GEOSCIENCE INFORMATION SOCIETY

Proceedings • Volume 8 • 1977

PROCEEDINGS OF THE TWELFTH MEETING
OF THE
GEOSCIENCE INFORMATION SOCIETY

November 7, 1977 Seattle, Washington

GEOSCIENCE INFORMATION RETRIEVAL UPDATE

Edited by

Richard D. Walker

PROCEEDINGS

Volume 8

Geoscience Information Society
1978

GEOSCIENCE INFORMATION SOCIETY

PUBLICATIONS

PROCEEDINGS, V.1	(covers 1966, 1967 meetings), published 1969 (ed. R. W. Graves) HANDLING GEOSCIENCE DATA AND INFORMATION	Out of Print
PROCEEDINGS, V.2	(1971 meeting), published 1972 (ed. R. W. Graves) TOWARD THE DEVELOPMENT OF A GEOSCIENCE INFORMATION SYSTEM.	$ 3.00
PROCEEDINGS, V.3	(1972 meeting, Minn., MN), published 1973 (ed. H.K. Phinney) (no title)	$ 6.00
PROCEEDINGS, V.4	(1973 meeting, Dallas, TX), published 1974 (ed. M.W. Wheeler) GEOSCIENCE INFORMATION	$12.00
PROCEEDINGS, V.5	(1974 meeting, Miami Beach, FL), published 1975 (ed. J.L. Morrison) GEOSCIENCE INFORMATION	$12.00
PROCEEDINGS, V.6	(1975 meeting, Salt Lake City, UT, 10th annual meeting), published 1976 (ed. V.S. Hall) RETRIEVAL OF GEOSCIENCE INFORMATION	$15.00
PROCEEDINGS, V.7	(1976 meeting, Denver, CO; 11th annual meeting). published 1977 (ed. J.G. Mulvihill) GEOSCIENCE INFORMATION	$15.00
PROCEEDINGS, V.8	(1977 meeting, Seattle, WA; 12th annual meeting), published 1978 (ed. R.D. Walker) GEOSCIENCE INFORMATION RETRIEVAL UPDATE	$15.00
DIRECTORY OF GEOSCIENCE LIBRARIES: U.S. & CANADA. (2nd ed.)	1974 (ed. R.W. Walker and D. Parker).	$ 5.00
GIS NEWSLETTER.	Subscription for current year (1977); 6 issues.	$12.00
ENVIRONMENTAL GEOLOGY IN THE PITTSBURGH AREA, ed. R. D. Thompson.	Geological Society of America, Field Trip Guidebook No. 6, 1971	$ 5.00
GEOLOGIC FIELD TRIP GUIDEBOOKS OF NORTH AMERICA, A UNION LIST	INCORPORATING MONOGRAPHIC TITLES, 2nd ed., 1971. (Available directly from the publisher; make checks payable to: Phil Wilson Publishing Co., Box 13197, Houston, TX, 77019)	$20.00

ORDERING INFORMATION

Geoscience Information Society
c/o American Geological Institute
5205 Leesburg Pike
Falls Church, VA 22041

Orders must be prepaid. Make check to Geoscience Information Society -- mail order to address at left.

PREFACE

The papers in the volume of the Society's Proceedings were presented orally at the Twelfth Annual Meeting in Seattle, Washington, on November 7, 1977, save one. The paper by Rosalind Walcott, "A Survey of the Holdings of a Sample of International Union of Geological Sciences (IUGS) Publications in Selected U.S. Geoscience Libraries," is a follow-up paper to one by Julie Bichteler, "Publications of the International Union of Geological Sciences: Their Influence on U.S. Geoscientists," which appeared in GIS Proceedings, vol. 7, 1977.

Papers appear herewith as submitted by the authors.

Richard Walker
Program Chairman, 1977

* *

Geoscience Information Society

HISTORY: founded November 5, 1965, in Kansas City; incorporated in March 1966 in the District of Columbia.

PURPOSE: to initiate, aid, and improve the exchange of information in the earth sciences through mutual cooperation among librarians, earth scientists, documentalists, editors, and information specialists.

MEMBERS: more than 200.

MEMBERSHIP: open to persons and organizations whose professional activities are related to geoscience or who are interested in the purpose of the Society.

DUES: $12 (individual), $25 (institutional), $100 (sustaining).

MEETINGS: annual, with that of the Geological Society of America.

PROGRAMS: the management, organization, and dissemination of geoscience information.

PUBLICATIONS: a directory of geoscience libraries, a union list of field-trip guidebooks, proceedings volumes, and the GIS newsletter (6 times a year).

ADDRESS: In care of the American Geological Institute, 5205 Leesburg Pike, Falls Church, VA 22041. Telephone: (703) 379-2480.

CONTENTS

THE NORTH DAKOTA REGIONAL ENVIRONMENTAL ASSESSMENT PROGRAM

(NORTH DAKOTA REAP)

Mary W. Scott
Chester Fritz Library
University of North Dakota

The Regional Environmental Assessment Program (REAP) is an innovative resource information and analysis program for the State of North Dakota. The increasing demand for the development of the lignite resources in North Dakota led to the realization in 1974 of the need for an accurate and coordinated information and forecasting tool for decision makers. To meet this need the North Dakota Legislative Council contracted with Battelle's Columbus Laboratories to conceptualize a comprehensive system. The concept developed by Battelle was drafted into a bill presented to the 1975 Legislature. On April 10, 1975, Governor Arthur Link signed House Bill no. 1004, which established the North Dakota Regional Environmental Assessment Program (REAP) and provided an appropriation of $2 million from a special coal development trust fund.

House Bill no. 1004 provides two mandates for REAP. The first is to establish and carry out research on North Dakota's resources, and the second is to develop the necessary data and information systems. The bill also set forth three specific purposes for REAP:

1. To provide facts and information to the citizens of the state

2. To assist in the development of new laws, policies, and governmental actions

3. To make known the alternatives available to the state in any use or development of resources in order that any such use will enhance the quality of life of the citizens of the state.

REAP is not to act in a decision-making capacity but rather to fill a supportive role for agencies and arms of the government of the state of North Dakota and of local governments. It is to be a centralized source of information and data which will have the capability of providing that information in usable and understandable form. It will gather together the results of many studies already underway and will initiate its own studies where appropriate. It will provide analyses of the information which is available, and will be able to forecast the results of various development alternatives.

In order to implement the two specific mandates contained within the legislation, four major tasks were identified:

1. The acquisition of baseline data from existing sources and through new REAP-sponsored studies

2. The design and development of an information system to provide for storage and retrieval of data and to execute appropriate analyses, summaries, and presentations of such data

3. The design and development of a modeling/assessment system

4. The design and development of a monitoring system.

The Legislative Council appointed a Resources Research Committee to be responsible for implementing the program. This committee consisted of three state senators, two state representatives, five representatives of state agencies in the executive branch, two representatives of the state universities, and one citizen at large. A director and technical staff were appointed. A Technical Advisory Council, made up of six specialists from state government and universities, was formed to provide advice to REAP on its total program. Eleven Technical Task Forces made up of ninety-two specialists from universities, state and federal agencies, and industry were convened to assess the availability of baseline data and to identify data gaps. The Technical Task Forces were organized by subject areas: air quality and meteorology; animals; geology; historical, archaeological and paleontological sites; land use; noise, radiation, and solid waste; social impact and quality of life; socio-economic impact; soils; vegetation; and water. The Task Force reports were completed in December of 1975. The REAP staff then identified priority study projects. Twenty baseline data acquisition projects were funded to fill data gaps identified by the Technical Task Forces.

REAP User Specification Teams (RUSTEAMS) were organized from the state agencies that would be using REAP-produced data. The RUSTEAMS were to identify and then describe the specific outputs and needs which they envision from REAP for their agencies, such as kinds of maps, tables, statistical summaries, and computer space. The needs identified by the REAP User Specification Teams and other requirements for the system were compiled, and IBM was contracted to design the total REAP system. In addition REAP conducted a complete land cover analysis of the state using satellite imagery, and county and state land analysis maps have been published.

An Econometric-Demographic model (E-D model) was designed and is now operational to make baseline projections and forecast the impact of potential development on the economy, population, employment and public-sector finance.

The REAP Resources Reference System or R^3S was implemented. This is a computer-based on-line interactive information system. It includes four different files: 1) people--for the identification of technical expertise in the state, 2) data source--identification and description of sources of data, 3) project--identification of current study projects in the state, and 4) bibliography--listing references to published materials containing data and analyses pertaining to North Dakota.

This last file is the part of REAP that I have been involved with this past year. One of the high priority needs identified by the Geology Technical Task Force was an up-to-date bibliography. A bibliography of the geology of the state had been compiled through 1959, but the late 50's saw a boom in geological research in North Dakota, triggered by the discovery of oil in the Williston Basin in 1951 and interest in uranium. The literature that had not been indexed was estimated to equal the volume indexed for the years 1806 through 1959. A water resources bibliography was also needed since water will play a large part in the coal development in the state. The North Dakota Water Resources Institute had funded a bibliography of the water resources of the state through 1970. It was decided that this should also be updated at the same time. The reason these two subjects were tied together as one project is that I had compiled the two bibliographies that had already been completed. A contract was negotiated between the North Dakota Geological Survey and REAP for funding of the project for one year. I was granted a one-year leave of absence from my position at the University of North Dakota Library, and the project was off and running.

The R^3S bibliographic file is set up to include environment-related published and unpublished papers, reports, books, articles, and manuscripts which pertain to North Dakota. The computer is capable of searching text, so each entry was to include an abstract. Key words could be used if they were taken from another data bank; otherwise an abstract had to be written. I immediately saw problems with having to write hundreds of abstracts, so a compromise was agreed upon. The author's abstract was used where possible; otherwise I assigned key words to the article using the Water Resources Thesaurus, published by the U. S. Department of the Interior, Office of Water Resources Research. Key words from the Department of

Defense's "Thesaurus of Engineering and Scientific Terms" and the American Geological Institute's "Glossary of Geology and Related Sciences" were used to select geologic key words.

The project started with a search of all the appropriate computerized bibliographic files I could think of. Searches were run against GEO-REF, CAIN, POLLUTION, ENVIRO LINE, BIOSIS, COMPENDEX, TULSA, NTIS, and the Office of Water Research file. I had a card file of several hundred geologic references that I had accumulated over the years, and there were some references left over from the earlier water bibliography project. The Bismarck office of REAP was to enter the two completed bibliographies into the file, the geology through 1959 from the published form, and the Water Resources through 1970 from a computer tape. As I located a reference and obtained a copy of the abstract or assigned key words to it, I completed a computer form which was sent to the Bismarck office to be entered into the computer file. I tried to look at every document indexed and not rely just on abstracts or key words from other files. One problem in the project has been discovering all the literature on the state that the bibliographic services miss but we need control of, such as North Dakota Water and Sewage Conference Official Bulletin. I spent considerable time going through files and piles of literature in the libraries and offices around the state. This is part of what makes REAP unique: it is designed for local needs and users. The unpublished literature is another problem. As a librarian, I really hesitated to include unpublished literature, but R^3S is set up to include the location of this material, so some manuscripts were included if I could locate them and could obtain permission to include them in the file with a location note. I added about 1,500 new references to the file during the project. The total file now has over 10,000 documents.

The R^3S file will also be unique in that it is not limited to one narrow subject like geology. All literature pertaining to the North Dakota environment will be included in one file. There may be problems with the lack of control over index terminology. This will have to be worked out by the R^3S Coordinator, who conducts the searches of the file. One problem was encountered after the computer tape of the water resources bibliography that have been compiled for the Water Resources Research Institute had been entered into the REAP R^3S file. Key words on the tape had been selected from the Water Resources Thesaurus, and REAP did not have a copy of this or of the modifications I had made in it. During a demonstration there was a request for references on salinity in water and the Garrison Diversion Project. Only about three references showed up, and they could not figure out why. The reason was that due to indexing restrictions on the first project a decision had been made to use the term Missouri River Diversion instead of Garrison Diversion. Cross references such as this are being added to the computer. Another problem in searching geological literature will be changes in stratigraphic nomenclature. There is currently a project as a thesis problem at the University of North Dakota Geology Department to try to unravel the history of stratigraphic nomenclature in North Dakota.

A third problem was one of defining the limits of the project. It is sometimes hard not to wander off into biology, soils, or other closely related fields, particularly when you feel that your project may be the only special bibliographic project funded. The plan is to build the bibliographic file in other areas by asking every REAP researcher to submit bibliographic forms for every reference used. The R^3S coordinator in Bismarck has the task of eliminating duplication in the file, standardizing the bibliographic formats and trying to fill in gaps in the file.

We have tried to make the file as versatile as possible. Additional key words to tailor the file to North Dakota were added to entries that already had an abstract or key words. County, city, and river names were common key words added. One problem may be that the indexing was too detailed and a person requesting general information like the geology of the Red River Valley will not get much information. The R^3S Coordinator is going to have to be very familiar with the file and knowledgeable in the terminology of the many areas of environmental studies in order to conduct good searches.

One of the three specific purposes for REAP was "To provide facts and information to the citizens of the State." The R^3S files are available to anyone in North Dakota. Access to the file is free for state agencies, local governments, and the institutions of higher education.

REAP R^3S is set up to be the one-stop shopping center for environmental information in North Dakota. Here a person with one phone call can find out what has been published on a subject, who the experts in the state are on that subject, what, if any, projects are now being conducted on that subject and when they will be completed and where data pertinent to the subject are located. When fully operational this file should save hours, if not days, of research time. And we hope it will enable our government officials to make more intelligent decisions concerning our environment.

THE NATIONAL WATER DATA EXCHANGE (NAWDEX)

By Melvin D. Edwards [1]

ABSTRACT

The National Water Data Exchange (NAWDEX) was established in 1976 to assist users of water data to identify, locate, and acquire needed data. NAWDEX is a confederation of water-oriented organizations working together to provide more timely and convenient access to their data. A directory of sources of water data and a nationwide index of available water data are maintained for the storage and dissemination of information on available water data. Assistance services are provided through a nationwide network of Local Assistance Centers.

INTRODUCTION

The need for a viable program for the improvement of the transfer of data from collecters to users is acknowledged throughout the water-data community. The National Water Data Exchange (NAWDEX) has been established in response to this need.

The design characteristics for NAWDEX were first presented by the Federal Interagency Water Data Handling Work Group, a task group of the Federal Interagency Advisory Committee on Water Data, in October 1971. The implementation of the program was recommended by the Federal Interagency Advisory Committee on Water Data and the non-Federal Advisory Committee on Water Data for Public Use. These committees work under the auspices of the U.S. Geological Survey's Office of Water Data Coordination which implements the guidelines for the coordination of Federal activities in the acquisition, storage, and dissemination of water data as prescribed by Circular A-67 issued by the Office of Management and Budget in 1964.

The U.S. Geological Survey accepted the lead-role responsibility for implementing NAWDEX and work began on this effort in January 1973. Two years of subsequent planning and design effort resulted in the operational implementation of NAWDEX in August 1975 and the formal establishment of the program in January 1976.

THE NAWDEX MISSION

The mission of NAWDEX is to identify sources of water data and to provide the linkage between those who acquire and those who use water data. This is being accomplished by establishing better communication between water-oriented organizations, acquiring as much information as

1/ Hydrologist, U.S. Geological Survey, Reston, VA 22092.

possible on available water data, and making this information readily available to water-data users. These efforts contribute to the improvement of the transfer of water data and information on water data between the collecter and user communities.

THE NAWDEX CONCEPT

NAWDEX is a confederation of water-oriented organizations working together to provide timely and convenient access to their data. This concept is directed at closer coordination of water-data dissemination activities, improving communication within the water-data community, more effectively utilizing the data resources, expertise, and talent within the community, improving the quality and efficiencies of techniques and procedures used for the exchange of water data, and providing water data to those who need these data in a cost-effective manner.

THE ORGANIZATION OF NAWDEX

NAWDEX is centrally managed by a Program Office located within the U.S. Geological Survey's Water Resources Division in Reston, Virginia. This office provides the coordination for all NAWDEX activities, develops the technical systems and procedures necessary for the operation of NAWDEX, maintains the data bases and other information resources required for operation, provides operational guidelines to all service centers, and serves as a liaison between members, external organizations, advisory committees, and the users of NAWDEX services and facilities.

The basic components of NAWDEX are organizations that become participating members. They are organizations that are collecters of water data, contributors from the water-data community who provide support toward program improvement, collecters of data related to water-resources activities, and users of water data. All members contribute support toward improvements in data exchange procedures and techniques.

NAWDEX MEMBERSHIP

Membership in NAWDEX is voluntary and open to any organization active in the field of water resources. This includes organizations from the Federal, State, interstate, local governmental, academic, and private sectors of the water-data community. There are no dues or fees associated with membership. Members are required, however, to sign a Memorandum of Understanding with the Program Office which defines that office's responsibilities and a general commitment of the member to take an active role in NAWDEX activities, to provide information on its data holdings for indexing purposes, and to provide data from its holdings upon request. The Program Office is not empowered to sign membership documents with foreign or international organizations. These organizations are invited, however, to participate in NAWDEX through an established relationship with the Program Office and will be identified in NAWDEX as foreign affiliates.

