<<<Back to Table of Contents
On to Part J: Human requirements that shape the evolving geoscience information system>>>

Geoscience after IT: Part I

A view of the conventional 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. A75-A85

Abstract - We need a strategy to cope with fundamental changes in our ways of working, based on a clear view of what we do and why. A systems view is essential to relate each part to the whole. This model of the geoscience information system should take into account: the need to separate data and process to enable reuse; the modes of thought of the geoscientist and how memory orders our thoughts; the shared geoscience record (knowledge base) and its interface with users; how ideas are linked as a network; how knowledge is organized; how a general overview can be maintained and linked to detail; how projects relate business objectives to the knowledge base and incentives keep the system alive.

Key Words - Systems view, thought modes, ideas network, business aspects.

1. A scheme of ideas

“It is a profoundly erroneous truism,” wrote Whitehead (1911), “that we should cultivate the habit of thinking about what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them.” But elsewhere (as quoted by Laszlo, 1972) he wrote: “in creative thought, common sense is a bad master. Its sole criterion for judgment is that the new ideas should look like the old ones.” and “. . the true method of philosophical construction is to frame a scheme of ideas, the best that one can, and unflinchingly to explore the interpretation of experience in terms of that scheme . . . all constructive thought, on the various topics of scientific interest, is dominated by some such scheme, unacknowledged, but no less influential in guiding the imagination. The importance of philosophy lies in its sustained effort to make such schemes explicit, and thereby capable of criticism and improvement” (Whitehead, 1929).

Kuhn (see part K section 1.2) resolves these apparent contradictions by distinguishing between “normal” science and revolutions in science. Normal science thrives on routine incremental additions to the body of knowledge, where the paradigm can be taken for granted. But revolutionary change calls for validation of ideas against basic principles. Our present concern is with changes to the supporting information technology that pervades all of geoscience. To grasp their significance, we need a broad view.

1.2 The need for a top-down view

Specialization helped the advance of science by overcoming the limited capacity of individuals to store and process information. By working in groups, each gains knowledge in depth of a specialized topic, complementing the knowledge of others. Difficulties arise, however, in communicating between specialist fields. Trapped in bubbles of specialization, we spin in our own eddies and lose track of the mainstream.

In day to day work a **bottom-up approach** is inevitable, concentrating on the detail and then fitting the pieces together. But rather than trying to assemble the jigsaw puzzle by building outwards from the pieces in your hand, it may be better to study the picture on the box lid. The system description should follow a **top-down approach**, seeking always the broad picture without which the detail might be misplaced. We therefore look at the behavior of the system as a whole and the structure and relationships of its specialized components. The objectives are to understand and control the massive changes stemming from the current advances in IT.

1.3 A glimpse of a broad panorama

We have looked in earlier sections at many aspects of IT. It offers many benefits. Information technology can help to deliver information efficiently in an appropriate form, where and when required. It can help to represent geoscience knowledge more comprehensively, linking ideas in a more fully connected network. It can help to manipulate the information more effectively with techniques to model, search, visualize, generalize and reconcile ideas.

We look next at the process of acquiring and recording knowledge. Geoscientists maintain in their minds a general model of the workings of geoscience and quite specific information about their own area of specialization. Each scientist has a unique view of the world, but in total, their knowledge overlaps to a large extent. The thought processes of the individual are modified by interactions within workgroups, and developed by training and study. They give rise to an explicit ordering of ideas as recorded knowledge.

We shall look further at the role of IT tools in supporting the thought processes of geoscientists. We usually take our thought processes for granted. However, by considering them explicitly, we may make better use of the tools and improve the information system. To understand the processes, we can draw on the work of brain specialists, on philosophers who study thought processes, and introspection of our own procedures. Change to one part can have unexpected consequences elsewhere (B 4), so we must study the system and its interrelated parts as a whole.

This leads to a view of future systems. IT networks connect many information sources, containing narrative text, spatial data and interpretations, structured databases, computer models, references to material and links to experts. This is the emerging global **hypermedia knowledge repository**, also known as **cyberspace**. It needs structures that guide contributors and users on where to put things and where to find them. Local structures are built and maintained by geoscience information communities and supported by shared metadata. Smaller projects can relate to these and remain in tune with global developments. The system must handle incentives (cash or kudos), because driving forces keep the system alive.

2. Systems

A **system** can be defined as a collection or set of interrelated and interacting objects or entities, including their relationships and behavior, which can usefully be studied as a whole. Early writers in this field, such as Beer (1967) and Laszlo (1972) stress the wide applicability of the systems approach. They note the importance of the **gestalt** principle - that the organized whole is more than the sum of its parts. More recent writers (such as Addis, 1985, Van Lehn 1991) have applied systems insights in specific fields, notably computer system design, and data and systems analysis. The object-oriented approach to design and analysis (H 5, J 2.4) owes much to this background.

