by J. Gravesteyn and G.N. Rassam

International cooperation in geoscience information: A global challenge

The open flow of information is critical to the producers, distributors, and users of geoscience information. The International Union of Geological Sciences' Commission on Geological Documentation has made great progress in stimulating cooperative programs, in contributing to the management of geoscience information, and in lessening the barriers to the open flow of information. The constraints and barriers to the flow of geoscience information include its small part of online distribution, variations in technology worldwide, language and cultural identity or integrity, national sovereignty and security, copyright protection, and the lack of integration of factual and bibliographic information.

International cooperation among the producers of data bases has resulted in the publication of a multilingual thesaurus, a directory of geoscience data bases, and global information exchange. More needs to be done, however, and the authors call for the Commission on Geological Documentation and the Commission on Storage, Automatic Processing, and Retrieval of Geological Data to move toward the creation of a single Commission that will be concerned with the exchange of information and the integration of factual and bibliographic data on a worldwide cooperative basis. (Ed.)

Introduction

The international exchange of information in scientific disciplines is as old as science itself. Direct personal contacts, circulation of publications, and sharing of experience have contributed, through the ages, to a better understanding of scientific phenomena and processes, and cooperative efforts were often at the origin of major discoveries.

The development of information technology has facilitated a faster exchange of information and easier access to better information media. The International Union of Geological Sciences' (IUGS') Commission on Geological Documentation (COGEODOC) has worked to stimulate cooperative programs, to contribute to working standards in geoscience information management, and to fight against barriers that impede the free flow of information.

The players on the geoscience information scene

To use a cinematic metaphor, we will look at the total "scene" of global geoscience information.

Producers

Geoscience information is produced by individual or collective research and field work that are conducted by geological surveys, universities, research institutions, and industry. In all cases, geoscientists are the producers of information.

Not all of the information produced is publicly available. Few figures exist on privately owned versus publicly available geoscience information, but for the earth sciences, a working hypothesis of 50 percent private and 50 percent public seems reasonable. This means statistically that the geoscientist has access to only one-half of the existing information or about 170,000 articles produced each year. For pure research topics, the proportion of publicly available information is probably better, but in mineral and oil exploration, it is much less. For factual and numerical data, public access is even more limited because of proprietary considerations and because of space considerations in books and journals. These figures also mean that, when we are talking about cooperation and global information needs, we are excluding more than one-half of the total volume of information that is produced in the world.

International cooperation is technically easier and economically more cost effective for bibliographic information than it is for factual data exchange. However, the increasing importance of integrated surveys, cartographic applications, and expert systems makes international coordination and cooperation even more imperative and the establishment of working standards and bridging mechanisms even more essential for factual data exchange in science.

Directors

The directors, or distributors, of information are mostly independent from the discipline-oriented producers of information. The publishers transform the ideas expressed in written form into a printed product that is available for acquisition. The data-base distribution services transform the data into their format, adapt the information to their search systems, and connect the data base to national or international transmission networks. Actually, the distribution services play an important role in the standardization of data elements and formats. On the other hand, they oblige the producer to accept the search system as it is, and this kind of standard searching can hinder the efficiency of the data base.

The public distribution of data bases involves both bibliographic and factual information. This is particularly true for financial and business data. In geoscience, however, the public distribution of factual data sets through online services is very limited because of the small demand. Use of the data sets is limited to the institutions or companies that produced the data or to their partners.

In contrast to factual data sets, international bibliographic data bases, such as GeoRef and PASCAL-GEODE, are installed on one or more online hosts and can be accessed by those who are connected to the transmission networks. In many cases, especially in countries that have online distribution, the role of the distributor is played by the national geological survey or by a university through offline services or manual files. On a global scale, offline and manual services are still the rule and online searching exists only for privileged customers.