Each member takes an active role in the planning, development, and execution of NAWDEX activities. Each member is an advisor to the Program Office and mechanisms exist to assure continued communication between members, to assure active participation of all members in an equal and equitable manner, and to maintain an awareness of the needs of the water-data community in order to achieve maximum program effectiveness.

NAWDEX SERVICES

A variety of services are provided by NAWDEX to assist its users to identify needed water data, to locate these data and to refer users to the proper sources for obtaining the data. It is not a function of NAWDEX to become a repository of water data. Instead, the data held by NAWDEX members are indexed by the Program Office to provide a central source of information on water data available from a large number of organizations.

Local Assistance Centers: NAWDEX services are provided through the Program Office and a nationwide network of Local Assistance Centers. This network was established in January 1977 and consists of 51 Centers located in 45 states and Puerto Rico. The Centers provide convenient access to NAWDEX services as well as making local-area expertise available in the identification and location of needed data. Additional Centers will be added to the network, as needed, to meet the demands of the user community.

Identification of Sources of Water Data: NAWDEX has extensive information available that identify domestic and foreign organizations that are sources of water data. This information is provided through a computerized Water Data Sources Directory maintained in the U.S. Geological Survey's computer system in Reston, Virginia and is accessible by nearly all Local Assistance Centers by remote computer terminal.

The Water Data Sources Directory identifies organizations that collect water data, locations within these organizations from which water data may be obtained, alternate sources from which an organization's water data may be obtained, the geographic areas in which an organization collects water data, and the types of water data collected and available. Information has been compiled for over 300 organizations. An additional 200 organizations are expected to be added to the Directory by the end of 1978. The contents of the Water Data Sources Directory will be periodically published.

Indexing of Water Data: A nationwide indexing service is provided by NAWDEX through its Master Water Data Index. This is a computerized index that identifies sites for which water data are available, the location of these sites, the data-collecting organization, the types of data available, the periods of time for which data are available, the major water-data parameters for which data are available, the frequency of measurement of these parameters, and the media in which the data are stored. The Master Water Data Index also is maintained on the U.S. Geological Survey's computer system in Reston, Virginia and is accessible by most Local Assistance Centers by remote computer terminal.

The Master Water Data Index was created in June 1976 with information on 62,000 water-data sites. Computerized interfaces being developed between the Index and data files of NAWDEX members will increase its contents to information on about 350,000 sites by the end of 1977. Information on nearly 1 million sites is anticipated by the end of 1978.

All NAWDEX members are encouraged to use the Master Water Data Index as their primary facility for water-data inventory and to contribute to its contents to the greatest extent possible. Foreign organizations also are encouraged to contribute information on their pertinent water-data holdings. Through the coordinated development of both the Master Water Data Index and the Water Data Sources Directory, NAWDEX is creating a viable and comprehensive base of information on available water data which contributes to the transfer of information on water resources throughout the water-data community.

Data Search Assistance: Through its Local Assistance Centers, Water Data Sources Directory, and Master Water Data Index, NAWDEX can readily assist users of water data in locating data meeting their specified criteria and in geographic areas of interest. Users can be quickly referred to water-data systems, bibliographic data services, other indexing services, or data-collecting organizations that can provide the information, data, or other services required. Hydrologists knowledgeable in the water resources of local areas are available to provide assistance at each Local Assistance Center. Each request is reviewed as it is received by NAWDEX to assure that the request accurately reflects the requester's needs. The request can then be referred to appropriate sources with sufficient information to assure a correct response in an efficient and timely manner. The amount of time required to locate needed information is, therefore, greatly reduced in most requests for data.

Charges for NAWDEX Service: Free services are provided to the greatest extent possible by the Program Office and Local Assistance Centers in their data-search assistance. Charges are applied, however, for services requiring computer costs, extensive personnel time, duplicating expenses, or costs accrued by NAWDEX from other sources in the course of providing services. In all cases, charges assessed by the Program Office and Local Assistance Centers will not exceed the direct costs incurred in responding to a data request. Estimates of cost are provided upon request and in all cases where costs are anticipated to be substantial.

Charges also may be assessed by organizations to which a data requester has been referred for data or services. These charges are assessed at the discretion of the responding organization and they are reimbursed directly to the assessing organization.

FUTURE PLANS FOR NAWDEX

Within the near future, several new projects are planned within NAWDEX. These include the preparation of a Directory of Sources of Water Related Data, the indexing of areally-related water data, and acquiring membership agreements to facilitate the handling of bibliographic references to published water data. Work will begin also on the development of recommended standards for the handling and transfer of water data. All NAWDEX members will be invited to participate in this important standardization activity. Input will be sought from all sectors of the collecter and user communities.

SUMMARY

NAWDEX, as a confederation of water-oriented organizations working together, can do much to contribute to the transfer of water data throughout the entire water resources community. It represents a concerted effort to improve the quality and efficiency of data transfer from collecters and users. This effort will improve communication throughout the water-data community, create a better awareness of existing water data, make data-search assistance more readily available to those who need it, reduce the time required to obtain data, increase the utilization of existing water data, and help to reduce the redundancy in data collection. All these benefits contribute toward making water resources programs more effective.

The support and response of member organizations has been excellent and requests for NAWDEX services are growing constantly. The Program Office, Local Assistance Centers, and members will continue to work to advance NAWDEX as rapidly as possible so that the user community may derive maximum benefit from its facilities. With continuing support and participation, the objectives of a viable and effective National Water Data Exchange will be achieved.

ADDITIONAL INFORMATION

For services or additional information concerning NAWDEX, please contact:

National Water Data Exchange
U.S. Geological Survey
421 National Center
Reston, Virginia 22092
Telephone: (703) 860-6031
FTS 928-6031

AN EVALUATION OF OASIS
NOAA's OCEANIC AND ATMOSPHERIC SCIENTIFIC INFORMATION SYSTEM

by Robert R. Freeman and James R. Stear
Environmental Science Information Center
Environmental Data Service
National Oceanic and Atmospheric Administration
Rockville, Maryland 20852

ABSTRACT: The OASIS concept was developed in 1971, shortly after the creation of NOAA. The original goals of the program were to improve nationwide access to the published literature of the oceanic, atmospheric, and related earth sciences through bibliographic data bases, an information service center, a document delivery system, and a network of regional information centers. The paper presents an evaluation of the effectiveness of the OASIS program, including an assessment of the impact of changing priorities, new information technology, an increasingly viable private information industry, and new user communities on the course of development. The program has played a role in the development of the national bibliographic information system by sponsoring the availability of three data bases, all containing coverage of some aspects of the earth sciences literature. These are Meteorological and Geoastrophysical Abstracts, Oceanic Abstracts, and Aquatic Sciences and Fisheries Abstracts. Using these and over 40 other data bases available through commercial, other government agencies, and its own computer systems, NOAA now provides nearly 3,000 searches each year. The majority are now provided as a part of normal library and information services. Users from outside NOAA pay a charge based on the marginal cost of the service. The regional information centers portion of the program at first was implemented only through making terminals available in NOAA field and regional offices. More recently, however, this concept has received a new impetus in the form of regional coastal information centers, a cooperative program of the Environmental Data Service and the Offices of Sea Grant and Coastal Zone Management.

Introduction. Evaluation of a public program implies a definite frame of reference, a set of criteria based on specific goals and objectives. In order to evaluate OASIS, we will need to trace its evolution -- why was it started, what were the original intentions, what became of them, what is it now, and why did it happen? One of the more important aspects of evaluation is answering the question: where are we headed? We will try to address this question too.

Origins of OASIS. Although there was active interest in information systems in some of NOAA's predecessor agencies, we take as a convenient point of inspiration a recommendation by the Stratton Commission in its 1969 report, Our Nation and the Sea.[1] The Commission recommended that the new agency that became NOAA develop a strong technical information system as one of the supporting services for its own mission and for the public.

Why should the Government do it? The reason is that information delivery systems have some characteristics of what economists call a public good. This means simply that once it is available, too many of the users are in a position to benefit from the service without paying for its full worth. As a result, private industry will not produce the service, and the field is left to Government and to voluntary associations such as professional societies if the service is to be rendered at all. This doesn't mean that there is no role for private industry nor that users will not pay anything at all, but simply that a significant proportion of the resources must come from the Government.

Having resolved that question, we faced the questions of what form NOAA's technical information system should take and how should it be implemented. There are in fact many types, and it's not at all unreasonable to say that large portions of NOAA's total program are there for the purpose of delivering technical information.

In our case, we defined a technical information system as a service that delivers specialized technical publications and references to published information related to a governmental mission as a complement to conventional library services on one hand and to data centers on the other.

From the beginning in 1971 we set out to provide a decentralized service. Using existing models, such as those of WRSIC or ERIC, we planned for:

(1) a central coordinating office;

(2) a set of input "centers of expertise;"

(3) a set of service centers located in major user community locations;

(4) a document delivery support system based on our facilities and those of NTIS; and

(5) an emphasis on providing computer-searchable bibliographic data bases.

Our original plan[2] especially focussed on secondary access to the published literature in fields of concern to the new agency. We found outstanding progress in developing computer-searchable data bases in most fields, but glaring exceptions in sciences of greatest concern to NOAA -- oceanography, fisheries, and meteorology. We decided to acquire those data bases of interest to us and to fill in the gaps by creating new data bases.

Our original estimate for implementing this plan was in the neighborhood of a million dollars a year. The decision was made to seek to initiate a pilot program. Congress appropriated funds, but the funds were impounded by the Office of Management and Budget. As a result, OASIS could go forward only with funds reprogrammed from lower priority activities. This was the reason for entering on phase 2.

In phase 2, we began to focus on developing the data bases and loading them into an information retrieval system (GIPSY) to support them on our own NOAA computer system. All the plans for decentralization went on the back burner, but OASIS went operational in a small way in 1974. We published a User's Guide to OASIS[3] and conducted a number of training sessions in 1974-75.

But now we began increasingly to feel the impact of external forces. A strong private information industry had begun to emerge. Among the capabilities offered were better facilities for nationwide on-line access to data bases. Not only could we access the data bases produced by others more easily, but we could provide others with better access to the data bases we support.

We entered phase 3 with a competitive procurement which resulted in two of our data bases going up on the Lockheed Dialog system, later to be joined by a third. By late 1975 OASIS had the capability of delivering both references directly to users equipped with terminals or to those who chose to contact our service centers. Service has been provided to the public on a cost recovery basis, while NOAA employees received requested services free of charge through a central funding arrangement.

The current or fourth phase began soon afterward with the increasing realization of the similarities between the now operational technical information services and the library services that had long preceded them. Beginning in 1976 we reorganized and consolidated the two, providing the ability to respond to requests with the full battery of published information resources.[4]

As a result, OASIS no longer exists as a distinct organizational entity or service, a point which causes some confusion. Nevertheless, one measure of success for a public program is institutionalization, that is, integration into a permanent operating service. This we have achieved.

Present Characteristics of OASIS. OASIS at present is best characterized as a program or a concept. It has eight recognizable elements.

These reflect our philosophy of providing the most effective service by using a combination of our own resources and those that private industry can provide.

First, as we indicated above, we currently use the services of Lockheed Dialog to support our data bases. This means that not only do we get the service we need without having to have a computer system maintenance staff, but more importantly anyone interested can search the data bases from locations in North America, Western Europe, Hawaii, and Australia without our having to intervene.

Second, we purchased one data base, Oceanic Abstracts, from the supplier, Data Courier, Inc. We began this arrangement with the judgment that there was not enough of a market to warrant making this data base available entirely through the private market. On-line searching has caught on so fast, however, that we may be able to withdraw public funding before long. Oceanic Abstracts now contains over 90,000 citations covering 1964 to present.

Third, we provide base support for abstracting, indexing, and data base creation to the American Meteorological Society from whom we acquire Meteorological and Geoastrophysical Abstracts. MGA now includes about 50,000 citations from 1970 to present.

Fourth, we participate actively in the international Aquatic Sciences and Fisheries Information System (ASFIS), which is coordinated by the FAO and the IOC of UNESCO.[5] ASFIS also now has centers in Canada, Britain, France, Germany, Mexico, Portugal, the Soviet Union, and the United Nations. A small but growing ASFA data base covering the science and technology of the marine and freshwater environments is now beginning to be produced. We expect to be adding to this data base at the rate of about 20,000 citations per year beginning in 1978. A unique feature of this system is the participation of a private firm, Information Retrieval Limited, as the editorial, production, and marketing arm under contract to FAO. NOAA supports the availability of the data base on Lockheed Dialog, while our Canadian Government counterparts have made it available through a Canadian firm, Q/L Systems.

Fifth, we support some small, specialized data bases on an internal NOAA computer system. These include files of information in aquaculture and marine biology.

Sixth, we provide a centralized NOAA service so that a growing number of NOAA locations can have access to environmentally related data bases available through SDC, BRS, Lockheed, NLM, NASA, and DoE.

Seventh, OASIS services are now coordinated with other library and information services so that users can be certain of obtaining the best that we can offer through a single source.

Eighth, as has always been our policy, OASIS and other library and information services are coordinated with the delivery of environmental data services, including both our ENDEX data referral program[7] and the resources of our data centers in oceanography, climatology, and geophysics and solar-terrestrial phenomena. Users can request services through any of these centers or through field offices in Woods Hole, Miami, La Jolla, Seattle, and Anchorage.

<u>Services Provided</u>. Approximately 45 data bases are accessed to provide referral to published literature in the environmental sciences.. The majority of these are scientifically oriented, each concerning either a specific

aspect of science or multiple scientific disciplines. Data bases are also included which cover publications relating to the legal and policy aspects of the environment. Referral to newspaper articles and popular writings have been a recent valuable addition to the scope of service.

The demand for service has continued to increase since the beginning of OASIS in July 1974. Currently, approximately 500 searches are performed/provided monthly. The service network includes 39 computer terminals at 35 NOAA locations throughout the United States. Searches are performed by librarians, information specialists, scientists, and technicians.

Three types of services are offered. Retrieval on a specific subject from an entire data base (Retrospective Search) for users who need to survey all that has been published and made a part of the data base being searched. Selective Dissemination of Information (SDI) searches provided from citations added to a data base at the time of update (usually monthly) or at a time interval convenient for the user. This type of automatic service keeps scientists and managers informed of new publications in their field of interest with minimal effort. A third type of service are published searches on topics of current or widespread interest. Typical topics (titles) include Oil Spills: Cleaning Up and The Coastal Zone. The published searches are reproduced directly from output from computerized searches of NOAA-sponsored data bases. Periodic updates are made to keep the published searches current.

The marked increase in number of referrals to published literature in NOAA's scientific disciplines has had the expected impact of a much greater demand (document delivery) for publications. During the past two years, NOAA has participated with other governmental libraries in sharing lists of publications. Listings of publications (Union List of Serials) have been produced which contain the holdings of NOAA, EPA, NBS, the U.S. Patent and Trademark Office, Bureau of Foods (FDA), and NASA. This allows for resource sharing and cooperation and assures a greater chance for document delivery through interlibrary loans. A new governmental interagency automation effort will place a Union List of Serials for NOAA, EPA, and NAL online for searching through the Bibliographic Retrieval Service, Inc., which is one of the three major vendors for on-line access to bibliographic data bases. This new service will allow for on-line Interlibrary Loan requests for the participating agencies.

Regional Information Services. Changing times and the evolution of NOAA's missions and emphases have led us to give more attention to the needs of a broader community of information users.[8] While our traditional clientele has been the scientific community, we are now trying to reach those whose primary interest is policy. These include, for example, Federal, state, and local legislators, legislative staffs. administrators, lawyers, and citizen action groups. Our mode of developing this new service, although not part of the OASIS concept, harkens back to the original OASIS plan for decentralized service centers.

Through joint action by the EDS, the Sea Grant Program, and the Office of Coastal Zone Management, we are establishing a network of Regional Coastal Information Centers, located in Sea Grant Institutions and linked to our national centers.[9] Eventually we hope that there will be nine of these centers. So far we have established two. If all goes well, we hope to solicit proposals for up to two more in 1978.

The present two include one for the Pacific Northwest and one for the

Northeast. The former is operated by a consortium of Oregon State University, the University of Washington, and the Oregon Estuarine Research Council. The latter is operated by the New England Marine Advisory Service with headquarters in Durham, New Hampshire, and its major facility at the University of Rhode Island.

Conclusion. To summarize, our "one sentence evaluation" of OASIS is that most of the original goals are being accomplished, but in ways we hardly imagined only seven years ago. OASIS is now firmly integrated into a comprehensive library and information program that provides a range of reference services, and document delivery services throughout NOAA, with many links established to international and U.S. federal, state, and regional information programs in the environmental sciences and marine and coastal resource utilization and management.