The **information system** is where geoscience and IT meet. It deals with recorded knowledge, and its associated tools, activities and procedures. It is concerned, therefore, with how and why we collect information, process and record it, structure it, analyze it, draw conclusions, and make the results available to others. It is concerned with the information industry - not just with the work of computing specialists, such as analysts, programmers, designers, database managers, systems and network managers, but also with the more traditional work of scientists, surveyors, authors, editors, referees, publishers, librarians, archivists, booksellers, reviewers and readers. It is concerned not just with the tools of modern IT but equally with products of the older technologies.

2.1 Designing change

We might think of the conventional information system as being like a set of books, maps and reports arranged on a shelf and cataloged in a card index. With full IT support, we need a new model. It may resemble more an interwoven fabric of objects and processes in cyberspace. Such major changes mean that we must look at the architecture of the geoscience information system. **Architecture**, in this context, is defined as the structure of components, their interrelationships, and the principles and guidelines governing their design and evolution over time. We will look at the metadata of dictionaries and models that define terms and relationships; the hierarchies of object classes; the systems by which the information is managed and manipulated; and the frameworks in which information is organized to tell a coherent story.

An **information system strategy** deals with the change from existing to new ways of working. It might apply to an individual, workgroup, project or company. Regardless of the scope, it addresses three questions (CCTA, 1989):

- Where are we now?
- Where do we want to be?
- How do we get there?

The IS strategy would normally apply at corporate level, and lead to an implementation plan that concentrated on actions to introduce new IT developments. These might involve specification of new software and the design of data models using CASE tools (H 3). But our immediate concern is different. We are looking at the science as a whole, as a broad background for more detailed studies.

2.2 Subsystems, interfaces, models and metadata

For descriptive purposes, the system can be broken down into **subsystems**. We can think of each of these as a system of smaller extent, selected to include objects and processes that naturally belong together. Complexity of behavior should be incorporated within the subsystems as far as possible, thus simplifying the interfaces between them. The **interface** is the shared boundary between systems or parts of a system, or the means of interaction between them that makes joint operation possible. An interface device, for example, provides compatibility by enabling one item of equipment to communicate with another. An example may clarify the value of such an approach.

It causes few external problems if we replace an established self-contained procedure by a new one. Indeed, this can be a good way to gain initial familiarity with computing. For example, a geologist might decide to experiment with a computer technique for drawing graphic logs. Nothing beyond that single task need be affected. If, however, the routine preparation of graphic logs is to be automated, the data must be available to the computer, and a database is a possibility. Standards for recording data exist (L 4) and maybe the data could be incorporated in a shared archive, available also for drawing cross-sections and map making. Rather than storing the paper logs, the images might be recreated when required, in forms appropriate to specific users, perhaps geophysicists, geochemists or soil scientists, each with their own needs.

The self-contained task has become open-ended, hinting at broad new possibilities. Meanwhile cartographers in the same organization might be automating their map data on different principles. Solutions to small computer tasks can grow haphazardly in this way (and as described in B 4) from many different points, and because they tend eventually to overlap and conflict, the outcome is generally unsatisfactory. If a good **system model**, or description of the system, is available, then standard interfaces between the subsystems can be defined. Provided the standard interfaces are maintained, operations in one specialized subsystem can be adjusted without affecting others.

In describing the system, however, we should bear in mind another issue (K 1.2). Kent (1978, page 93) comments: “A model is a basic system of constructs used in describing reality. It reflects a person's deepest assumptions regarding the elementary essence of things. It may be called a ‘world view’. It provides the building blocks, the vocabulary that pervades all of a person's descriptions. . . A model is more than a passive medium for recording our view of reality. It shapes that view, and limits our perceptions. If a mind is committed to a certain model, then it will perform amazing feats of distortion to see things structured that way, and it will simply be blind to the things which don't fit that structure.” It is unwise therefore to adopt a specific world view without serious thought, or to take a fixed view of a changing world.

There are many ways of looking at systems, and many possible models of varying extent. Hence there is a need for a higher-level description of information that enables it to be understood outside the context in which it was collected. This is the **metadata**, the data about data mentioned in A 1. Metadata might contain definitions of relevant objects and refer to a model (known as a **data model**) indicating the relationships between them. Individual models are likely to refer to specific subsystems, dealing with topics such as geophysics, stratigraphy, spatial aspects, or business aspects. Sections L 5, L 6.1 look at some attempts to relate these to one another and thus define geoscience metadata as a whole. The initial step, however, is analysis of the geoscience information system.

2.3 Scope and components of the system

System boundaries are somewhat arbitrary, and so we must define the **scope** of the system, that is, its extent, what it consists of and how it works. We need to identify the components of the geoscience information system and their interactions, the participants and their roles, activities and driving forces. For present purposes, it is important to identify aspects of the system that are essential for carrying out its task, rather than those which reflect the limitations or historical development of technology.