Spectators

The spectators, or users, are both king and victim. The user is king because he determines the success of an information service. If the user is satisfied, the product will survive and new services can be created; if he is disappointed, the data base or information service will disappear or transform itself into a living fossil. The user is also the victim. He wants complete information on a research topic, or he is looking for all of the existing papers and source data that are available for a certain geographic area. This information includes not only bibliographic material but also borehole sections, samples, analyses, modeling results, and maps. He knows and accepts that he cannot obtain everything, especially if earlier studies have been conducted by private organizations. However, the user is frustrated and society loses money when the missing information is the result of incomplete bibliographies, lack of information on existing boreholes, or the loss of a report from the library.

International cooperation is not a panacea, but global coverage of the world literature and the integration of factual and bibliographic information into one search system together will eliminate some of the reasons for user frustration. Also, the producers and distributors of information must be aware that not all of the users of geoscience information are geoscientists. Because of this, too high a degree of sophistication in search techniques or presentation will have a very negative impact on the use of geoscience information outside the profession.

Intermediary organizations, such as libraries and information brokers, also play an important role in the chain of information transfer. However, their international problems and challenges are not considered in this paper.

Constraints and barriers to the flow of geoscience information

The transfer of information across boundaries is not a smooth process. Impediments are many, and they vary in scale and importance.

Some of the barriers apply to all kinds of information (see table 1), and some are specific to the geosciences because of the political sensitivity of resource data.

International organizations such as COGEODOC tend to believe in the ideal of free flow of information. It is useful, therefore, to examine the existing barriers and the actions that might alleviate them.

Economy of scale

In 1990 the total cost of online information services probably will reach over US\$10 billion. And of the total information market, 80 percent will relate to the financial business sector and only 13 percent will relate to science and technology.

The total number of publicly available online data bases is now over 4,000. They are produced by more than 2,000 information services and are distributed by approximately 600 vendor organizations. All of these figures have been multiplied by a factor of 10 since 1979, which illustrates the rapid development of the market.

Geoscience information occupies an extremely small part of online distribution: less than 2 percent of the scientific sector and not even 0.1 percent of the total online industry. This is, in geochemical language, a few parts per billion of the information universe. The cumulated connect time for all geoscience data bases produced by vendor organizations in the United States and Europe probably does not exceed 1,000–1,200 hours a month. Obviously, in these circumstances, an online geoscience data base cannot recover its production costs solely from its online income.

Until recently, the bibliographic data-base producers could balance their accounts with the sale of printed bibliographies, but these products lose their attraction for the profession when online services (or offline services) are available. Therefore, the only solution to this budgetary squeeze, while maintaining quality, is either to increase subsidies or to decrease production costs. Clearly, increasing subsidies is not very popular in the present economic situation and under budget restrictions.

International cooperation has proved successful in cutting down production costs. Two different approaches have demonstrated this. The first approach was the creation of a worldwide network of individual professionals who contribute to a centralized data base. The other approach, promoted by COGEODOC in the 1970s, was the creation and maintenance of an international network of information centers. Each center is responsible for the geoscience information published in a certain geographic sector, and this both eliminates overlap in work and improves coverage.

The network of information centers started in 1968 with a French-German and a French-Czechoslovakian cooperation agreement. It now links geoscience information centers in the United

Table 1.—Barriers to geoscience information flow: problem areas

Information chain	International cooperation	
	North-north cooperation (including west-west and west-east cooperation)	North-south cooperation (including south-south cooperation)
Producers	1. Economy of scale 2. Language: cultural integrity, communication 3. Sovereignty: national security, protection of industry 4. Standardization	1. Human resources: manpower, know-how 2. Financial resources: equipment, personnel 3. Information technology: communication
Distributors	1. Integration of factual and bibliographic data 2. Search technique 3. Online access	1. Communication technology: access 2. Resources: human and financial
Users	1. Access to integrated information 2. Information on data bases 3. Language	1. Communication technology 2. Training 3. Language