REFERENCES

1. U.S. Commission on Marine Science, Engineering and Resources, Our Nation and the Sea: A Plan for National Action. Washington, Government Printing Office, January 1969

2. Freeman, Robert R., A Plan for an Oceanic and Atmospheric Scientific Information System. National Oceanic & Atmospheric Administration, Rockville, Md. Scientific Information & Documentation Division, April 30, 1971, COM-75-11424/9ST

3. Freeman, Robert R. and Stear, James R., User's Guide to OASIS: Oceanic and Atmospheric Scientific Information System National Oceanic & Atmospheric Administration, Washington, D.C. Environmental Data Service, Key to Oceanic and Atmosperic Information Sources No. 1, May 1974, COM-74-11524/7

4. Yeates, Elizabeth J., "NOAA's Marine Library and Information Services." EDS: Environmental Data Service, November 1977, pp 8-12

5. (Anonymous) "Aquatic Sciences and Fisheries Information System: An International Cooperative Exchange Mechanism." EDS: Environmental Data Service, January 1977, pp 24-25

6. Laughrun, May, "OASIS: A View from the Desert." EDS: Environmental Data Service, September 1976, pp 7-13

7. Gelfeld, Robert and McCann, Michael, "Endexing the Environment." EDS: Environmental Data Service, March 1977, pp 13-15

8. Austin, Thomas S., "Challenge in Meeting Public Needs for Science Information." EDS: Environmental Data Service, November 1977, pp 3-7

9. McGuire, Donald, "Regional Coastal Information Centers." EDS: Environmental Data Service (forthcoming)

Petroleum Data System--A Network of Energy Information

by Patricia A. Tracy
University of Oklahoma
Norman, Oklahoma

The Petroleum Data System consists of information on over 75,000 oil and gas fields in the United States and Canada. The Petroleum Data System or PDS, as the group of data bases is called, is a collection of publicly available information on all oil and gas fields and reservoirs. These data bases were developed over the past seven years by the University of Oklahoma under contract to the United States Geological Survey.

Information from hundreds of published reports have been incorporated into PDS, creating a library of computerized energy data. State annual reports are the primary source of data. Supplementary sources include the International Oil Scouts Association and state geologists. Federal agencies such as the Federal Energy Administration, Federal Power Commission, Energy Research and Development Administration, and Bureau of Mines are major contributors of data.

These files provide locations, geologic, engineering and production information. They are being used to analyze trends and risks, and perform economic evaluations and enhanced recovery studies.

The University has chosen to use several existing software systems to help it accomplish its goal of a complete energy network. These systems are generalized systems, which have proved themselves adaptable to a wide variety of applications. The use of existing software has freed the University of many problems in program design and implementation. It has also resulted in lower implementation costs and lower operating costs.

PDS uses a highly sophisticated storage and retrieval system called GIPSY. GIPSY was developed at the University of Oklahoma and has been operational since July, 1968. GIPSY is an acronym for General Information Processing System. The computer environment used by GIPSY is a mainframe of IBM 360 or 370 series operating under IBM OS. It is an excellent system for handling the diverse types of data in PDS, i.e., numbers, codes, and alphanumeric character strings, both fixed length and variable length.

GIPSY has certain characteristics which are prerequisite to the management of any large data base system. As I mentioned before, the Petroleum Data System has over 75,000 records. A single record may have as many as 600 different items of information. Records vary considerably in the degree of completeness. A typical record may have as few as 25 items of information or up to approximately 300 items of information. GIPSY uses economical storage techniques, compressing all empty spaces from the file. GIPSY takes advantage of binary searching techniques to locate data and to decrease costs and computer time.

GIPSY also has the flexibility necessary to handle the constantly changing facade of the PDS data. A set of utility programs handle standard file maintenance procedures, such as adding records to files, deleting records, creating backups, restoring files, and updating files. The update utility allows for the addition or replacement of data to PDS. It also allows for the deletion and for the extension of existing information such as a comments fields or a source document.

PDS is continually being updated. New discoveries are added to the file, and new sources of information are continually researched to fill in missing data. Frequently it is necessary to "restructure" the file by defining new data items, or by deleting items which were defined and have since been determined unnecessary. For example, we are currently adding the latitude and longitude of the centerpoints of the fields to the records in PDS. These items were not defined when the files were initially created. GIPSY handles these new data items with ease. All that is necessary is to add a new data definition to the file dictionary. No special programming is required.

GIPSY is a complete retrieval system. It allows the user to isolate from the mass of information that part which satisfies his immediate requirements. GIPSY has its own retrieval language. It is a simple command language which can be learned with a minimal degree of effort.

Two commands are available for searching data bases and retrieving selected subsets of records.

SELECT will peruse an entire record file or the index of the record file checking records for conditions which the user specifies. As many conditions as necessary may be specified. Conditions can be specified which simply check records for the existence of a particular data item. Numeric items can be checked to determine whether they are less than, greater than, or equal to a given number, or to determine if they fall within a given numeric range. An alphanumeric item can be checked for a specific character string, or for strings which fall within a given character range. Boolean logic (and, or, not) is used to describe how the conditions should be met in the records.

ITERATE will narrow a subset of records down to a smaller, more specific set of records.

A command called SORT will order records in whatever sequence is desired.

Using a command called PRINT, records in the system can be printed in their entirety in a predefined format. The command LIST can be used to print only selected information from the subset of records.

The SUM command will generate sums, averages, maximums, and minimums of specific items in a subset. The COUNT command generates one way frequency distributions.

GIPSY also provides a method of interfacing the data with other processing systems. The COPY command will output data in a fixed format defined by the user to tape or disk. The data can then be manipulated in any manner, either with special user-written programs, or interfaced with statistical packages, plotting routines or report-writers.

GIPSY gives the Petroleum Data System an economical, flexible, user-oriented retrieval mechanism.

The Petroleum Data System has been available to the public since August, 1975. Initially, the University attempted to make the file available through its own in-house computer time-sharing system. However, the response from the industry was so overwhelming that the University was unable to fulfill the needs of its customers.

In an effort to make the data more readily available to users of the Petroleum Data System, the University of Oklahoma has put the system on the General Electric time-sharing network. A validated customer may access the data via a local telephone call from any major city in the United States and cities in 20 foreign countries. The Petroleum Data System is now accessible to any group, large or small, government or private industry, on any scale they wish to operate, from a country-wide study to a small area such as a county or field.

The Petroleum Data System resides on the General Electric remote batch system. This provides a significant cost savings as opposed to the cost of trying to keep such a large volume of data available in an on-line interactive file.

Among the PDS users there is a large variation in user expertise. Typical users include geologists, consultants, programmers, systems analysts, and geological secretaries and librarians. To simplify the use of the remote batch computer, an interface was developed between the foreground interactive time-sharing service and the remote batch service. This interface, called a foreground driver, sets up the IBM Job Control Language and the GIPSY batch retrieval commands and automatically submits them to the remote batch computer. The driver program also monitors the job status, retrieves reports, and gives job statistics.

In a typical terminal session, the user submits a job to select a subset of records and to output certain information from those records. Generally, the user may log back on the terminal within 30 minutes to an hour after the job has been submitted to retrieve the results of the job. Printed reports may be listed totally or in part at the terminal. Large reports can be printed on the customer's in-house high-speed printer, or printed at General Electric and couriered to the customer's office. General Electric also provides many statistical packages and report generating programs which can be used for data manipulation.

PDS has been linked to an applications package developed by Amoco Production Company. The Applications Management System provides data processing, file interface, and display capabilities.

AMS is composed of many separate modules. These modules perform specific functions, which are data independent and often have general application. This allows the AMS user to develop custom solutions to a variety of problems, depending on the manner in which the modules are blended. Models have been developed for PDS which will generate scatterplots, histograms, regression analyses, trend analyses, and write reports.

Scatterplot, for example, will plot any two numeric items from any data base in PDS. The user responds to a series of questions asking for such information as the x,y coordinates and size of plots.

Sophisticated PDS/AMS users have the ability to develop models tailored to more specific needs. AMS provides a uniform syntax to aid the user in communicating with the data structure and the computer, thus simplifying and streamlining the computer process.

The next stage of development will be to incorporate a graphics package into the Petroleum Data System. This will give users the ability to plot locations and assign symbolic values to the graphics data.

Many other energy resource related data bases are being developed including coal, geothermal, and minerals. It is the plan of the University to make all of these data bases available on the General Electric time-sharing network. While the data bases differ substantially in file content and structure, the customer will be able to use the same operating procedure to access the data.

The result of all this has been a highly successful energy information network. Using developed software has been a distinct advantage. It has given us a wide range of flexibility by allowing us to develop applications as the demand arises. By molding these packages to meet the requirements of the Petroleum Data System and the many other energy related data bases, we have been able to provide our users with a large variety of output options.

Much of the success of the system can be attributed to the energy crisis and to an awakened realization by the industry for the need to immediate access to a variety of energy related data. It is our goal to encourage industry and government agencies to cooperate in the exchange of information which might be used in the conservation and development of our resources.

The National Data Referral System for Canadian Geoscience[1]

Cornelius F. Burk, Jr.
Canada Centre for Geoscience Data
Department of Energy, Mines and Resources
Ottawa, Canada K1A 0E4

ABSTRACT

During the past ten years "national" information systems in support of the geosciences have been proposed for Canada, the United States, the Federal Republic of Germany, Czechoslovakia, Australia and other countries. Implementation in Canada, and perhaps elsewhere, has been slowed by the lack of a realistic system model and by failure to apply certain basic managerial principles. Nevertheless, development of the "national system" recommended in 1967 for Canadian geoscience has evolved and progressed to a point where it shows promise of maturing over the next 1-2 years as a fully functional system.

The Canadian Geoscience Data Referral System (CGDRS) is the new name for this system. Its overall objective is to provide knowledge on the existence, location and nature of public geoscience data dealing with the Canadian landmass and its offshore regions. Thus CGDRS deals operationally with the management of secondary information rather than with primary data as proposed in 1967. The system concept involves eleven major elements for policy, international liaison, capture and acquisition of secondary data, national coordination and operations, the generation and use of referral tools and the user community.

CGDRS has been designed to operate within a structured national and international network of interdependent referral centres for scientific and technical information (STI), the full development of which is anticipated during the coming decade.

---

[1]Presented to the 12th Annual Meeting of the Geoscience Information Society, Seattle, Washington, 7 November 1977.

Introduction

At the Society's second annual meeting, held in New Orleans in 1967, I described a proposed "national system for geological data" for Canada (Burk, 1969). Has such a system in fact materialized during the past ten years? Proposals for other "national systems" in support of the geosciences have been enunciated in the United States (Hambleton, 1970, 1971; Paladino and Wachtman, 1976), the Federal Republic of Germany (Institute for Documentation, 1976), Czechoslovakia (Hruska, 1975) and Australia (Parkin and Tellis, 1977), among others. Taking into account these various national planning activities - and the apparently limited results, what have we learned of the characteristics of a so-called national system? Indeed, is it a realistic concept at all?

This paper attempts to provide a partial reply by means of an analysis of requirements and a summary description of the Canadian Geoscience Data Referral System, a developing operation which appears to meet the minimum requirements for a viable national system.

General Requirements

I would define a "national information system" as a set of coordinated policies, procedures and services for the management of specified information resources required to reach a national objective or to support a national activity. Canada has for many years approached the management of scientific and technical information (STI) from a national perspective (Brown, 1972) and among the various professional groups, geologists were probably the first in Canada to recognize clearly the benefits of this approach (Harrison, 1964) and to recommend specific action (Brisbin and Ediger, 1967).

However, it is one thing to recognize the desireability of a national system at the conceptual or "scientific" level and quite another to develop the strategy and tactics required to implement a successful operational system. Fiscal, institutional, organizational and other managerial constraints tend to be ignored at the outset or at best underestimated. To succeed, the system design must meet or exceed those general requirements demanded by any viable enterprise, among which the following are paramount:

1. Objective. The system must have an explicit, clearly understood objective, expressed in terms of the national needs to be served and the specific information resources to be included. A visible objective is essential for both development and management purposes.

2. System Concept and Model. Within the scope of the overall objective, the various major elements of the system

and their interrelationships must be identified at the conceptual level in order to derive a general model of the system. The model should also identify, in a generalized way, who and where the users are, areas of responsibility and the expected flow of information. As the system concept evolves in the course of practical development, it should be periodically reviewed and updated.

3. Management. The system must be manageable. This requires, in addition to an objective and model, an appropriate division of authority and responsibility, adequate fiscal and manpower resources, access to the STI involved, practical information technology, goodwill and cooperation among the various participating bodies, and a mechanism for monitoring results and progress; above all, the system must be compatible with the political framework in which it operates.

4. User Support. Recognition and support from the intended users is of course fundamental, yet it is common for system designers to take them for granted. The identification of potential users and their real needs is admittedly difficult because of differing perceptions, advancing technology, the effect of costs, product design etc., but a reasonable assessment must be made in order to take account of users in the system concept. If there are no users, there is no system.

5. Implementation Strategy. A successful system design will have taken account of the need for strategic planning in its implementation. The overall approach should be "top-down": First, the total information base should be explicitly identified and treated on a generalized basis; then as time and resources permit, individual sub-sets can be treated on a priority basis in increasingly greater levels of detail, i.e. general but comprehensive coverage is of more value than fragmentary detail.

Canadian Geoscience Data Referral System

1. Introduction. The Canadian Geoscience Data Referral System (CGDRS) is the new name given to the national information system that has evolved over the past ten years following the recommendation by the National Advisory Committee on Research in the Geological Sciences (NACRGS) to establish "a national system for storage and retrieval of geological data" (Brisbin and Ediger, 1967).

The focus of the 1967 proposal and subsequent operations was on the national coordination and utilization of computer-processable primary data, i.e. on computer-processable observations, measurements and other factual information. Following nine years' practical experience, including several shifts in emphasis and strategy (Burk, 1973, 1974), it became clear that little progress was being achieved in moving the geosciences toward development of an operational national system for primary data.

The reasons for lack of progress, as I was later to discover, could be traced to a failure to analyze and assess the original proposal in terms of meeting general system requirements, such as those listed above. For example, the objective was vague; key elements in the model were missing; some aspects could not be managed; and there was no strategy for implementation. It became necessary to take account of these and other deficiencies, face up to the constraints imposed by the "real world" and proceed accordingly.

This analysis of requirements and constraints led me to the logical conclusion that, for Canada, there was no prospect in the near term of developing a system for the coordination of primary data on a national basis. A choice then had to be made between pursuing a program that continued to deal directly with some aspect of primary data (but not on a national basis) or with development of a national information system that necessarily would assume a different form. The latter option was chosen and in mid-1976 the concept of CGDRS was put forward by the Canada Centre for Geoscience Data. The focus was shifted to secondary information. Overall, CGDRS addresses the question: "Where is the information?" and attempts to answer it through operation of a specialized national referral system.

2. Objective. The objective of CGDRS derives from a hierarchy of broader objectives of the Department of Energy, Mines and Resources, in particular an element called the Geoscience Information Transfer Sub-activity. Within this framework, the objective of CGDRS is:

> To provide, in a comprehensive, convenient and responsive manner, knowledge on the existence, location and nature of geoscience data available in the public domain which deal with the Canadian landmass and its offshore regions.

For purposes of this system, "geoscience" means the solid-earth sciences, including geology, geophysics, geochemistry, and various subfields thereof. Emphasis is placed on those geoscience data required for the study, exploration and assessment of non-renewable energy and mineral resources.

Use of the term "data" in the objective is intended to imply that, within the total spectrum of geoscience knowledge, priority is currently given to the identification and indexing of observations, measurements and other essentially factual information. However, other categories of geoscience information are not arbitrarily excluded and I expect that in future the more general term "information" will replace "data" in the system name.

The system concept through which the objective will be pursued includes the various elements blocked out in Figure 1. A summary description of each numbered element (3-13) follows.

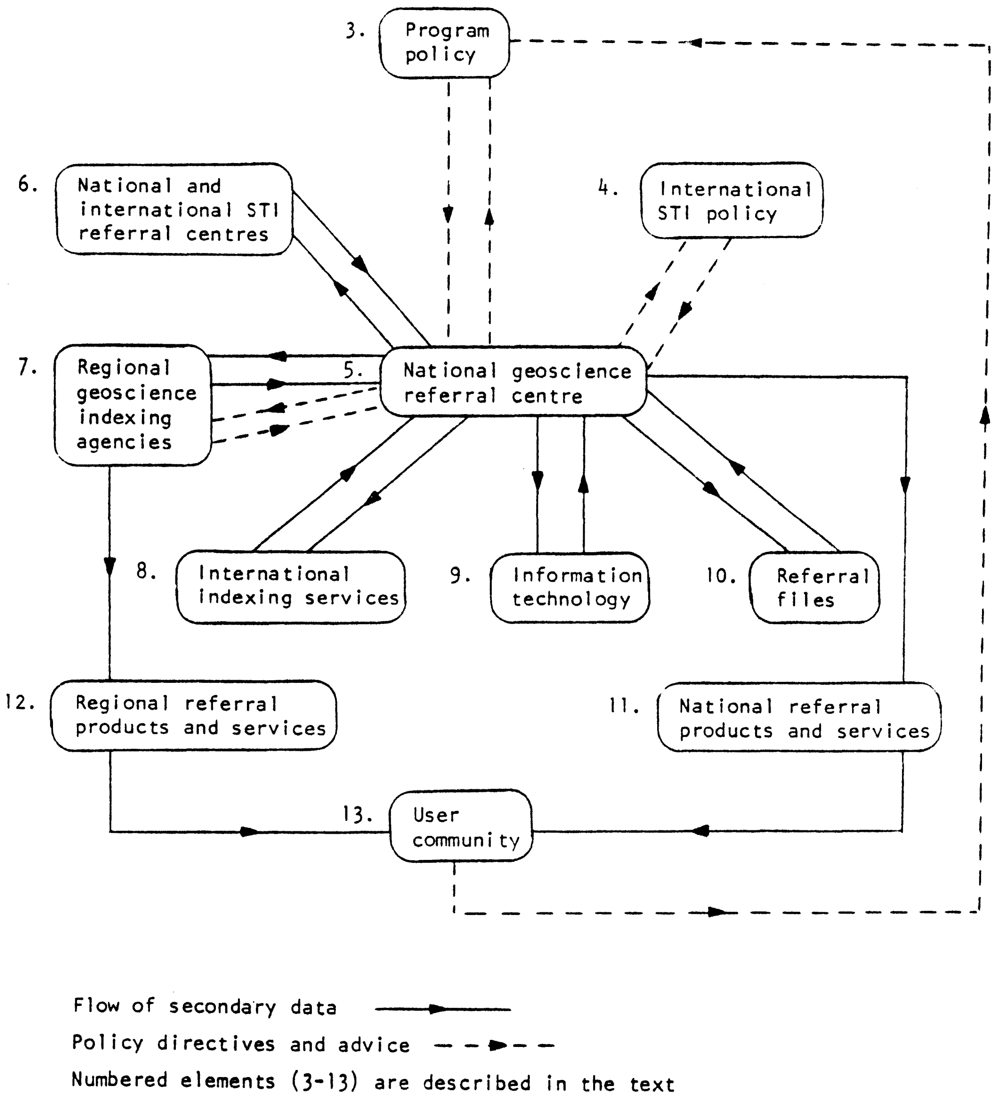


Figure 1. System concept: Canadian Geoscience Data Referral System (CGDRS)

3. Program Policy. Responsibility for overall system policy and operations rests with the Science and Technology Sector of the Department of Energy, Mines and Resources, on advice from the Canada Centre for Geoscience Data, the lead agency (EMR, 1976).