Fig. 1 depicts the real world at the base and recorded information at the top, with scientists as individuals and in groups moving to and fro between them. Recorded information is shown as a triangle. At the base is the detailed raw data, collected by observation and measurement. An important scientific activity is reworking those data by generalization, interpretation, explanation and indexing. These operations progressively reduce the information to more concise forms, thus ensuring that they can be more widely shared. In doing so, they alter the original data, creating new information. The triangle in Fig. 1 narrows as the volume reduces, as a reminder of these essential processes of **abstraction**. The process transfers ideas from observation to explanation, from data to metadata.

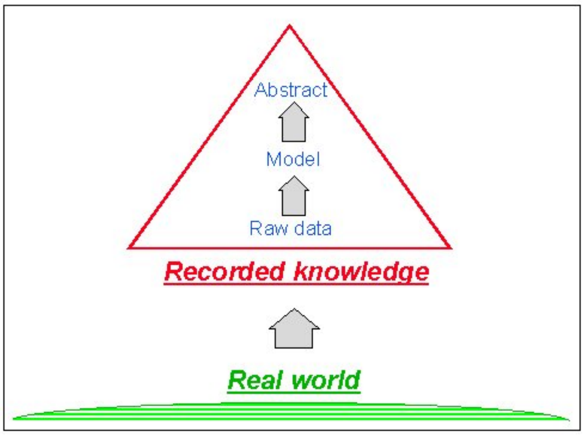

Fig. 1. Scientists collect and record information. From observations and measurements of objects in the real world, information is assembled, assessed and added to the store of recorded knowledge.

The real world as studied by geoscientists presumably exists independently of those who study it. It can therefore be treated as a system external to the information system. Scientists, who are thought by some to exist independently of their work, are again external systems. They interface with the real world through their investigations, and with the information system through their work in studying, improving and extending the information base. The **business environment**, and the **background theory** used to explain the data, much of it from other disciplines such as physics, chemistry, biology, mathematics and engineering, can also be regarded as separate systems. An attempt to show their relationships diagrammatically (Fig. 2), makes it clear that the geoscientists, not the information system, occupy the central position, and the interactions between the major systems are mediated by the scientists. The lasting record of the scientists' work is seen in **information repositories** - the vast body of recorded information in books, serials, reports and maps, as well as archived data, cores, samples and specimens. In addition, much unrecorded knowledge is held in the minds of individuals.

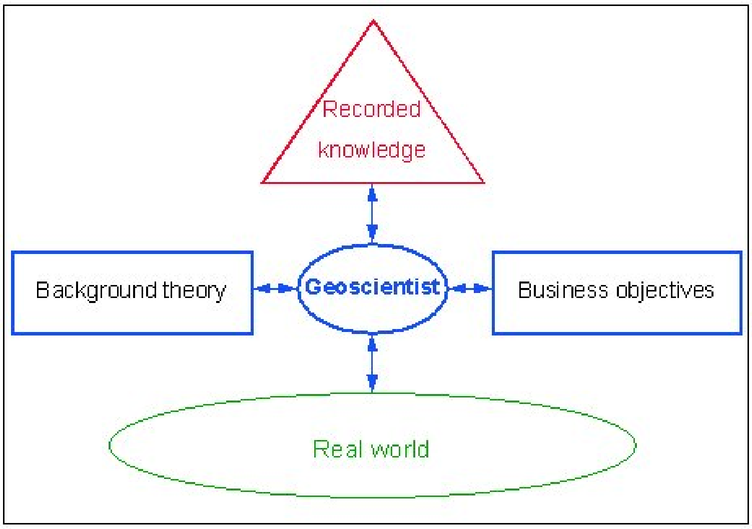

Fig. 2. Linking information to background theory and objectives. Geoscientists link the repositories of recorded geoscience knowledge and knowledge of other fields, business requirements, and their interactions with the real world.

Analysis of a system involves identifying components that can be studied separately in more detail, without losing sight of their interfaces and relationships. It soon becomes clear that many subdivisions, based on different criteria, are both possible and desirable. One possible basis for subdivision starts by separating the data (in the repositories) from the operations or **processes** that are applied to manage and manipulate them. The process of, say, depicting a number of features on a map, or of contouring a set of elevation values, is similar regardless of the area or dataset to which it is applied. When the processes are carried out manually, the distinction is between the data and the set of procedures and techniques. Where computer methods are used, the distinction is between the data and the program. An advantage of separating process and data is the prospect of **reuse**. A geologist who has learned the techniques of contouring can reuse them repeatedly in many different situations. Computer software can be prepared and maintained once, but used many times by many users. The same dataset can be analyzed by many different procedures. These benefits depend on clear, consistent interfaces. The means of linking separate data and processes are developed in the object-oriented approach (H 5). Following Kent (1978), this leads to a three-part division of the system: repositories, processes and their interface to the outside world.

