowledge from information.
2. *Coherence*. The interface should be compatible with: other systems in geoscience and other disciplines; the business context; any geoscientific instrumentation that passes data to the system. It should switch readily between browsing existing information; adding new information; and editing, analysis, manipulation and presentation.
3. *Control*. Originators of information, who best understand its significance and their procedures, should be able to determine the form and content of their contributions, and make them available without delay. Users, who best understand their own requirements, should be able to customize the interface to select information to meet their own specific needs and determine the form of presentation.
4. *Middleware*. To maintain a straightforward and familiar user interface, middleware (L 2) should hide the complexities of distributed systems and software such as GIS and DBMS. It should assist access to powerful tools such as SQL or graphical selection.
5. *Hypermedia*. Hypermedia links should allow different types of information (text, spatial, tacit, structured) to be closely associated at all levels of detail, both in collecting the information and presenting it to the user. Within a narrative account it should be possible to embed spatial information and quantitative evidence and reasoning, supported by visualization. It should also provide links to experts for advice, and pointers to archived cores, samples and specimens. An interwoven fabric of ideas should be supported through linkages between objects, processes and metadata.

3.2 Repository

The repository should provide safe, long-term custody of information with ready access to comprehensive, appropriate, current, coherent and testable records.

1. *Integration.* The repository may be partitioned according to information types, separating, say, documents, GIS and database for efficient management. But recognizing that only a shared framework can make communication possible, the partitions should share well-structured metadata, with models which link objects regardless of where they are stored or how they are represented.
2. *Connectivity.* The structure should support complex reasoning, including abstraction and generalization, through a network of links and cross-references among information in all its forms, including pointers to that held in the users' minds. From the palimpsest of overwritten and updated stories, it should be possible to extract material filtered by source and topic, and to drill down as required to detail, to supporting data, or to less popular, conflicting or older views. Dependencies between ideas should be recorded to ensure that change at any point can trigger knock-on effects.
3. *Redundancy and reusability.* The system should rely on connectivity rather than replication, for efficiency and to minimize confusion when changes are made. Later versions should be able to incorporate parts of the earlier by reference rather than repetition. Separation of metadata, objects and processes, should reduce redundancy and increase reusability.

4. *Granularity*. Microdocuments and markup languages, such as XML (L 6.2), should make it possible to handle narrative information in smaller discrete portions (finer granularity). Existing systems for handling spatial and structured data (GIS and DBMS) lend themselves to fine granularity, and can thus complement the detail of subdivided text.
5. *Flexibility*. It should be possible to identify rival paradigms, versions and views and to discover their different implications. From the same knowledge base, a range of software systems (interpreters) should support such activities as training; retrieving observations, interpretations or processes; developing ideas; and exploring analogies.
6. *Integrity.* The evolving structure must cope with past, present and future knowledge. Versions should be frozen on acceptance and retained as necessary for historical reasons, preferably with linkages to show their relationships with the metadata of the time. The system should be able to maintain valid current and historical references while coping with changing ideas, alternative versions and new information (including knock-on effects). Links should not be left dangling when objects are deleted.
7. *Legacy information*. Legacy information should be accommodated in its original form, together with any updated version where value has been added, for example by digitization and markup. The user should be able to inspect current views or views at some previous time and explore the development of ideas.
8. *Disposal*. A clear disposal policy, which does not compromise the integrity of the system, should be defined and followed for ephemeral and obsolete material. Access should not be compromised by changing technology or safe custody by business priorities.
9. *Context and framework.* Project objectives and design features that assist interpretation and evaluation should be recorded. A document should be put in context by indicating the standards followed, together with a note of any divergence, or else carry a full data description. It should be possible to identify the background theory and previous work on which a document depends. These indicate the knowledge required to understand it, and thus its comprehensibility for a particular user. Information assembled by information communities and editorial boards (M 2) should provide coherent frameworks that strengthen the structure of the knowledge base as a whole.
10. *Metadata*. Standards, and the metadata in the computer repository, are analogous to human semantic memory. They create guidelines for organizing the information, and are essential for retrieval and coordination. Widely accepted standards should be followed, so that information can be more readily and more widely shared to greater effect. Metadata should reflect the ideas of the geoscience community as a whole, appropriately controlled through committees that consult widely (L 5, L 6.1).
11. *Evaluation*. The results of evaluation should be generally available, supported by techniques such as quality assessment and branding. It should be possible to record different evaluations, and thus reflect changing opinions. Although older ideas should be retained, it is the fittest ideas that should survive and be the most obvious and accessible to the user.