4. International STI Policy. Supplementary policy advice is received through participation in various international programs, particularly those of the Committee on Storage, Automatic Processing and Retrieval of Geological Data (COGEODATA), the Committee on Geological Documentation, both of the International Union of Geological Sciences, and the Committee on Data for Science and Technology (CODATA) of the International Council of Scientific Unions.

5. National Geoscience Referral Centre. The Canada Centre for Geoscience Data (CCGD), an independent division of the Science and Technology Sector, Department of Energy, Mines and Resources, serves as the national referral centre within CGDRS (Burk, in press). In this context it aspires to maintain a comprehensive overview of secondary sources and services of significance to Canadian geoscience; to develop and manage one of the major secondary files, namely the Canadian Index to Geoscience Data; and to offer a range of referral services, products and advice.

Resources available to CCGD during 1977/78 include a full-time staff of 6 and a budget of Cdn. $263,000. In addition to overall management of CGDRS, the Centre is responsible for various departmental STI projects, including the development and operation of text-oriented data management systems.

6. National and International STI Referral Centres. The national geoscience referral centre operates within a network of centres operating at various levels and in different fields. For example, at the national level, the Canada Institute for Scientific and Technical Information serves as the basic node for all STI (CISTI, 1977) while the recently established World Data Referral Centre operated by CODATA in Paris provides an international link.

The design of CGDRS presupposes such a network of interdependent referral centres and I anticipate that such a model will in fact form the framework for the operational national system to be developed within the scope of overall federal policy for STI in Canada (Katz, 1969).

7. Regional Geoscience Indexing Agencies. The identification and indexing of geoscience publications, reports, open-files, computer files and other sources is carried out on a voluntary, de-centralized basis by various Canadian organizations, which currently include federal geoscience agencies (5), provincial departments of natural

resources/mines/energy (6), research councils (1) and professional societies(1) (Gunn and Burk, 1975). Overall indexing priority has been placed on government-held reports and publications, but each agency establishes its own internal priorities and transmits the secondary data to CCGD in machine-processable form. As the quid pro quo, the Canada Centre for Geoscience Data provides coordination, indexing training, vocabulary control, methodology, file management services and product design assistance. Agency representatives meet regularly as members of the Geoscience Index Advisory Committee to deal with common problems and to advise CCGD.

8. International Indexing Services. Following the model suggested by Lea et al. (1972), CGDRS is exploring possibilities for augmenting the Canadian Index through exchange of bibliographic data with the general international services such as GeoRef (American Geological Institute) and Geoarchive (Geosystems). In the absence of an international program for geoscience documentation, CCGD would negotiate such exchanges on a bilateral basis.

This prospect raises the basic issue of responsibility for bibliographic control of Canadian-produced information in the geosciences. At present it is only partially exercised within Canada, and as pointed out by Woolston (1977), such a policy benefits neither Canada nor the rest of the world. CCGD will be developing policy proposals on this matter for consideration by Canadian authorities.

9. Information Technology. Bimonthly updating and day-to-day management of the secondary data produced by the indexing agencies is carried out centrally by the Canada Centre for Geoscience Data using an in-house file management system (PRIM-1) running on the departmental CDC Cyber 74 computer. Other information technology used routinely includes text-editing and photocomposition (Alphatext Limited), optical character recognition (OCR), computer-produced microfiche (Computrex Centres Ltd.) and computer-controlled photoreduction and printing (Campbell Corporation). Telecommunication facilities include desk-top and portable terminals which utilize the common carrier networks.

Current development activity includes the design of a computer-based system which will take account of all the various present and future processing and management functions carried out by CCGD. Following participation in a number of projects during 1974/76 (Vallée et al., 1977), the Centre is investigating the operation of a "computer conferencing" network for CGDRS agencies, possibly with links to documentation centres in the United States and the United Kingdom.

10. Referral Files. The Canadian Index to Geoscience Data is the principal referral tool produced by CGDRS (Gunn and Burk, 1975). In the "national system" concept of 1967 (Brisbin and Ediger, 1967, Fig. 4, p. 23-34) the Index was considered a peripheral element but it now serves as a central focus. From its beginnings in 1967 as a pilot project of the Geological Survey of Canada (McGee, 1968), the file has grown to 63,397 titles (November 1977) indexed by 13 agencies. Included are indexed references to published reports and maps (11,633), aeromagnetic maps (7,000), theses (194), resource assessment reports (26,090), mineral inventory summaries (15,008), open files (911), internal reports (2,202) and unpublished maps (359). Unpublished (but public) titles constitute 70 per cent of the Canadian Index file.

As implied by the file name, indexing has emphasized the identification of factual information dealing with the Canadian landmass and its offshore regions. Indexing concepts are carefully controlled by a thesaurus and geographic location by use of the National Topographic System (NTS) grid, which permits systematic geographic resolution of areas measuring about 64 x 48 km, depending on location.

11. National Referral Products and Services. A full range of output products and services from the national referral centre is not yet available, this being the last of the major system elements to be developed. Currently, the principal products are custom indexes derived from the Canadian Index to Geoscience Data for private clients and public agencies. Output is provided on a cost-of-service basis as photoreduced printout, photocomposed type, computer-generated microfiche or magnetic tape. Thesauri and other file management tools are provided to the contributing agencies.

During the remainder of 1977/78 and on into 1978/79, priority will be placed on development of products and services, notably public "online" access to the Canadian Index file, publication of topical and regional indexes, publication of a directory of data sources and operation of a service desk to provide advice on the use of secondary tools and services of value to Canadian geoscience.

12. Regional Referral Products and Services. All geoscience agencies participating in CGDRS have been encouraged to provide products and services for their own clientele parallel to those provided nationally. Provincial agencies in Ontario (Geoscience Data Centre), Nova Scotia (Department of Mines), Alberta (Alberta Research Council), and a federal agency (Department of Indian and Northern Affairs) have published indexes and provide support services for their respective regions. As more products are produced by CCGD,

more referral tools will become available for use by the regional agencies.

13. User Community. Current users of CGDRS are from the resource industries, including those exploring for petroleum, coal and metals, from research and development establishments in government, industry and the universities, and from other nodes within the referral network - the contributing agencies, libraries and international referral services. A study to identify users and their needs in specific terms will be carried out by CCGD during 1977/78.

Conclusions

1. A model incorporating a referral network provides the only basis for developing a viable national information system for Canadian geoscience.

2. The Canadian Geoscience Data Referral System (CGDRS), following ten years of planning and development, shows promise of maturing over the next 1-2 years as a fully functional, national information system, with potential for expanding beyond referral into other areas of information management.

3. Difficulty and delay in developing a viable national information system for the geosciences in Canada, and perhaps elsewhere, has been due mainly to lack of an explicit and realistic system model and to failure to apply certain basic managerial principles required by all successful enterprises.

4. The experience and results gained over the past ten years in Canadian geoscience suggest that establishment of a discipline-oriented national referral network, following the West German example (Institute for Documentation, 1976), should be investigated as a possible strategic basis for continuing development of the Canadian national system for scientific and technical information.

Acknowledgements. Any success or achievement associated with development of CGDRS, and hence the substance of this report, is the result of the encouragement, dedication and efforts of many individuals extending over a long period of time. Principal among these are the original proponents of the "national system" concept, led by Dr. S.C. Robinson (then of the Geological Survey of Canada); the committee that began implementation, led by Mr. D.A. Sharp (then of the Ontario Department of Mines and Northern Affairs); employees past and present of the Canada Centre for Geoscience Data, particularly Mr. B.A. McGee, Miss K.L. Gunn and Mr. G.A. Chappell; and colleagues from participating Canadian agencies, notably Mr. R. Paquet (Québec Department of Natural Resources), Mr. H.A. Groen and Mr. S. Sun (Ontario Ministry of Natural Resources), Mrs. M.A. Fuzesy (Saskatchewan Department of Mineral Resources) and Mr. H. Missan (Newfoundland Department of Mines and Energy). However, responsibility for the contents of this paper, and the views expressed, are solely my own.

References

Brisbin, W.C., and Ediger, N.M., Editors 1967, A national system for storage and and retrieval of geological data in Canada: Nat. Adv. Comm. Res. Geological Sciences, Ottawa (available from Geol. Survey of Canada), 175 p.

Brown, J.E., 1972, The Canadian national scientific and technical information (STI) system; a progress report: Nat. Fed. Science Abstracting and Indexing Services, Philadelphia, Rept. 4, May 1972, 26 p.

Burk, C.F., Jr., 1969, The national system for storage and retrieval of geological data in Canada: Proc. Geoscience Info. Soc., v. 1, p. 1-7.

Burk, C.F., Jr., 1973, Evolution of Canadian system for geoscience data, 1964-73: Proc. First Open Conference on Info. Science in Canada, Montebello, Can. Assoc. Info. Science, p. 138-147.

Burk, C.F., Jr., 1974, Planning document for the earth resources data system (ERDS): Science and Technology Sector Rept. 74-2, Dept. Energy, Mines and Resources, 33 p. plus app., Ottawa (unpublished).

Burk, C.F., Jr. (in press), Canada Centre for Geoscience Data, in Activities of the Science and Technology Sector 1976/77: Science and Tech. Sector, Dept. Energy, Mines and Resources, Ottawa.

CISTI, 1977, Canada Institute for Scientific and Technical Information, report 1974-77: Nat. Res. Council Canada, Ottawa, no. 16014, 42 p.

EMR, 1977, Annual report 1976/77: Dept. Energy, Mines and Resources, Ottawa, Cat. no. MI-5/1977, 36 p.

Gunn, K.L., and Burk, C.F., Jr., 1975, The Canadian Index to Geoscience Data: a decentralized, cooperative indexing project: Proc. Can. Assoc. Info. Science, Québec, p. 242-253.

Hambleton, W.W., Chairman, 1970, A concept of an information system for the geosciences: American Geol. Inst., Washington, D.C., Dec. 1970, 23 p.

Hambleton, W.W., Chairman, 1971, Toward a concept for a national geoscience information program: Geoscience Info. Soc. Newsletter, no. 15, Jan. 1971, p. 10-13.

Harrison, J.M., Chairman, 1964, Data compilation and processing, in Fourteenth Ann. Rept. 1963-64: Nat. Adv. Comm. Res. Geological Sciences, Ottawa, p. 12-13.

Hruska, J., 1975, Data-base management system for environmental geological data in Czechoslovakia, in Computer-based systems for geological field data: Can. Centre Geoscience Data, Geol. Survey Canada Paper 74-63, p. 94-98.

Institute for Documentation, 1976, The programme of the federal government [of West Germany] for the promotion of information and documentation (I & D - programme) 1974-1977: Inst. for Documentation (IDW), Frankfurt a. M., 1976, 125 p.

Katz, L., Chairman, 1969, A policy for scientific and technical information dissemination: Science Council Canada, Ottawa, Rept. 6, 35 p.

Lea, G., Diment, J., and Harvey, A.P., 1972, Geological literature, in Encyclopaedia of Library and Information Science, v. 9: Marcel Dekker, New York, p. 309-364.

McGee, B.A., 1968, National data index: Geol. Survey Canada Paper 68-1, Pt. B, p. 47.

Paladino, A.E., and Wachtman, J.B., 1976, An assessment of information systems capabilities required to support U.S. materials policy decisions: U.S. Congress, Office of Technology Assessment, Washington, D.C., OTA-M-40, 249 p.

Parkin, L.W., and Tellis, D.A. 1977, Australian earth sciences information system: Proc. Australian Inst. Mining and Metallurgy, no. 262, June 1977, p. 7-23.

Vallée, J., Askevold, G., and Wilson, T., 1977, Computer conferencing in the geosciences: Inst. for the Future Spec. Rept. SR-78, Menlo Park, California, May 1977, 75 p.

Woolston, J.E., 1977, The North American role in international information systems: Proc. Can. Assoc. Info. Science, 5th Can. Conf. on Info. Science, Ottawa, p. 8-20.

G E O A R C H I V E :

GEOSYSTEMS INDEXING POLICY

by

Graham Lea
Director
Geosystems
PO Box 1024
Westminster
London SW1
England

Geoscience Information Society Symposium
"Geoscience Information Retrieval Update"
November 7, 1977
Seattle, Washington, USA

GeoArchive: Geosystems Indexing Policy

Graham Lea, Director, Geosystems
PO Box 1024, Westminster, London SW1, England

The purpose of this paper is to outline, for the first time publicly, Geosystems' indexing philosophy. This is both easy and difficult. It is easy because the philosophical basis is rather straightforward, yet it is at the same time difficult because of the degree of indexing elaboration that has been achieved in the past ten years.

First, some clarifications. Geosystems are international geoscience information specialists and publishers of geoscience bibliographies derived from their computerised database, called GeoArchive. GeoArchive contains something over half a million references, and is growing at about 100,000 items/year. Access to GeoArchive is online via the Lockheed Dialog® Information Retrieval Service, or through Geosystems' profile service (1).

Before discussing in some detail the concept of indexing levels being used for GeoArchive, it is perhaps relevant to say a few words about how Geosystems see the role of secondary geoscience services, and Geosystems' attitude towards scope and sources. In a paper on Geological Literature in the Encyclopedia of Library and Information Science (2) we introduced a model for geoscience secondary services that has, at least, the virtue of simplicity. (Figure 1).

1. GENERAL SERVICES
   e.g. Geotitles Weekly. Referativnyi Zhurnal
2. NATIONAL AND REGIONAL SERVICES
   e.g. Abstracts of Belgian Geology and Physical Geology.
   Recent Polar Literature
3. SPECIALISED SERVICES
   e.g. Mineralogical Abstracts.
   Bibliography of Vertebrate Paleontology

Figure 1. GEOSCIENCE SECONDARY SERVICES

1. The general services aim at worldwide coverage of the whole literature, although naturally the services differ in their interpretation of this aim.
2. The national and regional services have as their objective the documentation of the geology of a region or a country.
3. The specialised services cover the literature of a particular field of geology.

The suggestion made in that paper was that the general services supply information to the national and regional services which should in turn inform the general services of missing sources. Specialised abstracting services cannot alone achieve the coverage of the general indexing services, so again there is a complementary role.

Geosystems' belief in this philosophy has lead to an approach to indexing that might be described as micro-indexing at a macrolevel of everything that has relevance to geoscience. The intention is therefore to provide highly accurate indexing that is useful to specialists, but lacks the volatility of ultimate detail and does not require dozens of specialists.

Now, a word or two on the scope of Geosystems' coverage. Geosystems have adopted a broad definition of geoscience and have used abstraction-complexity series to define the limits of geological knowledge. This approach is illustrated in Figures 2 and 3. Figure 2 shows the pure science series, whereas Figure 3 shows the effect of combining geology with mathematics, physics, chemistry, geology _sensu lato_, and biology in the interdisciplinary geoscience series. Geology is more narrowly defined (geology _sensu stricto_) and would include stratigraphy and structural geology -- perhaps the core of pure geology.

```
PURE SCIENCE SERIES
PURE MATHEMATICS abstraction
PHYSICS                ↑
CHEMISTRY              |
GEOLOGY sensu lato     ↓
BIOLOGY            complexity
```

Figure 2. ABSTRACTION-COMPLEXITY SERIES

Further discussion of this approach, including some recent work on epistemological models, is given in Chapter 1 of the lecture notes of a Geoscience Information Workshop given to the Australian Mineral Foundation (3).