The **user interface** (which can have more specific titles like machine interface, human-computer interface or HCI) is concerned with communication between the information system and the user (L 2). It includes output from the computer to screen or printer, and input to the computer from scanner, keyboard, mouse, or other device. With our broad definition of the information system, we must include traditional devices, such as pen and paper, typewriters and typesetters.

The user interface must give access to such functions as:

- input
- data management
- data manipulation
- modeling
- output

It should handle these in a consistent manner. Furthermore, as mentioned in parts C - H, it must deal consistently with activities involving varying degrees of collaboration:

- personal computing, individual thought processes and concepts developed and inherited from training and study
- interactions within a workgroup, preparation of project-centered documents and explicit metadata
- developing the main corpus of knowledge at the corporate and global level to advance the science as a whole, maintaining the canon, and sharing knowledge with known and unknown colleagues of past, present and future

3. A student looks at the real world

Moving on from general systems aspects, we focus first on the conventional (non-IT) geoscience information system and how it handles recorded knowledge. As we work through from this to IT-based systems, it may help to have in mind a geoscience investigation in which you were personally involved. My own choice of an example is close to my home.

Near the center of Edinburgh, Scotland, there is a tiny mountain, not 250m high, with the unusual name of Arthur's Seat. It is the remnant of a Carboniferous volcano. Generations of geology students have been brought here to meet the realities on which the science is based. From a viewpoint on Salisbury Crags, they can look across the grassy slopes of the Queen's Park, and see the smoke-blackened tenement houses of the Royal Mile, like spikes on a long tail stemming from a smaller volcanic crag to which the Castle clings. Beneath the slopes are sandstones, known from boreholes which once provided water to the breweries beside the Royal Mile. South of the Royal Mile is part of the ancient University. Here, intellectual battles between Neptunists and Plutonists once raged (Wyllie, 1999). The eventual paradigm shift, deciding in favor of igneous rocks cooling from a molten magma, rather than precipitating from a primeval ocean, presumably resulted in wholesale reconsideration of earlier observations.

From the viewpoint, the students can be led along a track at the base of a cliff of crystalline rock, examining the color and texture of freshly broken samples. They can observe the contact between this crystalline material and the underlying bedded sandstone, with its water-worn grains set in a silty matrix; can see the hardened, discolored sediment where it was baked in contact with the igneous rock, and observe the diminution of crystal size where the igneous rock cooled rapidly at the lower contact. Observations during a short walk along the base of the igneous body suggest that it lies parallel to the bedding of the sandstone. Then, vertigo permitting, a few observations at the top of the cliff reveal similar contact relationships in reverse. Having now identified the igneous intrusion as a sill, it can be positioned on a map, with measurements by compass and clinometer of its regional slope and that of the surrounding sediment.

The strange hollow at the base of the Crags unrelated to any apparent river erosion, and the crag and tail formation of the Castle Rock, can be explained to the student as glacial features, with an invitation to imagine the huge glaciers of the Ice Age grinding overhead on their way from the distant Highlands to the sea, gouging away the landscape, impeded here and there by harder volcanic rock. An ice-striated rock face is conveniently to hand. Now gather round and look at this map showing orientation and type of glacial features across central Scotland and observe our position in the regional picture.

The volcano itself is a little more difficult. The outcrops on Arthur's Seat show the now-familiar crystalline rock. The landforms suggest sheet-like lava flows, all with a regional dip to the east. Microscopic examination of thin sections back in the laboratory confirms that successive sheets have distinct petrographic characteristics which can be conveniently recorded on data sheets. Some careful mapping in that rough terrain and a coherent story emerges of a sequence of events (deposition, intrusion, volcanic activity, tilting, erosion, glaciation) consistent with the evidence and with what is known from the surroundings and elsewhere.

From the summit looking eastwards one can see the estuary of the River Forth opening out into the North Sea. A separate excursion will take us 60 kms east to the coast at Siccar Point, noting on the way the Old Red Sandstone of Devonian age, exposed along the coastal cliffs as thick near-horizontal beds. From the car park, we will stumble across the beach on rough vertical ridges of Lower Paleozoic gray shale and sandstone, etched by waves and partly obscured by sand and boulders brought in by last winter's storms. We come to a point where the vertical strata are plastered over by boulders and sand, as elsewhere but now consolidated. Following the consolidated beds we realize that we are looking at the local base of the Old Red Sandstone. The more imaginative students will perceive, as Hutton did on this very spot in 1788, that, had they been old enough, they could have stumbled across a rather similar beach in Devonian times. His visit is vividly recorded by Playfair (1805). As for the now-vertical Silurian strata, they too must once have been deposited as horizontal beds. Across Playfair’s giddy abyss of time, the present is seen as the key to the past.

From the cliff-tops we can look out across the North Sea, now criss-crossed by seismic tracks, the underlying strata penetrated by thousands of wells as it developed into a significant source of oil and gas. Its three-dimensional subsurface geology is known in great detail thanks to modern instrumentation, with digital records that can be displayed on the computer screen.