3.3 Processes

Comprehensive procedures should be available for acquisition, storage, processing, delivery and presentation of information.

1. *Search techniques*. The system should simplify the process of identifying and reaching all recorded information relevant to the individual's needs, by organizing material within a clear browsable structure, and by offering comprehensive search procedures with indexes, summaries, keywords, spatial search, structured query language, and hyperlinks.
2. *Interaction*. The advantages of the computer's precise rules-based activities and the user's fuzzy but extensive background knowledge should be combined in interactive processing.
3. *Analogies*. Explanation by analogy relies on the human mind, with its background knowledge and capacity for inference and intuition. A lengthy learning process is involved, imprecision and ambiguity being the price of flexibility of thought. The computer should help the scientist to detect and explore a wide range of analogous situations under interactive control.
4. *Reconciliation*. The system should respond to the opinions and views of individual users, recognizing that these overlap extensively but seldom coincide. It should support negotiation to align and reconcile (but not obliterate) alternative versions.
5. *Representation*. Computer systems should make it possible to express conceptual models more fully, tied more clearly to the evidence on which they are based, in a form shared by the geoscience community as a whole. They should provide effective representation of geoscience knowledge through computer-based models and processes, such as visualization and statistical and spatial models.
6. *Tacit knowledge*. Processes should be available to communicate tacit knowledge by showing the learner how to do things by example, demonstration and practice. For example, annotated photographs and video clips can help to show the procedures and locations of observation at an outcrop, allowing the reader to repeat the original procedures, and confirm the results or otherwise.
7. *Abstraction*. Information should be available at different levels of detail or abstraction. Where possible, the process of abstraction should be automated, but in many cases will require human judgment and intervention. Standard levels of detail should be available to simplify comparison and integration with other datasets, as is current practice with maps at standard scales.

3.4 Business aspects

The geoscience information system should be relevant, profitable, and efficient in meeting the business needs. The business context of a project is relevant to the scientific interpretation and should be recorded.

1. *Reduced costs*. The need for paper publications and their management in numerous libraries should be reduced through client/server communication.
2. *Disintermediation*. Dependence on intermediaries and consequent delays should be reduced by computer support for word, data and image processing,

and for search and retrieval. The complex tasks of managing a store of scientific information should be eased by computer support and indexing.

3. *Delayering.* A directed flow of business and scientific information should support better decisions, simpler management structures and more efficient working. By making relevant information available to all concerned, conclusions should be reached more rapidly, layers of management can be eliminated, and groups and individuals empowered to make decisions within the constraints of the system.
4. *Project management.* The information system should be brought closer to business requirements by linking it to project management and control. Project management and business aspects should be closely linked to the scientific system, as they bear on the planning and execution of each investigation. Descriptions of past, present and proposed projects should be available to help users to interpret the results and be aware of current developments. Within an organization, the information system strategy should be incorporated into the broader business plan.
5. *Standards.* The system should discourage unnecessary barriers to communication by recognizing the value added through compatibility and adherence to standards.
6. *Outsourcing.* It should be possible to delegate some activities, such as information management, to a specialist organization. Assessment for quality, including adherence to standards, should ensure an efficient service to many users.
7. *Intellectual property rights.* The reward system relies on intellectual property rights, which should be protected. Access should be controlled if necessary by entitlement indexes and encryption.
8. *Incentives*. Participants should be motivated to drive forward all aspects of the system in a coordinated manner, by appropriate incentives and giving credit where it is due. Charging systems should be implemented where appropriate.

4. References

Blackmore, S., 1999. The Meme Machine. Oxford University Press, New York, 264pp.

Chamberlin, T.C., 1897. The method of multiple working hypotheses. Journal of Geology. Reprinted in 1995, Journal of Geology, 103, 349-354.

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

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.

Minsky, M. 1981. A Framework for Representing Knowledge. Reprinted, pp. 95-128 in Haugeland, J. (Ed) Mind Design, MIT Press, Cambridge, 368pp.

Rudwick, M.J.S., 1976. The emergence of a visual language for geological science 1760-1840. History of Science, 14, 149-195.

4.1 Internet references

POSC, 1997. POSC Specifications - Epicentre 2.2. Petrotechnical Open Software Corporation, Houston, Texas. http://www.posc.org/Epicentre.2_2/SpecViewer.html

<<<Back to Table of Contents

On to Part L: Adjusting the emerging information system to new technology>>>