INTERDISCIPLINARY GEOSCIENCE SERIES

MATHEMATICAL GEOLOGY	abstraction ↑
GEOPHYSICS	│
GEOCHEMISTRY	│
GEOLOGY	│
PALEONTOLOGY	↓ complexity

Figure 3. ABSTRACTION-COMPLEXITY SERIES

These studies have proved useful, not only for relevance studies, but also for the development of our thesaurus, Geosaurus (4).

So far as information sources go, Geosystems try to include as many as possible, whether they are formal, semi-formal or informal. The criteria for selection are geoscience information content, and public availability. We have now identified more than 11,000 geoscience serial sources, of which about 5000 are current (5). This catholic policy with regard to sources presents some interesting problems that will be the subject of a forthcoming paper. From the indexing viewpoint, universality of sources tests the indexing system. Our observation of geologists is that many are very interested in literature other than that published in formal journals, and that they are more likely to read less formal literature. For many geologists, the last ten years are not necessarily the most important. For this reason, Geosystems have underway a retroindexing project to index the major part of the estimated 2 million items not at present included in GeoArchive. So far, something over 50,000 pre-1960 items have been indexed.

GEOARCHIVE INDEXING LEVELS

Geosystems are strong believers in a structural approach to indexing. This implies dedication to a hierarchical, controlled vocabulary with adequate provision for dealing with synonyms and homonyms. At the same time, we are aware that there is a fundamental difference, albeit through a gradual transition, between descriptors, identifiers, and numerical data: Figure 4 shows this transition from generality to specificity.

	generality ↑
1. DESCRIPTORS (Geosaurus, Geosystems thesaurus)	
2. IDENTIFIERS (fixed & free format)	
3. DATA (predefined sets, subsets, fields, subfields)	↓ specificity

Figure 4. GEOARCHIVE INDEXING: LEVELS

Bibliographers are accustomed to dealing with a rather limited set of bibliographical elements. Yet geologists are actually interested in tabulations of data contained in a paper rather than the reference itself; Geosystems have therefore started to index the data contained in the literature, as distinct from indexing bibliographical elements. At present, this is an exacting manual process requiring considerable intellectual effort. But techniques can and will be developed to automate this process to some extent, especially when full text becomes more readily available as a result of computer typesetting.

Geosystems have therefore defined three indexing levels - descriptors, identifiers, and data. These will be discussed in turn.

The fundamental indexing of GeoArchive is carried out using Geosaurus, a hierarchical 8-level thesaurus of geoscience. In the controlled vocabulary series seen in Figure 5 Geosaurus is a thesauro-classification, at the highest end of the series.

1. UNCONTROLLED TERMS
2. LEXICON OF UNGROUPED CONTROLLED TERMS
3. LEXICON OF GROUPED CONTROLLED TERMS
4. ALPHABETICAL THESAURUS
5. CLASSIFICATION
6. THESAUROCLASSIFICATION

Figure 5. CONTROLLED VOCABULARY SERIES

The advantages of such a structured vocabulary become apparent to users soon after they start serious use of GeoArchive. Many 'thesauri' have only an alphabetical presentation with one level up and one level down (i.e. broader and narrower terms) from a term in the alphabetical list, and no systematic, hierarchical presentation of the thesaurus. This leads to several problems:

1. Synonyms are not controllable.
2. The way in which a subject is organised is not apparent, so that users must either search the whole of the alphabetical list, or hope that the terms that come to mind will give exhaustive coverage of their topic.
3. The benefit of being able to include the whole of a subject area, broad and narrow terms, or just part

of it, in a search, is lost.

4. Most alphabetical thesauri do not have code numbers for each term; these are particularly useful in online searching since typing problems are reduced, spelling and language difficulties eliminated, and ranges of numbers can be given.

Figure 6 shows the two series and nine subseries which make up Geosaurus.

000000/999999	GEOSAURUS SUBJECT SERIES
100000/199999	Publication form subseries
200000/989999	Specific subject subseries
990000/994999	GeoArchive indexes subseries
(000000/999999)	GEOSAURUS AUXILIARY SERIES
(000010/000099)	Language subseries
(001000/009999)	General location subseries
(010000/099999)	Stratigraphic subseries
(100000/119999)	International geographical subseries
(120000/949999)	Geographical subseries
(950000/989999)	Paleogeographical subseries

Figure 6. GEOARCHIVE INDEXING: LEVEL 1 - DESCRIPTORS

Of interest to many people is the ultimate size that we envisage for Geosaurus. There are no theoretical limits of course, but to be realistic, all terms should be familiar to geologists with a good grounding in all disciplines of geology. We therefore do not expect Geosaurus to grow beyond 10,000 terms, although supplementary lists of minerals and taxa could make it larger.

We have recently extended our _Geosaurus_ code numbers to six digits in the 4th edition, mostly by adding zeroes to the end of the four digit numbers. Figure 7 from the older third edition of _Geosaurus_ shows part of the petrology section.

5460 Volcanic rocks

The classification used for 5461/5469 follows Turner and Verhoogen (1960), p.71-72.
See also: 4610 Endogenetic mineralisation; 6660 Igneous geochemistry

5461 Basalts
Undifferentiated basalts

5462 Olivine & trachybasalts
Includes olivine diabase, picrite basalts oceanites, ankaramites, mugearites

5463 Tholeiitic basalts
Includes diabases

5464 Spillites & keratophyres
Includes albite diabases and pillow laves

5464 Nepheline basalts
Includes nephelinites, nepheline basanites, nepheline tephrites, leucitites , leucite basalts, leucite basanites, leucite tephrites, melilite basalts, limburgites

5466 Lamprophyres

5467 Trachytes & phonolites
Includes porphyries, tinguaites

5468 Andesites & rhyolites
Includes porphyrites, latites, trachyandesites, dacites, quartz porphyries, obsidians, pantellerites

5469 Pyroclastics
Includes vent agglomerates, bedded agglomerates, volcanic breccias, ashes, tuffs, crystal tuffs, lithic tuffs, vitric tuffs, welded tuffs, ignimbrites.

Figure 7. GEOSAURUS GEOLOGICI

Identifiers, the second level of GeoArchive indexing shown in Figure 4, are intermediate between thesaurus terms (descriptors) and pure data. In GeoArchive, identifiers can either be of fixed format, in which case they are put in brackets as title extensions with an identifier code, or completely free format. The version of GeoArchive available in the Lockheed Dialog system has all identifier codes translated, so that the codes shown in Figure 8 are translated into conference preprint, new species, new mineral, new stratigraphic name, obituary and portrait, and biography respectively.

FIXED FORMAT IDENTIFIERS

[*CPRE[14-9/13:9/15] 77]
[*NSPE Urocythereis phantastica]
[*NMIN fedorovskite]
[*NSTR Kirkby Range Member]
[*OBIT *PORT HEINTZ, A 1898-1975]

Figure 8. GEOARCHIVE INDEXING: LEVEL 2 - IDENTIFIERS

This technique allows many special indexes to be produced and automatically typeset in publications derived from GeoArchive. Figure 9 shows examples of two of these indexes: one is an index of new taxa and the other is a biographical index.

Free format identifers are uncontrolled, as their name suggests. Figure 10 shows some examples. One use of identifiers is to provide lists of candidate terms for fixed format identifiers, or for Geosaurus.

INDEX OF NEW TAXA

Apodops pricei gen. nov. sp. nov 146
Ardynomys glambus sp. nov 249
Ardynomys vinogradovi sp. nov 249

Beautomns bisus gen. nov. sp. nov 249

Chroniosaurus dongusenis gen. nov. sp. nov 108

BIOGRAPHICAL INDEX

BERTAUT, EF biography 726
BROILI, F biography 793

NACKEN, R obituary 801

SOWERBY, J 1757-1822 biography portrait 796

Figure 9. SPECIAL INDEXES

Equipment:	FEU-28 photomultiplier
Trade names:	Dr Dollar's clinometer
Organisations:	University of London
Nomenclature :	Calymene, vesuvianite, Piha Formation
Code names:	UNISIST
Acronyms:	Fortran
Abbreviations:	NERC

Figure 10. FREE FORMAT IDENTIFIERS

In many ways the most interesting development in GeoArchive indexing has been a quiet revolution: the extraction of certain categories of data from documents. The procedure is seen in Figure 11. Data of a certain category are arranged in a set, consisting of subsets, fields and subfields (and we have already distinguished two additional smaller divisions). The procedure is being expanded rapidly at present, but two examples will show what can be done. The first is a more familiar semi-biographical data set for geological maps, seen in Figure 12. Each element is separately retrievable. A detailed paper describing Geosystems' work on a collection of 100,000 geological maps in the Institute of Geological Sciences Libraries was given to the Geography and Map Division of the Special Libraries Association in New York last June.

{SET 1} {SET 2}

{SUBSET 1; SUBSET 2; etc}

{field 1, field 2, field 3; etc}

{subfield 1, subfield 2, subfield 3; etc}

Figure 11. GEOARCHIVE INDEXING: LEVEL 3 - DATA

{Accession number, location, issue, year;
form; scale; edition; related items; latitude:
longitude; publication number; abnormal conditions}

{A121189, V2480, 30/14-4, 1976;
F7;63360; 2; 120636, 120637, 120638;
52.20N-53.00N : 4.30W-5.00W; ;Z2}

Figure 12. GEOLOGICAL MAP DATA EXAMPLE

Figure 13 shows an example of the kind of data which can be obtained from the literature at the same time as the paper is indexed.

{material; Geosaurus location code, area; quantity, units, time; ore, quality...}

{*Cr; (690000), Ingessana Hills; 16000, t, 1976; chromite, 54, 3.5, 0.1}

Figure 13. MINERAL PRODUCTION DATA EXAMPLE

The retrieval possibilities are exciting. Questions could be structured to ask:

1. Where are the chromium minerals being produced?
2. What minerals are produced in the Sudan?
3. What is the variation of iron-chromium ratio in chromite, worldwide, and the amount produced?
4. Is chromite the only commercial chromium mineral?
5. What are the major serial sources of information about chromium production?

If this three-level indexing approach covered all forseeable requirements, there would still be many years of work to be done in refining it -- the progress of geology would ensure that this would never stop. But it is already clear that there is a need for a fourth level of indexing. This level is the automatic extraction of new ideas from the full text of papers, using semantic analysis and artifical intelligence techniques. Putting aside the very real problems of producing good abstracts, it is apparent that an abstract is likely to contain an unstructured mixture of data and ideas. The data indexing technique which Geosystems have developed is a rather superior alternative to conventional data abstracting in that the data can be manipulated by computer. Research is now in progress on methods of automatically distinguishing and extracting new ideas from the full text of papers. As a progress report, it can be said that there is little doubt about the feasibility of the project: what must be refined is the quality and efficiency of the process. When this fourth level of indexing has been realised, the symbolisation of knowledge will be a reality.

REFERENCES

1. LEA, G et al
   GeoArchive Users' Guide
   Geosystems, London, 1978, 2nd Edition

2. LEA, G; DIMENT, J & HARVEY, AP
   Geological literature
   In: Encyclopedia of Library and Information Science
   Marcel Dekker, New York, 1973, p309-364

3. LEA, G
   Geoscience Information Workshop
   Australian Mineral Foundation, Adelaide, 1976

4. CHARLES, R et al
Geosaurus: Geosystems' thesaurus of geoscience
Geosystems, London, 1978, 3rd Edition

5. SHEARER, J et al
Geosources: Geoserials and Geopublishers
Geosystems, London, 1978, 3rd edition

A COMPUTER AID TO THE DISTRIBUTION OF GEOLOGIC PUBLICATIONS

Jo Anne DeGraffenreid[1] and Harold W. Cable
Kansas Geological Survey, Lawrence, Kansas 66044

ABSTRACT

Publications issued by specialized groups such as regional geologic societies may have limited circulations. These smaller publishers have not used computers to maintain mailing lists, prepare labels, or issue invoices because the small volume makes automation seem impractical. Increased postal costs and decreased computer charges now make computer-aided distribution of publications cost effective for the smallest geological organizations.

The Publications Control System (PCS) is a mini-computer program to maintain and update mailing lists, generate labels, and create invoices for publications of the Geologic Research Section of the Kansas Geological Survey. PCS includes an interactive text editor, allowing a typist to build or modify a file of mailing addresses with a remote terminal. The file may be sorted by identification number, alphabetically by name or state and country, by zip code, or by a special code. The latter allows different mailing lists to be combined into a single file and later retrieved by type. Any subset of a file can produce a membership or subscription list, pressure-sensitive mailing labels on pin-feed forms, or invoices. Files currently are maintained for three categories of recipients, including the membership of the International Association for Mathematical Geology.

PCS is used with a mailing procedure which heat-seals publications in plastic bags, providing some protection against postal system vicissitudes. Cost of preparing a label and invoice and bagging a publication is below 15¢ per copy--less than conventional procedures. PCS will operate on most mini-computers having FORTRAN, a disc, and lineprinter.

INTRODUCTION

The biggest jobs most small scientific societies do are collect dues, mail publications or arrange journal subscriptions, and periodically contact their members with ballots, newsletters, and announcements. The maintenance of membership lists, preparation of invoices,

[1] Speaker.

and printing of mailing labels can all be done by computer, with great savings in time and effort. This is routinely done by commercial firms, universities, and the larger scientific societies. However, few small societies have utilized computers for these tasks, because they usually have no professional staff, no headquarters with great resources, and officers are usually "amateurs" who change frequently. Most small groups have not had the time to invest in automated mailing systems, nor the facilities on which to implement them. Now, however, computers--especially minicomputers--are widely available and accessible to almost any group at a modest price. Unfortunately, most small groups still do not have the time or talent necessary to develop an automated mailing system--especially if the program must be rewritten for different computers every few years as officers change.

What is needed is a software package, usable on almost any computer, including the very smallest minicomputers, that is adequate for the needs of a typical small scientific society. Although many mailing systems are available, most are "too much" for a small society. Program requirements for an automated mailing/billing system for a small society are: (a) it must be simple and efficient, to fit within the smallest of machines, and use the most general forms of computer coding; (b) it must be easily transferable from small to large machines, from one brand to another, from place to place; and (c) it must be extremely simple to use, by nonspecialists, with an absolute minimum of instruction and training.

These requirements could be met in various ways. Our approach was to design the software to operate in ANSI standard FORTRAN widely used on 32K and larger minicomputers. Make extensive use of mass storage devices for the lists--preserving core for manipulation and I/O. Such a system would be directly utilizable on bigger computers and would be transferable to most minicomputer systems as well. Make the system modular, separating editing, sorting, and listing functions. Use a query/command structure so the user can interact with the program, and if necessary, learn from it.

It seems appropriate that the International Association for Mathematical Geology should commission the writing of a program for automated mailing and handling of its membership lists. This organization is deeply committed to the application of computer methods to the solution of geologic problems--and keeping *their* membership rolls straight is definitely a problem!

In addition, the Office of the Western Treasurer of the IAMG is located at the Kansas Geological Survey, which has a large staff devoted to computer technology. The Survey has recently installed its own minicomputer system which is similar to those used in geology departments and governmental and industrial offices. Software developed for our system should be usable elsewhere.

The computer program which the IAMG desired would keep a file of all members' names and addresses, coded to indicate the amount of dues paid and the journal to which each member subscribes. Names and addresses would be added, deleted, or modified as desired. Upon request, the program would prepare invoices or receipts, and print mailing labels and membership lists in a variety of formats. This program would also serve as a prototype of a much larger system to handle the entire publication distribution and billing operation of the Kansas Geological Survey. The Survey distributes hundreds of thousands of publications a year.

THE PUBLICATIONS CONTROL SYSTEM

PCS (Publications Control System) is a FORTRAN IV program designed to meet the requirements just stated. It runs on a Data General Nova 1200 computer with 32,000 words (32K) of memory. The computer is equipped with two tape drives and a disc for mass storage of data files and programs. Data, including mailing lists, are entered directly into the computer through CRT remote terminals. At the present time, we are using a Tektronix 1440 display scope attached to a Varian electrostatic printer/plotter; we can obtain hard copy automatically. Eventually we'll have terminals at the desks of users, so payments, orders, and changes of address can be entered immediately as they are received by mail or by telephone.

The Publications Control System actually consists of two separate programs. The first, called MAILIST, builds the mailing lists and is used to edit and maintain the computer files. It consists of 298 FORTRAN statements, many of which are messages that are transmitted interactively to the user. The second program is called MAILABEL. It prints out membership lists on the line printer or generates mailing labels. MAILABEL contains 449 FORTRAN statements, mostly in sorting routines and I/O commands.