After these experiences, the students may be able to draw some tentative conclusions about how geoscientists work and think, some more obvious than others. Points relevant to IT will be discussed as they arise. First, we need to review the process of investigation just described.

4. How memory orders our thoughts

We observe much, remember little, and record less. We direct our attention towards information which clarifies the developing geological picture. The more relevant the material, the longer we remember it. We extract ideas from the geological picture to modify or reinforce our background knowledge. For example, observations of heaps of loose rock fragments at the base of the cliff might enter **short-term memory** (which enables us to remember small amounts of information very accurately for a brief period). But having avoided tripping over them, we soon forget them, just as we remember in detail the flow of words from the professor just long enough to interpret their significance. **Episodic memory** holds information much longer (though not always reliably) and allows us, in our mind's eye, to recreate sequences of past experiences and events. Although the episodes are typically an autobiographical view of our own experiences, we may similarly build in memory an internal narrative of processes, states and events in the geological past (the conceptual model of B 4.1). Thus the observations which enable us to build a picture of the intrusion of the sill might be passed on from short-term to episodic memory by tying them to events in its geological history. These can be linked in turn to the broader history of the volcano, the region, and more general aspects of the geoscience model. Recognizing that the rocks involved in the intrusive episode apparently followed well-established principles relating, say, crystal size to rate of cooling, confirms the value of ideas such as the present being the key to the past. This reinforces concepts in **semantic memory**, which deals with acquiring and using general knowledge, and with background understanding of what is true and what is significant.

This suggests that our brains have a built-in ability to abstract and generalize, building summaries that are remembered when the detailed observations are long forgotten. In activities involving short-term memory, thought may provide an instant response. At the other extreme, activities involving episodic and semantic memory may give rise to deep cogitation and reflection, with repeated review of the options, reaching conclusions only after a long period. The first may lead to entries in a field notebook, the second to publication of a considered opinion.

Psychologists and neurophysiologists have identified several levels of memory and have been able by experimental investigations to map them to distinct regions of the brain (Pinker, 1997). As well as those already mentioned, there are other levels relevant to our present purposes. **Procedural memory** involves learning motor and cognitive skills, sometimes executed subconsciously, as in trimming a hand specimen with a hammer or sketching in a notebook. **Spatial memory** refers to spatial pattern and the relative location of objects in space. These concepts have been carried across to studies in machine intelligence (in Brachman and Levesque, 1985) and to data analysis (Kent, 1978).

The different levels of memory are not wholly distinct, however. Recent research (see McCrone, 1997, 1999), which studies the working of the conscious brain with non-invasive techniques, emphasizes the intricate interconnections of the brain. It concludes that events influence, and their perception is influenced by, the state of the brain as a whole. The brain's responses to input are dynamic, non-linear and thus largely unpredictable, quite unlike those of the digital computer. Specialized regions of the brain do perform specific functions. But the brain was not designed component by component. It evolved as a whole in response to its environment over past generations and developed in response to events in its own lifetime. This fits the common-sense view that we should regard the computer not as an extension to our brains, but as a shared tool for managing and manipulating stored knowledge more effectively for the benefit of the human users.

It is clear that we must record information and knowledge if we wish to share them widely. We must store records to make them available in the future to others and to refresh our own uncertain memories. In these records, we can detect narrative text recording stories from episodic memory, data files and field notes that communicate with short-term memory, maps and diagrams that match our spatial memory, textbooks that feed our semantic memory. The immediate reasons for considering **conventional systems** (that is, those not designed with modern IT in mind) are to see where improvements are desirable, what knock-on effects they might cause, and how our legacy of recorded information can be carried forward into a new environment. The description, therefore, follows the system subdivisions mentioned in I 2.3 of interface, repository and process, to which business aspects are added.

5. Interfaces in a conventional system

5.1 Access

In conventional systems, users can interface directly with the knowledge base. This includes the formal and informal literature, in-house records and archives, computer files and the vast pool of knowledge held in scientists' minds. Access to recorded information may involve intermediaries, such as librarians, curators, record clerks or booksellers. These powerful figures are equivalent to the middleware (L 2) of a software system and influence what is visible to the user.

Conventional publications in geoscience assume that readers are familiar with current thinking. Using them effectively depends on the user knowing, or finding out, which books and serials deal with relevant topics at an appropriate level of generality and complexity. Unpublished knowledge may be passed on formally through managers and supervisors, and informally through a network of colleagues, by discussion, presentation and demonstration.

The interface is also involved when users contribute to the knowledge base. Informally, information can be passed on without delay. A typical published scientific paper, however, is prepared and edited with great care over a period of some months or years, and a similar time may elapse between submission of the completed paper and its final publication. Many copies of the paper are distributed and printed, although only a fraction of one per cent of the published copies may be read in any detail. Each contribution to the published record is identifiable, permanent and unchanging.