Table 2.—International cooperation: actions and impact

Information chain	International cooperation	
	North-north cooperation	North-south cooperation
Producers.....	Bibliographic networks Documentation standards Automatic translation Multilingual thesaurus Integration of factual and bibliographic data	Bibliographic network (PANGIS) Coordination of factual and bibliographic data bases Multilingual thesaurus Training
Users.....	Common search strategy Multilingual access Information on existing data bases Data base on research projects	Common search strategy Multilingual access Training Information on existing data bases Data base on research projects

States and Europe (Czechoslovakia, Finland, France, Germany, Hungary, Italy, Poland, Romania, and Spain). The data bases produced by the partners through an exchange program are updated in accordance with the technical economic principles defined in several bilateral agreements.

Technology

The fast development of information technology since the 1960s has resolved many of the problems that were related to the free flow of information. Technology is, in itself, certainly not a barrier to the international exchange of information. However, the differences in various parts of the world in the development stages and the use of computer and communications technology provide a serious handicap in data flow across borders and international cooperation on a global scale.

In the United States and some other Western nations, society is evolving into a real "informatics culture." The earth scientist has immediate access to a mainframe or a microcomputer. When he has a problem, he can call for assistance within his direct professional environment. In other industrial countries, the "information culture" has yet to invade society to the same extent, but the available information technology can satisfy most needs.

The situation is completely different in the developing countries where many geologists do not have access to any computer equipment. Online searching is still a mirage, and participation in an operational data-exchange program is a dream. There is a dramatic need for "north-south" cooperation (table 2).

Language

The language barrier is a problem in itself because of the direct links between the language and the cultural identity of the individual scientist or the cultural integrity of a linguistic community. Problems associated with cultural integrity certainly exist, but the maintenance of the language as a tool for communication and comprehension between scientists who belong to the same linguistic and cultural environment is a right and a pragmatic need.

The language barrier problem was illustrated by a study in 1989 of the European Community on the use of domestic and foreign online data bases. Of the total users, 80 percent were searching data bases in foreign languages, and nearly 60 percent considered that the language itself was not an essential factor in the choice of the host service. However, more than 50 percent considered language to be a key issue in the choice of the data base. This seems to indicate that there is not a major problem in handling a search strategy in a foreign language. However, the use of the mother tongue is felt to be important in the formulation of the query.

How do we protect language integrity intelligently in a multilingual and international environment? International linguistic cooperation for the development of multilingual information tools is one way that was chosen by COGEOCOD. The first edition of the "Multilingual Thesaurus of Geosciences" is now well known, and a second edition is being prepared, at least in a computer-readable form. The thesaurus is being used in the international United States-Europe cooperation network and in the Pan-African Network of Geological Information Systems (PANGIS). Another outfall of international linguistic cooperation is the creation of bilingual data bases. The United States-Europe cooperation network made it possible to translate automatically all geologic terms used formally to index the literature (descriptors) and the geologic terms used informally (nondescriptors) and to implement bilingual data bases in Europe: French-English for PASCAL-GEODE and German-English for Geoline.

Certainly more can be done in the sector of linguistic barriers, as, for example, the publication of user guides in several languages or multilingual information on available online data bases. This process is more of an economic problem than it is an informational problem. The increased use of geoscience data bases probably will stimulate the information providers to assist their users in this area.

National sovereignty and security

Obviously, each country is sovereign and each nation wishes to protect itself against technical and economic plunder. Problems and barriers arise when these legitimate rights are misused and prevent the export or import of publicly available data, or when the national information industry is protected exceedingly through subsidies, national equipment-acquisition policies, and fiscal measures. National and international information bodies must be vigilant and must react firmly in extreme cases by using market mechanisms to defeat such attempts.

Restrictive or protective measures concerning data flow across borders affect especially defense, financial, business, and engineering information. Fortunately, the geosciences are not much involved. However, those who have tried to gather aerial photographs for geological surveys in strategic areas or to retrieve mineral exploration reports for followup operations know how restrictive national information policies can be.