MAILIST uses a large random disc file to store names and addresses. No ordering is done on this file; it is simply maintained as it is read in. The program communicates only with the remote terminal and with

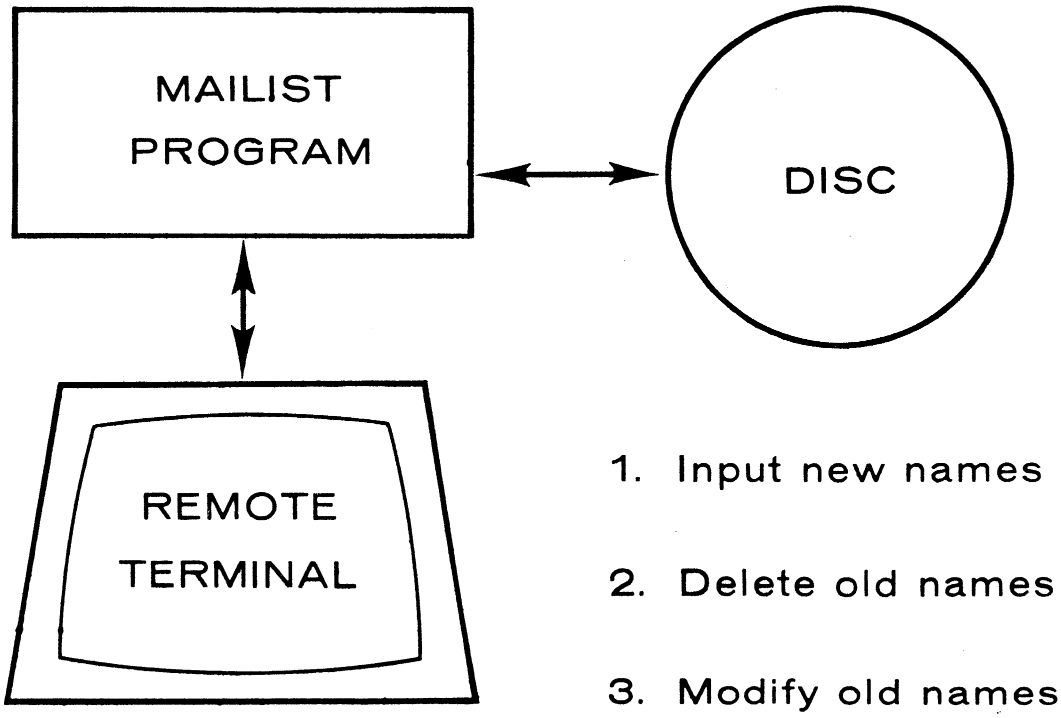


FIGURE 1. Devices accessed by MAILIST and three basic functions performed.

the disc unit (Fig. 1). MAILIST allows three basic functions to be performed: insert a new name and address; delete a previous name and address; modify a previous name and/or address.

MAILIST data are entered in a nine-line format; each line may contain up to 40 characters. The data are read one character at a time in A-format, so numbers, letters, and special characters may be used anywhere within the address. The nine-line structure is outlined in Table 1. The only error checking done by the program at this point is on the second line. Characters are scanned until a comma is encountered, followed by a space. If all characters preceding the comma are letters, the character string is accepted as a last name. [Names must be entered LAST NAME, COMMA, FIRST NAME, MIDDLE NAME(S)/INITIAL(S).]

Within the computer, the lines are grouped in sets of three, forming a character string 120 characters long. The strings are stored one to a record on the disc. This speeds FORTRAN I/O on our particular minicomputer; other record lengths might be more appropriate for other machines.

TABLE 1. Nine-line structure for MAILIST data entry. Each line may contain up to 40 alphanumeric characters, including special symbols.

FORMAT	DATA
1. Membership code	*1A*
2. Name	*Jones, Adam B.*
3. Address (Part 1)	*Dept. Geology*
4. Address (Part 2)	*114 Deike Building*
5. Address (Part 3)	*Suburbia University*
6. City	*Suburbia*
7. State/Province	*KS*
8. Country	*USA*
9. Zip/Postal Code	*60006*

MAILABEL uses the terminal as a control device to accept the user's instructions. Output, in the form of printed lists, printed invoices, or mailing labels, is written on magnetic tape or produced on a line printer. The MAILABEL user must answer three series of questions which the program asks through the remote terminal (Fig. 2). The first determines which names and addresses are to be put in the output file. The second establishes how these are to be sorted (alphabetically, by state or country, etc.). The third determines what type of listing is desired and how it should be produced.

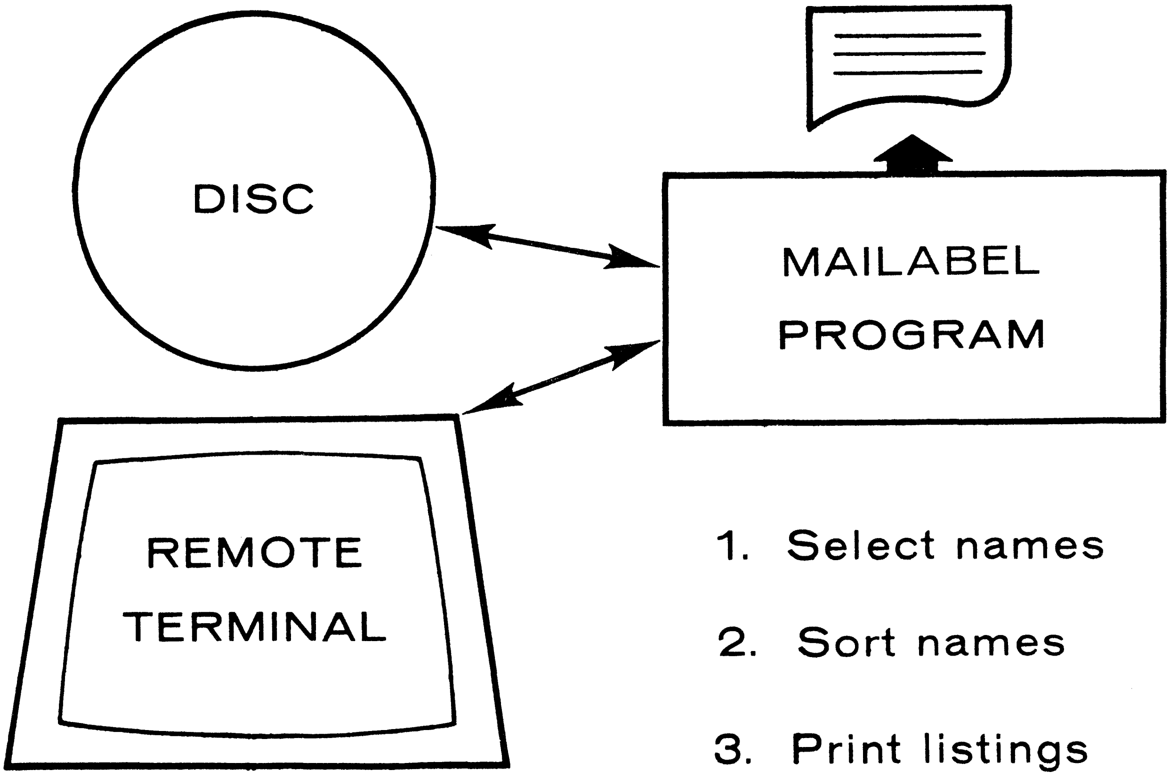


FIGURE 2. Devices accessed by MAILABEL and three basic functions performed. Output may take the form of printed listings, invoices, or mailing labels.

The first line of an entry can contain up to 40 alphanumeric code values, each representing a specific category of membership or mailing. In this way, the IAMG membership list can be separated into those who receive either or both of two journals. The mailing lists of the Geologic Research Section of the Survey are also randomly intermixed with the IAMG files. If desired, a complete list of persons interested in mathematical geology can be generated by combining all files. Lists can also be generated of those persons common to two or more files, or unique to a single file, or characterized by any other logical combination. In theory, as many as 1440 different categories can be represented in the code line.

In addition to selecting by code category, MAILABEL will sort the entries found in several different ways: alphabetically by member's last name, using the information from line two of each entry; alphabetically by country; or by state or province. Names can be arranged alphabetically within each country or state. It is also possible to sort by Zip code. [A word of warning! You'll never retrieve all your Dutch members if their addresses are recorded "The Netherlands," "Netherlands," and "Holland"! A lot of time can be spent weeding out inconsistencies. Decide at the onset, for instance, if you want to use two-letter state designations for U.S. addresses.]

Once a list of names and addresses has been selected from the random disc file and sorted into the desired sequence, MAILABEL will print this information in a variety of forms. Invoices can be prepared, incorporating pre-specified billing messages. Membership lists can be generated sequentially or arranged in multiple columns on a page. The code line may be suppressed, if desired, and names can be listed first name or last name first.

Perhaps the most useful form of output is that printed on die-cut, pressure-sensitive mailing labels arranged on a pin-feed paper backing. MAILABEL will accommodate various forms and sizes of labels; we commonly use a 1 1/2" x 4" label for mailing of newsletters and flyers. Survey journals are heat-sealed in transparent plastic bags for mailing. An oversize label, 3" x 6", with a special adhesive back, is used with these bags. The large size permits us to put the postage on the label, as it will not stick to the bags. Incidentally, we have found this method to be almost post office-proof. The plastic bags will not rupture or tear and must be opened with knife or scissors. If the publication fits tightly within the bag, it is rigid and fairly resistant to having its corners knocked off. The special labels have an adhesive used in the frozen food packaging industry and will not peel off, even when frozen or subjected to high temperatures. The seamless plastic tubing costs less than conventional Kraft envelopes, the plastic bags are easier and safer to handle, and publications mailed in them have a significantly higher survival rate. Although the IAMG and the Kansas Geological Survey are pleased with PCS, our automated mailing system, I imagine the post office will outlaw it--it works too well!

THE USE OF LANGUAGE PROCESSOR CONCEPTS TO DESIGN AND IMPLEMENT GEOGRAPHICAL INFORMATION SYSTEMS

(Presented at Ninetieth Annual Meeting of the Geological Society of America and its associated societies)

by

William A. Gates*

Brent H. McCown

B.J. Niemann, Jr.

November 1977

*Author's Address: Institute for Environmental Studies
610 N. Walnut Street
University of Wisconsin
Madison, Wisconsin 53706

Acknowledgments

This paper and work reported in it were made possible through support from several sources: (1) the Rockefeller Foundation, Grant #RF74099, to the Institute for Environmental Studies, University of Wisconsin; (2) U.S. Department of Agriculture, HATCH Multidisciplinary Grant #2173 to College of Agriculture and Life Sciences, Department of Landscape Architecture, University of Wisconsin; (3) University of Wisconsin Graduate School; and (4) State of Wisconsin, Department of Administration, Office of Energy and Planning.

Curtis Wilcott, now with U.S. Fish and Wildlife Service, Fort Collins, Colorado, working on GIS, deserves credit for encouraging and promoting incorporation of language processor concepts into work on GIS. As original principal investigator of the HATCH Project, his support was essential.

Introduction

The application of computers to questions requiring spatial analysis has a long history. Tomlinson is perhaps the best single source of information on the subject. The University of Wisconsin has many notable examples of such analyses where the computer has been an indispensable tool. As an example, in 1972 resource data were collected and analyzed to illustrate optimal routing, as criteria varied, of an interstate highway from Milwaukee to Green Bay (Miller, et al.). More recently, an interactive grid system, GRASP (Geographical Resource Analysis Software Package), has been used to evaluate different waste disposal technologies in northern Wisconsin (Butler, et al.). By providing the decisionmaker with quick response and feedback, and with a comprehensive database (McCown, et al.), GRASP also demonstrated the contributions that interactive computing could make to problem solving by using a simple, consistent user language supported by advanced data structures for efficient execution.

What is common to efforts such as these is that a limited geographical area is studied intensively and when the study terminates the data collected seldom serves other purposes. It may also be observed that most of the effort goes into data collection, often from inadequate sources. These observations may also be made of non-research institutions, such as the state and federal governments. Enormous expenditures are made for data collection for legislatively mandated purposes, but the application of the collected data to other purposes is fraught with difficulties. Geographical information systems (GIS) represent an attempt to provide a general framework into which data may be deposited and subsequently interrelated, and a means to answer the unanticipated, urgent question.

Our research (Wilcott and Gates) into the nature of GIS has revealed two primary problems: (1) inadequate or incomplete capabilities, and (2) a lack of evolvability. For example, few systems were complete because they did not provide for keeping collected data up-to-date. And in that new data could not be added, they were not evolvable. We also concluded that the users were left out of the design and implementation phase and that this omission contributed to incomplete or inadequate systems. Everest provides an excellent discussion of the desirable qualities of information systems and the consequences that await systems without such qualities. These qualities include sharability, integrity, availability and evolvability. Ross summarizes evidence that suggests that "analysis and design errors" account for 64% of user objections to large software systems.

Recent work on GRASP has been combined with research to produce the geographical information system capability of handling and manipulating line-drawing data that may be taken from maps, aerial photographs, satellite coverage, etc. The result is to be a hybrid system, representing the best features and capabilities of both grid and line-drawing systems. The language processor concepts that we describe, albeit briefly, in this paper have been essential. They have allowed us to remain flexible and evolvable.

Language Processor Concepts

Many computer programs are difficult to use. Users have long complained about how frustrating it is to get the "input" right. This situation has been changing gradually, but there are still as many difficult-to-use programs as there are easy-to-use programs.

The improvement has come about as programmers have been provided with better and better "high-level" programming languages. These languages provide superior facilities for problem solution. They also remove many tedious details that have to do with making the computer work and have nothing to do with expressing a solution to the problem. The original computer programming language--machine language--required expressing an entire solution as a string of ones or zeros. The next language--assembly language--provided for mnemonic labeling of machine instructions such as elementary operations like addition and subtraction. FORTRAN, regarded by most as the first high-level language, improved upon machine language in that it provided FORmula TRANslation facilities that made it possible to write computer programs in an algebra-like notation. A FORTRAN program would be translated to machine language by a compiler. As the "high-level" language made it easier for the programmer to write programs, more attention was given to making things easier for the users of those programs. But now an extra step was necessary, that of compiling (translating). Development of the early compilers took a team of ten persons two years to complete, for a 20-year total development effort. This investment was quickly repaid by the savings realized in implementing programs more quickly that were easier to use.

The formal specification of high-level languages was at first ad hoc but now relies heavily on formal mathematical concepts of language. The compiler writers use the specification to determine what language their compiler must translate to machine language. The programmers use the specification to determine how they can write in the language. Chomsky's (1956) formal definition of basic language types is in many ways responsible for this positive state of affairs. These definitions, among others, have led to further improvements so that now a sophisticated programming language can be implemented by a single individual in a single year.

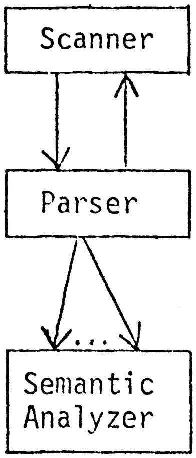
In the progression from machine language to assemblers to high-level languages, the recognition has been made of how important language is to speedy and correct implementation of computer programs. It seems reasonable that the utilization and application of the concepts that produced a 15- to 20-fold improvement in compiler construction should do the same for constructing GIS. Additionally, the application of the concepts should produce easier-to-use systems and realize the same benefits for users that programmers have realized in high-level programming languages.

A complete reconstruction of the developments that have led to the availability of language constructor tools can be gained from Aho and Ullman. Presently, two quite commonly available tools--scanner-generators (SG) and parser-generators (PG)--can automatically construct tables that can in turn be used by a language processor to recognize a language of great power and flexibility of expression, yet one that is economical in terms of human understanding and use. The language processor handles all the tasks of checking and correcting user input.

The SG requires as input a simple description of the "words" in the language to be recognized. The PG requires as input a description of the grammar of the language. The grammar specifies word order and relationships among "sentences" of the language. Gries (pp. 12-15) provides a good orientation to the grammar notation called Backus-Normal Form (BNF) with an example applied to English sentences. This notation is used as input to many parser-generators for language definition.

The tables produced by these two generators are then used by the language processor. Very simply, the language processor has three parts: scanner, parser, and semantic analyzer. The parser controls the language processor. Under parser control the scanner will, by using the tables produced by the SG, isolate and identify the next input "word" and return it to the parser. The parser solicites input from the scanner until it has recognized, by using the tables produced by the PG, a "sentence" in the language, somewhat similarly to diagramming sentences in English. The parser, having recognized a "sentence," calls the semantic analyzer providing it with the "sentence." The semantic analyzer takes the appropriate action. These various actions have been coded by the programmer, who is now concentrating only on problem solving. Figure 1 summarizes this description.

In the latter discussion we did not mention GIS--an omission that was intentional. The concepts have wide applicability and should not be restricted to GIS any more than to the design and implementation of programming languages. In the next section, however, we shall discuss language development process as applied to GIS.



Reads user input, identifies "words," and returns them to parser.

Accumulates "words" until sentence type recognized, then calls appropriate semantic routine in semantic analyzer.

Carries out actions as dictated by recognized sentence.