5.2 Connectivity

Knowledge has its source in the highly connected networks of the human mind, with many links that cannot be carried through to written records. Selections of the same material are therefore repeated in different forms for different readers, such as geophysicists, geologists, or engineers. Some connections, however, can be recorded. They include references to related papers for background, corroboration or detail; to specific points within a paper by page number or quotation; or to points on a map by geographical reference. Amendments and comments may be published later, perhaps unknown to readers.

Much information is repeated from earlier sources, probably reorganized and with added comments relating it to the specific project in hand. Each paper is like a patchwork of pieces drawn from many sources, cut to shape, augmented, and sewn together to produce something new and largely self-contained.

The reworking means that the author can repeat the story with changes, introducing a personal opinion. This gives room for local dissent within the world view, and thus for evolution of ideas. In the process, the provenance of ideas may be blurred, and the precise changes and their implications may be apparent to the reader only on making a detailed comparison with the earlier work. The information is organized and arranged to follow long-established conventions enforced by editors and publishers, but, within the house-style of the journal, each paper has a different background and distinct viewpoint. The apparent simplicity thus conceals the vast labor, perhaps the greater part of the project, spent on searching and evaluating earlier work and either reconciling it with new ideas and observations (K 1.4) or ignoring or contradicting it.

6. Conventional repositories

When an author records information for use by the geoscience community, there is a need:

- to ensure that it meet standards which enable users to understand it
- to ensure that it is connected with other information on which it depends
- to store it safely in suitable repositories
- to identify it for reference purposes
- to catalog it for subsequent identification and retrieval

International standards define the codes and procedures for identifying each contribution, and cataloging rules which provide metadata for retrieval by author or topic (H 2). The system must be able to attract and accept contributions, evaluate them, ensure that intellectual property rights are upheld, and reward all concerned (I 8.2).

Unpublished work may be restricted to a local file. Data records are generally held by the originators, or the commissioning group, or in archives established by geological societies or other organizations. International sharing of data on, say, global seismology, geomagnetism or aspects of oceanography, relies on networks, with collaborating groups exchanging computer files to agreed standards (L 5). Cores, samples and specimens are stored in museums or in company or specialist archives, such as state-funded core stores.

6.1 Repetition

A striking feature of the geoscience knowledge base is the extraordinary amount of **redundancy**, that is, repetition of the same information. Scientists all undergo lengthy training to develop a shared understanding. The courses necessarily cover much of the same ground, and many of the differences may be due to the fact that they are taught, not by the finest teacher of that topic, but by the one who happens to be in post at that particular place. Understanding is instilled through numerous examples and expositions, and it may be left to students to arrive at their own general conclusions or world views. Whether this is an important part of their training or merely reflects the bottom-up view of the instructors no doubt depends on the circumstances. The wide availability of textbooks does much to ensure that the best ideas are available for sharing. In due course, computer-aided instruction and distance learning (Butler, 1966) will no doubt also improve standards by making the best teaching methods more widely available.

Published information is stored in hundreds, if not thousands, of identical copies distributed through many general, specialist and personal libraries, and may be independently cataloged in many of them. Repetition is also a feature of the detail, such as descriptions of vertical sections or well samples, where ditto marks are ubiquitous. Much basic information is laboriously copied from informal reports to include in published papers which repeat part of what has already appeared in other publications. The material used for teaching is likely to have been reworked many times, perhaps from an original research report to part of a regional assessment to an account of a new method to a paragraph in a textbook. Text accounts overlap with information shown on a map and vice versa. The reader who consults several sources is likely to encounter the same information in many subtly different forms.

Redundancy is inevitable, and much of it is probably desirable. Scientists, for example, could not communicate without shared understanding based on similar training. Measurements may be repeated to check their validity. Information is clarified and reinforced by repetition, which is why you may reread the same passage several times. The reader will be influenced by frequently encountering similar ideas, and repetition may therefore be a means of indicating what is important. Undesirable features of redundancy are that it increases costs and makes change difficult to control. A significant part of copying information is unnecessary and unhelpful and might well be reduced by appropriate information technology. Hypertext can link an item to any relevant context. Explicit evaluation could avoid the need to repeat for emphasis. A long-term goal may be to create an IT knowledge base in which the same recorded information can be filtered and presented in different ways for many purposes from, say, teaching to risk assessment.

6.2 Organization of content

The literature deals with entities of interest, which might be referred to as objects (J 2.4), although they are seldom identified as such and their definitions may differ subtly between authors. The object classes, such as stratigraphic units, localities, fossil or rock types, may be linked implicitly or explicitly to published definitions, or may be defined within the documents which use them. Lexicons and dictionaries provide metadata to standardize the vocabulary. Textbooks and monographs record standard procedures, methods and classifications. The relationships among objects are unlikely to be formalized in a data model, but instead are developed in a text narrative.