Copyright

The issue of copyright has become urgent over the last 20 years. Just as in life people become more reasonable with age, it seems that, since the birth of information technology, society has needed some time to find a new equilibrium in the chain of information transfer. The dialog among publishers, secondary information providers,

librarians, and legal advisers, together with some court decisions on both sides of the Atlantic Ocean, has led to a "consensus" on some basic principles:

- Protection of texts against copying and exploitation in textual data bases
- Freedom of indexing and reasonable abstracting of articles for inclusion in bibliographic data bases
- Protection of data bases as a whole
- Limited free use of bibliographic references

In this area, COGEODOC is representing IUGS in the Working Group on Legal Aspects of Information Transfer and is operating under the umbrella of the International Council for Scientific and Technical Information (ICSTI).

Factual-bibliographic interface

At the GEOINFO III meeting in Adelaide, Australia, in 1986, COGEODOC made a plea for better integration of factual and bibliographic information in order to respond to the stronger demand for the complete and user-friendly supply of information.

Since the Adelaide meeting, we have not observed much operational progress in this field. More sophisticated software packages have made it possible to resolve local problems. Positive results also can be reported from some European countries concerning the online distribution of drilling location records and areal bibliographic data. However, the use of Geographic Information Systems (GIS) in publicly available bibliographic data bases is still very limited. On the international scene, IUGS made the decision to create a special Sub-commission on Global Data Management and Information Systems (SGDMIS) under the Commission on Storage, Automatic Processing, and Retrieval of Geological Data (COGEODATA). SGDMIS is to work as a link between the Union and national geological surveys in developing an appropriate Global Geographic Information System.

The COGEODOC program of action

COGEODOC's overall objectives are (1) to promote international cooperation among data-base producers, (2) to facilitate access to more complete and integrated geoscience information, and (3) to contribute to the elimination of barriers in both the information chain and data flow across borders.

The COGEODOC program of action is conducted along three axes: (1) to improve compatibility among indexing, coding, and searching techniques on a worldwide scale; (2) to produce working standards, directories, and manuals for international geoscience data handling, exchange, and use; and (3) to promote more extensive use of geoscience information in the Third World.

International cooperation among data-base producers

The cooperative effort that started in 1968 with French-Czechoslovakian and French-German agreements has proved quite successful. The network now groups 10 countries, and it operates in a linked structure that has 2 focal points, one in the United States and the other in France.

The volume and procedures of exchange vary from one bilateral agreement to another, but all the partners use common indexing techniques and practically the same vocabulary. This use of a common indexing practice not only saves time and money for the producers but also enables the user to switch from one data base to another without worrying about different indexing philosophies or classification schemes.

This type of network is open for further extension, and one could dream about a cost-effective worldwide network to which each nation or group of nations could contribute its part of the literature in exchange for the complete world geoscience data base. A possible next step could be a better integration between the United States and European network and GEOINFORM, the cooperation network of the Council for Mutual Economic Assistance (CMEA) countries (Eastern Europe, Mongolia, Vietnam, and Cuba). COGEODOC is ready to act as a platform for discussion about further international cooperation.

Production of working tools

The major event since the last GEOINFO meeting was the publication in 1988 by Pergamon Press of the first edition of the "Multilingual Thesaurus of Geosciences" (MT). We expect to report at the next GEOINFO in Prague, Czechoslovakia, that the updated data base of the MT's second edition is ready.

At present we are reviewing the fields of systematics in mineralogy; igneous, metamorphic, and sedimentary rocks; and paleontology and stratigraphy. In 1991 and 1992 we will deal with the introduction of the missing terms of general geoscience concepts outside the systematic fields. From a statistical analysis on the frequency of use of MT and non-MT terms in the PASCAL-GEODE data base, we can estimate the volume of missing terms. For the systematic fields, the volume of new indexing terms (mostly nondescriptors) to be introduced is important: there are more than 2,000 terms. For the other conceptual indexing terms that frequently are used in the data base and are missing in the MT, the updating will add approximately 650 terms. Some 350 terms are to be added to describe methods, general concepts, and equipment. Additional terms will address geographic and geologic-geographic concepts.