Figure 1. Relationships between scanner, parser, and semantic analyzer

Language Development Process

"Step-wise refinement" is a popular term used in computer science today to describe the process of program development. The concept suggests that a programmer, in constructing a solution to a programming problem, begins with a very simple model or statement of the problem: given these inputs, manipulations will produce the following outputs. The solution is arrived at by iteratively refining the previous stage of problem and solution statement. At each iteration the programmer is aware of many ways to proceed to solve the problem. Choices must be made based on constraints in time and space. This trade-off is made both in presenting a solution to the problem and in the actual programmed solution. In the most general sense the step-wise refinement model is a familiar one. The language development process we have used is quite similar except it involves many individuals over a long period of time. Figure 2 summarizes the process. After a GIS development group

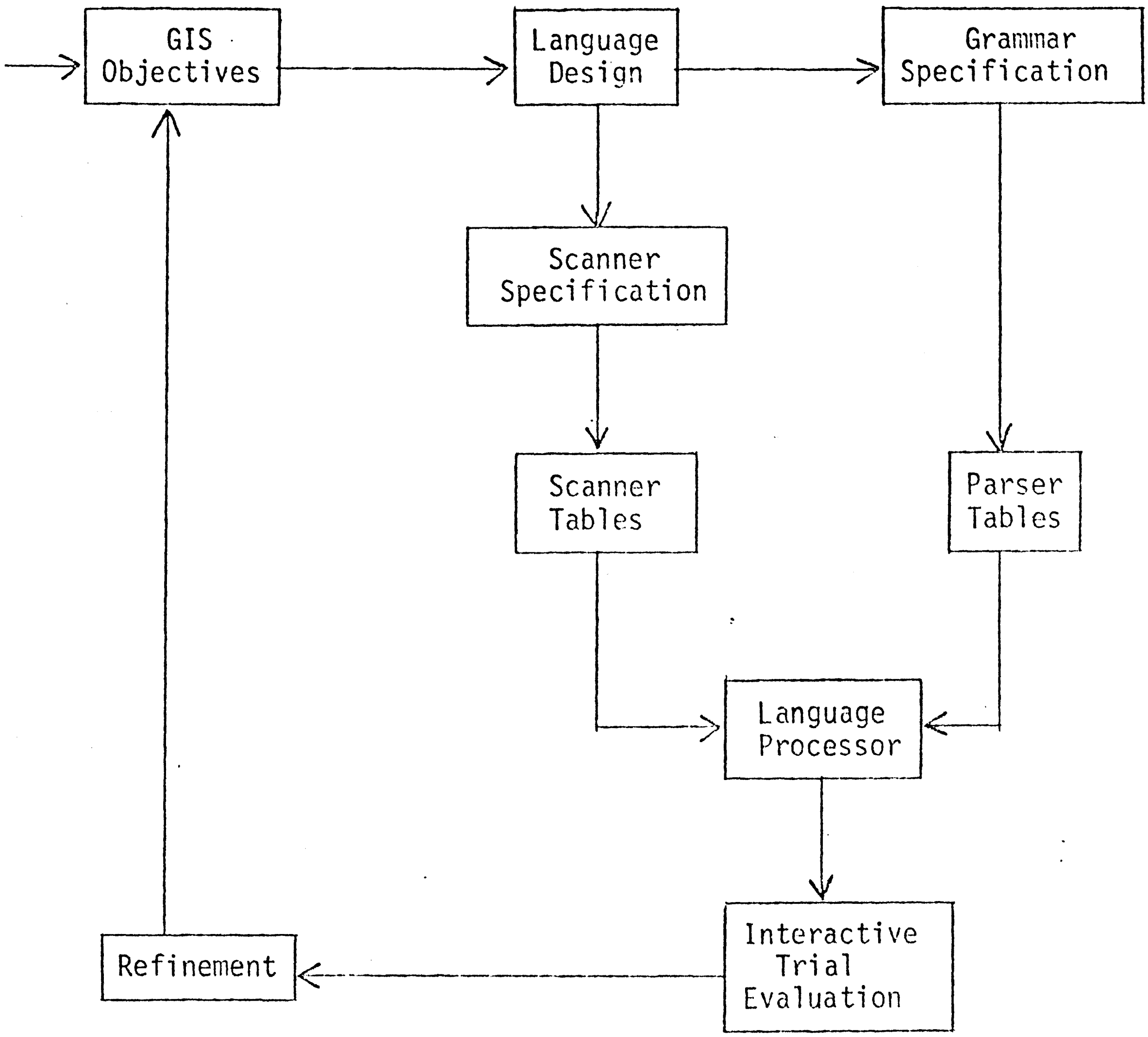


Figure 2. Language development process

has devoted suitable effort to establishing its objectives, a grammar is specified that embodies the manner in which users wish to go about dealing with the system. A few hypothetical examples detailing possible steps in the process are examined for their consistency. The implementation team provides some direction both in consideration of data structures and algorithms required and in knowledge of what types of language features can be easily used by people and computers. What then exists is a fairly well-understood idea of what is to be accomplished. The challenge is to keep the level of understanding the same while the refinements become more numerous.

Frequently, at this point the designers and implementors finish the job without further interaction with the users. The language development process, however, permits the designers and implementors to spend a few hours (1) formalizing a scanner and a grammar, (2) producing tables, and (3) conducting their own trials of the language. No time is spent on implementation. Most energies are directed toward discussion and research of algorithms and data structures to support solution. Thus, in a very short time the users are presented with a working model of the language they helped to formulate. Simultaneously, an analysis of the difficulty of implementation plus new ideas for extension or modification can be considered. The users also conduct language trials of their own. The designers and implementors provide interpretation of what is happening as the various language features are tried out.

The first step thus concludes with joint evaluation, analysis, and new language specification, and a second iteration may begin. Generally, this new language will be closely related to the trial version, although insights into the problem gained from trial-experience may have caused it to be abandoned. Since no investment beyond a few hours has taken place, this abandonment presents no real problem. Documentation supports each iteration, especially as the details of solution become more numerous.

Contributions of Language Processor Concepts to Solving Design and Implementation Problems

There are four major points to be appreciated about the language development process. First, very little time or effort is required to produce a clear understanding of what the result is to be. Second, no resources are devoted to implementation until the user's view of what is to result is well understood. Third, the implementation process has become much simpler and organized according to the rules of the grammar. And finally, the language development process can be used during implementation to further assure the correctness of the result.

Involving the users systematically in the process of language development avoids many design errors. Repeated language trials prior to implementation, where a clear understanding between the various groups can be documented, reduces costs of changes to a minimum. It is very easy to change a language, but very expensive to alter an implementation because algorithms and data structures take substantial energy to develop and consequently to change. The final product has been implemented and tested parallelly by all groups. This will certainly make it more acceptable to users. The language itself is an excellent way of managing the problem of new individuals becoming involved at different stages of the process. A number of rules of the grammar being applicable throughout the language make it fairly simple to learn.

The most significant contribution of language processor concepts may be the way they permit the language development process to guide GIS design, but they also provide enormous technical benefits to implementation. Absolutely no implementation effort is devoted to handling and checking user input. At any stage where semantic action is called for, the input is valid. The language processor detects any errors automatically and directs the user to correct them. Another advantage is that the data and software are quite independent of each other. The pieces of machine-dependent code can be isolated in utility routines quite easily, thus assuring greater portability. The language is definitely independent of any particular machine because the scanner and parser tables are.

The grammar also provides the capability of implementing sophisticated language features without presenting the implementors with immense tasks. Program coding can be broken into small units, semantic routines. Coding tasks correspond on a one-for-one basis with the rules of the grammar. These rules may be grouped to recognize sentences but many parts of the grammar are semi-independent and therefore form a basis for division of labor while enhancing parallel implementation. Similarly, the language may change in various ways but have little overall impact on any implementation already begun.

Conclusion

The implementation of software for GIS is immensely difficult. The formal ideas imbedded in scanner-generators, parser-generators, and language processors organize and make the implementation process much more manageable. Furthermore, the coordination required to be certain that the GIS can indeed meet the needs of a very diverse user group is perhaps even more difficult. The language development process presents a method incorporating language processor concepts to maintain user involvement throughout GIS development. A major advantage is the capability to provide users immediate experience with languages that they have helped to design. Their participation in this manner assures that the GIS will indeed meet their needs.

GIS have the potential to help us plan for and wisely use the resources we have. The massive amounts of data being collected must be made more interrelatable. The systems into which we place this data must make the data, as in Everest, more sharable, available, and lend integrity and must do so while remaining evolvable. We present language processor concepts and the language development process as a means to these ends.

BIBLIOGRAPHY

Aho, A.V. and S.C. Johnson. "LR Parsing." Computing Surveys, Vol. 6 (1974), pp. 99-424.

Aho, A.V. and J.D. Ullman. Principles of Compiler Design, Reading, Mass: Addison-Wesley Pub. Co., 1977.

Backus, J.W. "The Syntax and Semantics of the Proposed International Algebraic Language of the Zurich ACM-GAMM Conference." Proc. International Conference on Information Processing, UNESCO (1959), pp. 125-132.

Butler, K.S., B.H. McCown and W.A. Gates. "Use of a Computer-Based Resource Data System to Evaluate On-Site Waste Disposal Alternatives." Journal of Soil and Water Conservation (in press).

Chomsky, N. "Three Models for the Description of Language." IREE Tranactions Information Theory, Vol. IT2 (1956), pp. 113-124.

Everest, G.C. "The Objective of Database Management." Information Systems COINS IV (J.T. Tou, ed.) New York: Plenum Press, 1974.

Gries, D. Compiler Construction for Digital Computers. New York, New York: John Wiley & Sons, Inc., 1971.

McCown, B.H., K.S. Butler and W.A. Gates. A Resource Data Management System (GRASP). Part 1. Introduction and Application to the Lake Superior Shorelands. Madison, Wis., Institute for Environmental Studies, Report 87, 1977.

Miller, A.H., B.J. Niemann, Jr., D.C. Bunde, T.M. Krauskopf, and B.E. Frazier. Applications of Computer Technology to Highway Location Dynamics: Phase II. Madison, Wisconsin, Department of Landscape Architecture (1972).

Ross, D.T. "Homilies for Humble Standards." Communications ACM, Vol. 19 (November 1976), pp. 595-600.

"Special Issue: Data-base Management Systems." Computing Surveys, Vol. 8 (1976), pp. 1-151.

Sterling, T.D. "Guidelines for Humanizing Computerized Information Systems: A Report from Stanley House." Communications ACM, Vol. 17 (November 1974), pp. 609-613.

Tomlinson, R.F., ed. Geographical Data Handling. Ottawa, Canada: International Geographical Union Commission on Geographical Data Sensing and Processing, 1972.

Wilcott, J.C. and W.A. Gates. "A Review of Existing Geographical Information Systems and Some Recommendations for Future Systems." Symposium on Classification, Inventory, and Analysis of Fish and Wildlife Habitat, January 1977.

Information Education in the Geosciences: Anatomy of a Course

by

Joseph J. Kohut
Science Librarian
Portland State University
Portland, Oregon

OBJECTIVES & SCOPE OF COURSE

Since World War II--especially in the post-Sputnik era--effective scientific and technical communication has been a matter of national concern. The Committee on Scientific and Technical Information was established in 1966 under the aegis of the National Academy of Science and the National Academy of Engineering to investigate the status and future requirements of the scientific and engineering communities with respect to the flow and transfer of information. Fundamental to its philosophy is the proposition that the strength of science and technology depends on personal committment by scientists and engineers to the development and use of their communication systems. Corollary to this effort is the role of education. An oft-cited study is the Weinberg Report which also recognizes that those concerned with research and development must accept responsibility for the transfer of information in the same degree and spirit that they accept responsibility for research and development itself. This is self evident in the face of the fact that the later stages in the information-transfer process, such as retrieval, are strongly affected by the values and behavior of the originators of scientific information.

Exponential growth and increasingly complex information-communication patterns within and among disciplines demand practical solutions. Though such problems are primarily of interest to professional documentalists, the scientific societies have traditionally been sponsors of the primary and secondary literature. It may be inevitable, certainly appropriate, that shortcomings be resolved--if they are resolved at all--by the scientific and technical societies. Given then a sense of responsibility and at least a superficial understanding of information and documentation, members ought to be willing and better able to assume an active role in resolving communication problems in their discipline.

Clearly, information education for scientists and engineers must include more than simple bibliographic skills. For example, the information transfer process must be taught in its entirety--from generation through retrieval. In particular, the student should understand the nature of scientific communication and the central role it plays in research and development and, in fact, will play in his career. I am convinced that traditional courses in scientific bibliography, typically narrow in scope and biased in perspective, fail in these broad objectives.

The conventional course tends to be of the "where-to/how-to-find-out" variety, often with some "guide-to-the-literature" as a textbook. More specifically, conventional courses deal almost solely with retrieval, the final link in the information-transfer chain. Moreover, they stress only that channel for retrieval, the secondary literature, to which the student normally has access in his search of the creative literature. Though such bibliographic skills are, of course, essential, the conventional course tends to be directed more to the student as a student than to the student as a potential scientist. The bulk of his information-related activities as a practicing scientist continues to be left to chance, intuition, or example.

As noted in the Weinberg Report, the attitudes and practices of authors, editors and publishers predetermine the later steps in the information-trasfer process. Yet conventional courses generally ignore such matters as the attitudes and practices of scientists as producers as well as users of information.

The fact is that scientists <u>do not</u> work in isolation. They are members of communities of interacting, that is, communicating, scientists. Thus, personal success and the advancement of science in general depend in part on effective communication within scientific communities. Effective communication depends in part on adherence to certain social norms and position in the scientific network of communication. In other words, social as well as ideational factors are significant in the development of science and scientific careers.

There is another dimension, the historical, which is appropriate in a course in scientific communication. The student's conception of the development of science is biased by textbooks, the primary channel of recorded information to which he is exposed at least on the undergraduate level. The student ordinarily does not understand, and may never <u>fully</u> appreciate even as a scientist, the workings of science between the publication of one textbook and the next, or how the textbooks create a frame of reference that color his interpretation of the literature and, in fact, nature itself. Kuhn, for example, has proposed a highly regarded interpretation of scientific growth, which shows how it is often irrational, illogical and non-cumulative. Rarely do conventional courses teach scientific communication as a process; never, as far as I know, as it is related to the sociology and history of science.

Finally, conventional courses vary widely in emphasis depending on whether they are taught by professional librarians, documentalists, or scientists themselves. To librarians and documentalists information retrieval is largely an end in itself, hence a preoccupation with the mechanics of information and document retrieval systems. To the scientist information retrieval is only a means to an end. Constrasting points of view may affect course content in many, often subtle, ways. For example, librarians tend to value comprehensiveness in information channels; scientists, convenience. And so forth. The challenge in designing a course in information education is to find a best fit solution; that is, one that is multi-faceted but selective and carefully balanced according to these various perspectives.

In short, information education should range well beyond the bounds of the traditional course. A minimally acceptable course should include the notion of science as a human activity, information as a phenomenon central to that activity, and communication as a process dealing with that phenomenon.

GENERAL OUTLINE

Given the broad-ranging objectives I have suggested, organization of such a course is a critical matter. In the first instance, topics range from the abstruse (e.g. nature of scientific knowledge) to the mundane (e.g. how to use a particular bibliographic tool). Moreover, selected elements must be drawn from a diversity of disciplines that impinge on the information-communication problem. All these aspects must be integrated into a meaningful sequence of instruction.

Case histories and examples drawn from geology are used to introduce basic concepts and develop information-retrieval skills. In addition to providing a framework in which to present ideas concerning the growth and development of the geosciences, case histories serve as examples of information-retrieval strategies and of how individual scientists typically acquire and communicate information.

With case histories as a point of departure the following kinds of questions are discussed.

1. How does science really work and how are its workings reflected in the literature?

2. What motivates scientists and what observations can be made from the literature to suggest these motivations?

3. What is the social organization of science and how is it related to scientific communication?

4. What is the scope and nature of the science-information problem?

5. What are the various channels for acquiring information? What are their limitations? How are they perceived by scientists, librarians, information scientists, etc.?

6. What are some of the characteristics of scientific literatures and how do these affect information retrieval?

7. What are some of the problems of organizing scientific knowledge?

8. What operational problems commonly plague information and document delivery systems and how do these affect the user?

Though it is impossible to isolate topics into neat categories, the outline for the course consists of four major parts which are:

1. science and scientific communication as a human activity.

2. principles and channels of information and document retrieval.

3. scientific literatures and basic search strategies.

4. bibliographic sources and services for the earth sciences.

PART 1: SCIENCE AS A HUMAN ACTIVITY

Part one begins with a discussion of the characteristics of the geoscience literature. Size and growth rate are discussed, including the nature of exponential growth, the logistic curve, and the relationship between literature size, growth rate and number of living scientists. Growth patterns of individual subdisciplines are also discussed with emphasis on research fronts and irregular, amoeba-like patterns of growth. The concept of fast (e.g. solid-earth geophysics) and slow (e.g. vertebrate palaeontology) fields of study is introduced and comparisons are made.

This introduction to the primary literature lays the groundwork for a look at those activities that underlie its evolution. Particular topics include the nature of scientific information vs. scientific knowledge, the growth of science, and the nature of scientific revolutions. I borrow from the observations of historians and sociologists of science as well as personal experience, especially familiarity with a mini-revolution, the emergence of multielement-conodont taxonomy. The basic theme, however, is the development of the theory of continental drift, a superb example of Kuhn's theory of scientific growth (i.e. appearance of a paradigm, normal science, anomalies, crises, emergence of a new paradigm). Finally, Crane has shown how Kuhn's stages of development of scientific concepts are related to segments of the logistic curve for scientific literatures.

While summarizing and interpreting the history of continental drift, the impact of sociological factors cannot be ignored. The story is replete with anecdotes that demonstrate the effects of such social phenomena as the reward system in science, competition, social stratification, opinion leaders in science, etc. A particularly useful example is described in an article by Menard on the publication history of the concept of sea-floor spreading. I use this scenario to introduce such ideas as preprints vs. formal publication, invisible colleges, priority in science and near simultaneous discovery. Another example is the article by Vine and Matthews published in Nature in which they linked the notion of sea-floor spreading to geomagnetic reversals and magnetic anomalies parallel to mid-oceanic ridges. The Vine-Matthews hypothesis is similar to that of L.W. Morley whose paper was earlier rejected by the same journal. There is a natural lead-in here to a discussion of peer review of scientific papers and organized scepticism vs. speculation in science.