A geological document is concerned with more than one narrative thread of events. It weaves together a set of stories concerning various aspects of the geology (J 1.1). An account of the reasons for undertaking the investigation and of the methods and instrumentation might be woven in with separate threads giving conclusions about the geological history of the area, aspects of the stratigraphy, petrography, structural and economic geology, and so on, all related to one another in a single coherent document.

Text narrative is generally concerned with describing observations, sequences of actions during investigation, and the processes of explanation, such as a chain of cause and effect. Diagrams and photographs provide some local spatial context, but the main spatial framework may be handled as a map, possibly separated from the text. Spatial information locates observations and interpretations and describes form and shape, spatial patterns and relationships, and movements in space and time. In geoscience, the narrative accounts are also likely to refer to spatial objects in a spatial environment. Narrative and spatial information, although analyzed in distinct areas of the brain, must therefore be closely linked. In their presentation, however, aspects may be separated because information types can be difficult to combine in one publication. Maps are most useful in large format, so that the user can see the overall pattern, can trace variation or search for detail, while aware of the context through peripheral vision. They are therefore published separately, and may be produced independently of text documents for the same area. Tacit knowledge is passed on interactively through training, discussion and demonstration. It may be reinforced by informal papers, but by definition cannot be part of the conventional literature.

7. Processes in the conventional system

7.1 Generalization

As specialists, we may wish to study parts of a paper in detail. For the rest we require only an abstract or summary of salient points. It is therefore helpful to have an explanatory title, a table of contents, abstract and index that provide a quick overview and an easy route to topics of interest. Also, it is only a small number of papers that interest any one reader, and they are easily missed. A significant part of the literature is concerned with reviewing and summarizing earlier work from a variety of viewpoints. This increases the amount of redundant information, but by reading the review articles we are less likely to miss significant new ideas, regardless of where they were first published. Similarly, maps are available at various scales, so that we can maintain a general overview, as well as finding detail for areas of particular interest.

A geological survey, for example, may provide illustrated written accounts at several levels of detail, from a regional guide covering a large area to an archived description of a microfossil. It may also provide maps at different scales with different levels of discrimination and resolution, from a postcard map of the entire country to a field sketch of a single outcrop. Although geological processes differ at different scales, detail and summary are inevitably interrelated. Abstraction and generalization apply to all these aspects and to all information types. By reducing the volume of information, they make it more widely available beyond its own specialized area.

The scientific literature sifts ideas from earlier papers by reference and quotation. As readers tend to concentrate on recent work, the effect is constantly to revive the more important ideas. They are implicitly placed, amended if need be, within the current paradigm. Other, probably less significant, ideas are allowed to sink into the sludge of old, unread papers. Some useful concepts, not recognized as relevant in their time, could also disappear without trace. However, the metadata (or metainformation) created by librarians, curators and catalogers may save them from oblivion. The metadata have a number of functions:

- to provide a structure or framework of topics to organize the contents of the information system
- thereby, to make it easier to find specific information
- to define terminology and describe methods and procedures, thus clarifying communication.

The classifications, definitions and the rules for applying them are generally agreed by committees of experts, defined by international standards, and applied by librarians or imposed by editors. The results are in dictionaries, lexicons, library catalogs, and the placement of books on library shelves (H 2). Specialist journals reflect, and may help to define, topic-based subsystems of the information system.

8. Business aspects

8.1 Projects

Contributions to the knowledge base generally stem from projects, and can be fully understood only in the light of their procedures. A **project** is a manageable activity which has objectives, resources, and structure. It may involve one individual, carrying out, say, a site investigation or an academic research study, or it may have a multimillion dollar budget and involve collaboration among many institutes. Each geoscience project has its own objectives, determined by its business setting, and therefore is unique in its methods and information content. Perhaps because of repetition, the conventional documents needed to provide background geoscience information for undertaking a project can be surprisingly few - a small number of maps and reports, perhaps access to a small part of an archive, and the use of a handful of well-known textbooks. The results of a project may be recorded in their turn as scientific papers. A paper generally refers to a single project, although several papers may deal with different aspects of the same project. Some details of how the investigation was designed, reflecting the project and its business setting, are generally necessary to interpret the results correctly, and are therefore described in the paper.

In designing an investigation, there is a trade-off between meeting specific project needs efficiently, and generating information that can be widely shared. Conventional procedures do not impose a solution, but do respond to market forces. For instance, a speculative seismic survey may be extended to cover a wider area, in the hope of selling the results to more oil companies. The scope of an academic study may be deliberately extended to reach a wider audience, or to make publication easier.