Another type of updating of the MT concerns the extension of the MT to other languages. The printed first edition included English, French, German, Italian, Russian, and Spanish. The MT data base, managed by the Consiglio Nazionale di Ricerche in Milan, Italy, contains also Finnish and Czechoslovakian versions. Our Brazilian colleagues are starting translation of the MT into Portuguese. We will welcome any initiative to broaden the linguistic scope of the MT to other languages, such as Arabic, Chinese, or Japanese.

A directory of geoscience data bases was published in 1989 by COGEODOC and COGEODATA through their joint working group. They plan to update this directory regularly. Also, automatic translation of geoscience abstracts is being studied by VIEMS, the Soviet Institute for Mineral Deposits Research. Plans exist for this to become a common COGEODOC-GEOINFORM project.

COGEODOC also has been asked to take the responsibility for implementing a data base on current international research projects. For financial reasons, the start of this project has been delayed. At present, information on international geoscience research projects is scattered and poorly distributed, although the most important program, the International Geological Correlation Programme (IGCP), is rather well known among geologists and information specialists.

Promotion of the use of geoscience information in Third World countries is not an easy task, in spite of the importance of natural resources to the economy of these countries. Efficient management of geologic field data and exploration reports is a key issue in the discovery and management of natural resources. Priorities may vary from one country to another. In some cases, mineral resources are essential to the economic and social situation in a country. In other cases, the management of water resources is vital for agriculture and industry. In all cases, the integration of basic geologic field data and applied exploration surveys is a key issue for development, as is the efficient exploitation of this information. A good example of

pragmatic cooperation programs is the African PANGIS project. PANGIS was launched in 1986 by the International Center for Training and Exchanges in the Geosciences (CIFEG) in Paris, France, with the support of the United Nations Educational, Scientific, and Cultural Organization (Unesco).

In the first stage, PANGIS has been concentrating its efforts on bibliographic material. Documentation centers are being set up in eight African countries. COGEODOC is concerned directly or indirectly in two ways: the multilingual thesaurus is being used for indexing by all partners involved in PANGIS, and the Commission organized, with the support of Unesco, an expert consultation meeting on the expansion of PANGIS toward factual data management. During this expert meeting, held in March 1990 in Paris, representatives of several geological surveys from Africa, North America, and Europe discussed the feasibility of creating factual data bases in Africa, in accordance with the needs and priorities of the PANGIS countries. This second stage of development of factual data bases is to be followed by a third stage where specific applications, such as cartographic output, statistical approaches, or modeling, will be implemented.

During the March meeting, the technical specifications were defined for the mineral deposits data base, the field and borehole data base, and the geochemical exploration data base. Data-base elements, worksheets, data-base infrastructure, software packages, and hardware and human resources requirements have been defined as well. We expect that the first experiments will start at the end of 1990.

International cooperation and global science

The growing interest in global science and global environmental change is no longer purely a scientific concern. Environmental issues like the effects of the Chernobyl disaster, greenhouse effects, and the destruction of the ozone layer, as well as the drama they have caused, are making global environmental problems a major concern for citizens and politicians around the world. This adds a social dimension to these problems. In some ways, global change touches each and every world citizen.

The International Geosphere/Biosphere Programme (IGBP), sponsored by the International Council of Scientific Unions (ICSU) and many other organizations, started in 1986. It is the scientific reaction to the "panic in the house" that occurred after recent anthropogenic disasters. Most of the projects are oriented to the biosphere and concern interaction between the atmosphere and the terrestrial or marine biosphere. The only accepted geoscience project in IGBP is "Global Change in the Past," which involves specialists in glaciology, paleoclimatology, and paleomagnetism, for example.