There is significant overlap in the basic information required in many projects. As a wide variety of investigations rely on the same basic information, it can be more efficient to collect it once in a single wide-ranging survey, than repeatedly whenever it is required. Organizations such as geological surveys and shared repositories are supported for this purpose, often with state funding. Information from a geological survey, a museum archive or in-house company records, is more highly structured, but less flexible, than the general scientific literature. It attempts to provide one coherent and consistent picture through the close collaboration of many workers. Individual topics, such as limestone resources or the findings of an aeromagnetic survey, may be the subject of separate reports, maps or series, or may be included as sections within accounts of specific areas, or both. All should tie together as aspects of a shared view - a cohesive but limited part of the literature. They constitute the results of a single, large project, and are the base data on which much else is erected (M 2.3).

8.2 Driving forces

The importance of business aspects permeates the information system (M 1, M 3). The development of specialized journals may be not unrelated to the identification by commercial publishers of two driving forces: the anxiety of authors to publish papers, even to the extent of paying for the privilege; and the need for librarians to purchase the result, with little sensitivity to the price. In contrast, some worthy computer projects have neglected the potential driving forces, and failed through not anticipating the passive hostility of those who were expected to contribute but not to benefit (Peuquet and Bacastow, 1991).

Authors are among the more highly motivated participants in the information system. Their career prospects are strongly correlated with the ability to produce a stream of valuable papers. The value of unpublished reports may be judged by a manager, with or without help. The manager, who is likely to have asked for the report for a specific purpose, is in a good position to judge the result. Publications are judged by their volume and by the prestige of the journals in which they appear, reflecting decisions by the editor and referees employed by that journal. The extent to which other authors cite the papers may also be taken into account, and with books, the comments of reviewers may carry some weight.

Likely motives of publishers, booksellers and editors are to earn an honest living and prosper, influenced or even dominated by enthusiasm for the science. The enthusiasm is likely to be shared by the referee and reviewer, possibly reinforced by a desire to remain in the network. Catalogers, curators and librarians are normally paid for their work, although this by no means rules out enthusiasm. They influence what is available and who sees it. Readers are presumably the most important participants, being the ultimate beneficiaries of the system’s existence. Their motives are presumably to gain information for a variety of reasons, including curiosity, which reflect their business concerns. In the circumstances, their role is a surprisingly passive one. They seem to have little direct influence on the evaluation, operating instead through intermediaries or in their other roles such as authors or referees. The cost of the system is largely paid, not directly by the readers, but by government or institutional support for libraries. Publication costs may be recovered, but seldom the cost of the underlying research, which is paid for by other means.

In due course, these tasks will be handled differently. Any new structure must however incorporate the legacy of earlier information and work practices. The alterations and extensions which new technology is introducing need a solid foundation in the conventional knowledge base.

9. References

Addis, T. R., 1985. Designing Knowledge-based Systems. Kogan Page Ltd., London, 322pp.

Beer, S., 1967. Cybernetics and Management, 2nd edn. English Universities Press, London, 240pp.

Brachman, R.J. and Levesque, H.J. (Eds.), 1985. Readings in Knowledge Representation. Kaufmann, Los Altos, 571pp.

CCTA, 1989. The Information Systems Guides. John Wiley & Sons, Chichester.

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

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

McCrone, J., 1997. Wild minds. New Scientist, 156(2112), 26-30.

Peuquet, D.J., Bacastow, T., 1991. Organizational issues in the development of Geographical Information Systems: a case study of U.S. Army topographic information automation. International Journal of Geographical Information Systems, 5(3), 303-319.

Pinker, S., 1997. How the Mind Works. Norton, New York, 660pp.

Playfair, J., 1805. Biographical account of the late Dr James Hutton, F.R.S.Edin. Transactions of the Royal Society of Edinburgh, Vol. V.-P.III. Reprinted 1997, in James Hutton & Joseph Black. RSE Scotland Foundation, Edinburgh, Scotland.

Van Lehn, K. (editor), 1991. Architecture for Intelligence - the 22nd Carnegie Mellon Symposium on Cognition. Lawrence Erlbaum Associates, Hillsdale, NJ.

Whitehead, A.N., 1911. An Introduction to Mathematics. Thornton Butterworth, London, 256pp.

Whitehead, A.N., 1929. Process and Reality. Cambridge University Press, 429pp. Reprinted 1969, Free Press, New York.

Wyllie, P.J., 1999. Hot little crucibles are pressured to reveal and calibrate igneous processes. In: Craig, G.Y., Hull, J.H. (Eds.) James Hutton - Present and Future. Geological Society, London, Special Publications, 150, pp. 37-57.

9.1 Internet references

Butler, J.C., 1996. Another node on the Internet for those with interests in geosciences, mathematics and computing. http://www.uh.edu/~jbutler/anon/anon.html

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

Disclaimer: The views expressed by the author are not necessarily those of the British Geological Survey or any other organization. I thank those providing examples, but should point out that the mention of proprietary products does not imply a recommendation or endorsement of the product.

<<<Back to Table of Contents

On to Part J: Human requirements that shape the evolving geoscience information system>>>