The dispersion of the earth sciences in the international scientific structure may have played a role in the selection of IGBP projects. In the ICSU family, three unions represent the earth sciences: the International Geographical Union, the International Union

of Geodesy and Geophysics, and the International Union of Geological Sciences. It is not easy in these circumstances to present homogeneous core projects. Also, the existence of international geosphere programs, such as IGCP or the International Decade for Natural Disaster Reduction, may have been considered adequate to cover major geosphere risks and global changes. In any case, geoscience information can play an important role in global change studies.

Earth science information is less dispersed than the scientific activities themselves, at least from the standpoint of bibliographic data bases. All publicly available, earth science online data bases include the complete spectrum of solid Earth subfields from geomorphology to geologic hazards, hydrology, and geophysics. In science in general, factual data sets are oriented to projects or topics, and they are highly diverse and geographically dispersed. Geoscience is not an exception to that rule.

IGBP has a working group on data and information systems (IGBP WG 1), which calls for an IGBP directory of existing data sets and a clearinghouse structure for their dissemination. For factual data, the World Data Center can certainly serve the geoscience sector. However, bibliographic information is not mentioned in the description of the working group's activities.

Global science means global research and survey activities. Remote sensing, robotic systems, and modeling are marvelous applications of human skill, and these tools must be fully exploited in order to learn more about global processes. However, no machine or software package can replace personal observations and the interpretation of a biologist or a geoscientist. We are reminded here of a British professor visiting at the university of one of the authors. The gentleman could not understand why anyone would study rocks in the laboratory without having seen and collected them in the field. When his visit ended, he left at the department of geology a box of rocks that he had collected personally as a pointed reminder that original observation is the basic tool of geology. And here again, we are touching on the barrier of different scales of economic development on our planet. The risk exists that remote techniques will take over the task of recovering scientific data directly from Third World countries and that this will broaden the "north-south" gap in the development of information technology (table 2).

IGBP tries hard to involve the developing countries in global change projects and information exchange through the creation of regional research centers. Humanity cannot allow global projects to ignore the scientific potential of three-quarters of all humans, people who inhabit three-fifths of the land surface of the Earth. The motto of IUGS is "*Mente et Malleo*." A freestyle translation could be "Skill and Observation." Both facets should be shared on a global scale. The move toward integration exists in many human activities. For geoscience information and for IUGS specifically, the time has come for the two Commissions concerned with information, COGEODOC and COGEODATA, to move toward the creation of a single Commission dealing with both aspects of information data and documentation. Let us commit ourselves to transform international exchange into global cooperation, to integrate factual and bibliographic data, and to link the "north" and the "south," as well as the "east" and the "west." That is the challenge for the nineties. □

Before early retirement, **Jack Gravesteijn** was Charge de Mission at the Bureau de Recherches Géologiques et Minières (BRGM) in Orléans, France. His primary responsibility there was the International Cooperation Program in Geoscience Information, which involves especially the PASCAL-GEODE network. Previously he held management positions in BRGM's documentation department after a 10-year career as a geologist in mineral exploration projects conducted for BRGM and the United Nations Development Programme (UNDP) in western Africa. He graduated from Utrecht University in The Netherlands. He is Secretary of the IUGS Commission on Geological Documentation, and he devotes part of his time to activities of the Commission, as well as to maintenance of the "Multilingual Thesaurus of Geosciences."

Born in Iraq, **Dr. Rassam** finished college at the University of Baghdad in geology. He received his master's degree from Miami University of Ohio (USA) and his doctorate from the University of Minnesota (USA) in 1967. His professional experience has included teaching at several universities in the USA, oil company experience in France and the Middle East, and nuclear energy research in Iraq. For many years Dr. Rassam was associated with the GeoRef geoscience data base and was Editor-in-Chief of the Bibliography and Index of Geology. He is now Group Director at the American Geophysical Union, where he is responsible for Data Processing, Marketing, and Electronic Products. He also is Chairman of the Commission on Geological Documentation.