PART 2: INFORMATION AND DOCUMENT RETRIEVAL

Whereas Part one dwells on the nature, growth and development of the earth-science literature, Part two is primarily concerned with information retrieval from that literature. Whereas the perspective in Part one is largely that of the historian and sociologist of science; that of Part two, the professional documentalist.

Part two is organized according to channels for acquiring information, both formal and informal. Advantages and limitations of each channel are compared from the points of view of the scientist, documentalist and librarian.

For example, scanning the current literature is a popular and effective method by which scientists acquire information. Given the magnitude of the information problem, this form of retrieval seems a hopelessly inefficient means for "keeping up with the literature". However, the structure and usage of the literature, for which the scientist has an intuitive feel, can be studied to show how potentially useful journal articles tend to be concentrated in the literature. Literature scatter, analyzed according to source of publication, leads to a discussion of Bradford-Zipf distributions, literature obsolescence, and the concept of core journals in science. These measures of literature scatter can be related to research fronts, classical and ephemeral literatures, etc. Citation-delay curves--a product of my own research--show, for example, that the post-World War II literature explosion in solid-earth geophysics occurs simultaneously with an implosion according to source of publication.

The secondary literature is, of course, also evaluated as an information channel. A considerable amount of time is spent on the problem of expressing and clustering concepts by index terms, also hierarchical structures of knowledge, depth of indexing, scientific jargon, etc. There is an endless number of examples of geologic terms depicting synonomy, obsolescence, local usage, spelling variations and similarities, multiple meanings, etc. Several other topics are introduced here (*e.g.*, free text search and natural language, controlled and entry vocabularies, etc.).

The matter of subject-term linkages between concepts and documents, and between intermediaries (*i.e.*, indexers) and users is given special emphasis. The importance of building a pre-search file of potentially relevant terms is stressed. I use the concept of radiometric dating of rocks as illustration. As specific bibiographic tools are mentioned throughout the course, I continually return to the radiometric dating search in order to show the various, sometimes unexpected, ways used to express this standard geologic concept.

An information-retrieval technique frequently employed by scientists is to follow up references cited. The citation approach to retrieval is compared with the subject-term approach. A very important distinction is made between concept-concept linkages of documents by citing authors and concept-term clusters of papers by indexers.

Other channels of informal and formal communication are: preprints and reprints, correspondence, textbooks, librarians or information officers, personal indexes or records, internal organizational reports, subject bibliographies, and, of course, scientific meetings.

The library is studied as an example of a document-retrieval system; the card catalog, as a channel for acquiring information. Topics on classification range from the organic nature of knowledge and its implications for classification systems, to the meaning of classification and cutter numbers. Other topics covered are file organization and basic filing rules, browsibility and serendipity as considerations in collection organization, and classification problems relating to multidisciplinary documents and form of publication. Library cards are examined as sources of bibliographic information and as leads, via tracings, to other pertinent subject headings and documents.

The unit on document retrieval concludes with some comments on system effectiveness. Topics briefly covered are user error and frustration, system malfunction, circulation, acquisition and interlibrary loan. A search for monographs on radiometric dating is used to demonstrate many of the points made above.

PART 3: SEARCH STRATEGIES

It has been my experience that a bare handful of bibliographic tools provides access to the bulk of the geoscience literature. Part 3 deals with three of these tools, each quite different from the other. As part of the instruction in the use of these sources, they are evaluated and compared according to the principles presented earlier in the course. This section is divided into three units which are:

Unit 1. Subject-search approach to the geoscience literature using the Bibliography and Index of Geology.

Unit 2. Multidisciplinary citation-search approach to the scientific literature using Science Citation Index.

Unit 3. Manual and on-line searches of the technical report literature using Government Reports Announcements/Index and the NTIS computer readable data base.

The development of the Wilson-Morgan hypothesis, an explanation of the driving mechanism behind plate tectonics, is used as a case history. The immediate objective is threefold: (1) to study the channels for acquiring and communicating information used by two prominent scientists, (2) to look at its implications in terms of the sociology of science, and (3) to gain insight and skills in the use of The Bibliography and Index of Geology.

The communication pattern of the concept of mantle plumes yields several noteworthy observations: the importance of information exchange at scientific meetings, the inordinate delay which is common for monographs with many contributors, the lag time between publication and indexing, and the role of such magazines as Nature and Science. Quick-publication expedites priority; high standards lend prestige to articles; and wide circulation promotes visibility. All three factors are, of course, designed to maximize the potential for reward; that is, peer recognition. And so on.

As the English language index for the geosciences, the Bibliography and Index of Geology merits a close look, not only as an example of subject indexes in general, but for its intrinsic value to geoscientists. In conjunction with this discussion I raise such problems as the organization of geological knowledge, subject vs. geographic vs. process approaches to searching, and the time lag between appearance of new scientific terms and their incorporation into index vocabularies. Mantle plumes and radiometric dating, are used as search topics in the Bibliography and the results evaluated. The objective is to reinforce those points that distinguish the sophisticated from the casual search, and to set reasonable expectations for subject indexes as guides to information.

One of the most powerful, but least understood, bibliographic tools is the Science Citation Index. A reasonable introduction is by way of comparison of subject-search vs. citation-search strategies. Morgan's first three papers on mantle plumes are taken as a point of departure for a search of the Science Citation Index for 1973. The contrast with a search of the Bibliography and Index of Geology is impressive: 13 "hits" for the Bibliography; 85 unique "hits" for Science Citation Index, 9 of which cite all three of Morgan's papers. Among these is Molnar and Atwater's paper in Nature that disputes the fixity of hot spots.

Molnar & Atwater's article, a "typical" paper at the research front of a fast moving subfield, is shown to rely on a relatively small number of core journals published over a relatively brief interval of time. The references cited are matched against the Bibliography up to the time Molnar and Atwater's paper was submitted for publication. The fallout gives a rough idea of the limited degree to which such researchers could rely on standard bibliographic sources while maintaining their current pace of research and publication. The lesson is clear: volatile areas of research tend to outrun the ability of the secondary literature to provide leads to up-to-date information.

Special attention is paid to the "Source Index" as an author bibliography. Scientists tend to publish time and again in the same area of research so that the author-bibliography strategy of information gathering is very often productive. Moreover, there are some interesting relationships between quality and quantity of published research. Lotka's Law, a power-law relationship between number of scientists and productivity, is discussed in order to impress students with the fact that the bulk of the scientific literature is produced by a relatively few scientists (i.e., about 10% of the scientists produce about 50% of all scientific papers). These deliberations on author productivity lead naturally to a discussion of the Ortega hypothesis and the counter-arguments of Cole and Cole.

Although the federal government is the largest publisher in the world, much of its output, particularly the technical report literature, is not fully exploited. Because technical reports are frequently state-of-the-art sources of information on applied geology (e.g. geothermal resources, remote sensing, earthquake prediction) I devote a unit of instruction to the activities of the National Technical Information Service (NTIS). This unit deals with the nature of technical reports, structure of the NTIS bibliographic data base, and manual search of Government Reports Announcements/Index. The technical report literature is compared with the formally published literature in terms of referring, orientation, timeliness, content, etc.

The NTIS data base is an especially good one to demonstrate on-line bibliographic retrieval. Topics fundamental to understanding on-line searching are presented (i.e. Boolean logic and term coordination, postings, uniterm cards, computer file organization, etc.). Students are walked through the design of a search strategy on geothermal resources of the Pacific Northwest. The search includes NOT as well as AND and OR logic. Because the manual and on-line search topics are identical, a direct in-class comparison is made. The advantages of on-line retrieval become obvious; for example, the flexibility of term coordination, computer scan of text. etc. Moreover, students are impressed by the off-line print of complete references and abstracts for 118 hits, most of which are relevant. Incidently, NTIS is one of the least expensive data bases, therefore, the benefit per cost is high.

Another source of government information is, of course, the Government Printing Office. Much material of geologic interest turns up in government documents, sometimes in quite unexpected places (e.g. committee prints, hearings, etc.). The geothermal resources question is carried through a search of the Monthly Catalog of U.S. Government Publications. A major portion of this unit is also reserved for the publications and publication history of the federal and state geological surveys. Part three concludes with a brief survey of the role of the U.S. Geological Survey in the geosciences.

PART 4: EARTH-SCIENCE BIBLIOGRAPHY

The geoscience literature was produced and is maintained at great expense. Much of it is multidisciplinary and international in scope. It is, therefore, unfortunate that so little time--if any--is devoted to training geologists to more effectively search the literature. Part Four of this course is designed to develop some of the basic skills that I think pertain to this real, but widely neglected problem.

In content and organization this part of the course is very much like that of traditional discipline-oriented courses in bibliographic methods. Instruction is organized along two lines: (1) sources of information and search strategies for common types of geologic questions and (2) kinds of publications that are in large part unique to geology. The former includes mainly "how-to-find-out" questions about such things as stratigraphic units (Lexicon of Geologic Names), fossils (Treatise on Invertebrate Paleontology), minerals (Mineralogical Abstracts), etc. Special kinds of publications discussed include field-trip guidebooks and geologic maps.

There is, of course, a wealth of detailed bibliographic information that can be explored. I am, therefore, emphatic about the value of two recent guides to the earth-science literature: Ward and Wheeler's Geologic Reference Sources and Wood's The Use of Earth Sciences Literature. Given these two resources and an understanding of the first three parts of the course I believe the student can find his way reasonably well through the geoscience-information maze.

CONCLUSION

In conclusion, although the impact of the course is difficult to measure, student response is enthusiastic. In particular they are fascinated by the historical and sociological aspects of the course. Nonetheless, they are primarily interested in geology _per se_, not scientific communication as a separate field of study. In view of the other demands of their major their concern is the immediate pay-off of bibliographic skills in other courses in geology and graduate school. It is, therefore, gratifying that on the basis of their course assignment they seem to compare remarkably well with students in graduate schools of library science.

I acknowledge the Center for the Utilization of Scientific Knowledge at the Institute for Social Research, University of Michigan, who with National Science Foundation funds encouraged me to write the report on which this paper is based.

A Survey of the Holdings of a Sample of International Union of Geological Sciences (IUGS) Publications in Selected U.S. Geoscience Libraries

by

Rosalind Walcott
Earth and Space Sciences Library
State University of New York at Stony Brook

Abstract

A survey of holdings of a sample of IUGS publications in selected U.S. geoscience libraries showed that university libraries had the highest ownership rates of these publications, with company, governmental and public libraries having significantly lower ownership rates. Most librarians who responded to the questionnaire indicated that they considered the IUGS publications that they owned to be valuable additions to their collections. The comments made by the librarians pointed to a need for more advertising by IUGS to ensure that librarians are aware of the publications available.

Introduction

This survey was conducted as the second part of a study suggested to the Geoscience Information Society (GIS) by the Committee on International Scientific and Technical Information Programs of the National Academy of Sciences. The first part of the study was an investigation of citation rates of selected IUGS publications. The results of this study have already been reported in the Proceedings of the eleventh meeting of GIS (Bichteler, 1977).

The survey of holdings was conducted to determine distribution and availability of IUGS publications in the United States geoscience community. An attempt to gather information on the usage and value of IUGS publications in United States geoscience libraries was also made.

Methodology

The survey of holdings was conducted using the same list of IUGS publications selected by Bichteler to investigate citation rates. The criteria she used for inclusion of particular publications in the list are discussed in her paper (p. 5). The list of publications was divided into three parts; conferences, symposia and proceedings, reviews, and other publications. This list was circulated, together with a cover letter from GIS (Appendix A), asking librarians to indicate which of these publications they own, and also asking them to comment on the usage and value of these publications in their libraries. This questionnaire, together with a return addressed envelope, was sent to 210 geoscience libraries in the United States. The libraries of all United States universities listed in the 1976-77 Directory of Geoscience Departments for the United States and Canada as having a doctoral program in geoscience received a copy of the questionnaire (111 libraries), and also all company, public and governmental libraries listed in the second edition of Directory of Geoscience Libraries (99 libraries).

Results

Of 210 questionnaires mailed, 3 were not delivered owing to out-of-date addresses, and 106 questionnaires were returned. This represents a return rate of 51% which can be considered to be very high for a mailing such as this. Of the 106 returned questionnaires, 68 were from university libraries, 16 from company libraries, 15 from governmental libraries (USGS, state survey libraries etc.) and 7 from public libraries (museum libraries, public institutions etc.).

The results are presented in Table I. Clearly, university libraries have the highest ownership rate for these selected IUGS publications, with company

TABLE I

COMPILATION OF RESULTS FROM RETURNED QUESTIONNAIRES

Item	% owned by all libraries (106)	% owned by university libraries (68)	% owned by company libraries (16)	% owned by governmental libraries (15)	% owned by public libraries (7)
CONFERENCES, SYMPOSIA & PROCEEDINGS					
Antarctic Geology	59	74	31	33	43
Antarctic Geology and Solid Earth Geophysics	56	68	38	27	43
Deep-seated Foundations of Geological Phenomena	54	65	44	27	29
Genesis of Precambrian Iron and Manganese Deposits	42	59	6	27	0
Geochronology of Precambrian Stratified Rocks	54	65	44	33	43
Geology of Saline Deposits	56	69	38	27	29
Gondwana Stratigraphy	63	74	56	40	29
International Clay Conference, Stockholm	57	71	38	33	14
International Clay Conference, Jerusalem	50	69	19	27	0
IUGS Committee on Mediterranean Neogene Stratigraphy, Proceedings 3rd Session, Berne	39	49	25	20	14
Ores in Sediments	63	82	31	33	14

Item	% owned by all libraries (106)	% owned by university libraries (68)	% owned by company libraries (16)	% owned by governmental libraries (15)	% owned by public libraries (7)
Origin and Distribution of the Elements	59	82	0	40	29
Problems of Hydrothermal Ore Desposition	44	60	13	27	29
Recent Crustal Movements and Associated Seismicity	34	38	19	33	14
Second Gondwana Symposium	28	31	38	20	0
Sexual Dimorphism in Fossil Metazoa and Taxonomic Implications	35	47	6	20	14
Spilites and Spilitic Rocks	52	69	13	33	14
Trace Fossils	67	87	38	27	29
Upper Mantle Symposium	45	56	25	33	14
REVIEWS					
Almeida	13	15	6	13	14
Didier	8	12	0	7	0
Dylik	4	3	0	7	0
Evans	21	29	0	20	0
Maarleveld	5	5	0	13	0

Item	% owned by all libraries (106)	% owned by university libraries (68)	% owned by company libraries (16)	% owned by governmental libraries (15)	% owned by public libraries (7)
OTHER PUBLICATIONS					
Hruska	44	50	25	40	43
Lexique Stratigraphique International	49	54	69	27	14

and governmental libraries having significantly lower rates of ownership. As only 7 public libraries returned questionnaires, it is difficult to draw any valid conclusions from these data.

Higher rates of ownership for all types of libraries were shown for the conferences, symposia and proceedings volumes in comparison to the reviews section. The average rate of ownership for the conferences, symposia and proceedings section for university libraries was 63.9%, for governmental libraries 29.5%, and for company libraries 27.5%. The average rate of ownership for the reviews section for university libraries was 12.8%, for governmental libraries 12.0%, and for company libraries 1.2%.

Higest rates of ownership for university libraries were shown for the titles Trace Fossils (87%), Ores in Sediments (82%), and Origin and Distribution of the Elements (82%). Six university librarians who responded to the questionnaire supplied usage statistics for the titles that they owned. These same three titles were also cited by these six librarians as receiving heaviest use. In addition the two International Clay Conference volumes were mentioned as receiving heavy use. In the comments given by librarians who answered the questionnaires (see Appendix B), the titles Origin and Distribution of the Elements, Ores in Sediments, Trace Fossils, Gondwana Stratigraphy and the two International Clay Conferences were mentioned as receiving heaviest use.

Highest rates of ownership for company libraries were shown for Lexique Stratigraphique International (69%) and Gondwana Stratigraphy (56%). Comments from company librarians showed the Lexique to be a most useful reference for them.

For governmental libraries, highest rates of ownership were shown for Gondwana Stratigraphy, Origin and Distribution of the Elements and Hruska, Computer Based Storage and Retrieval of Geoscience Information (40% for all three).

It is obvious from the comments written on the questionnaires that many librarians find IUGS publications difficult to identify for purchase and often difficult to acquire. Many librarians mentioned how useful an annual bibliography of IUGS publications would be for them, and several were grateful for this list of publications as it pointed out gaps in their collections. It seemed that many librarians who received this list would have purchased more IUGS titles had they been aware of prices and ordering information for these titles. These facts point to a weakness in IUGS publication procedures. More effort should be made by IUGS to ensure that librarians are aware of new IUGS publications and of where they can be ordered. The majority of librarians rated IUGS publications valuable to their libraries. The Lexique Stratigraphique International seems to be a key publication for company libraries, while university librarians rate many of the IUGS publications very valuable to their faculty and graduate students (see comments, Appendix B).

The majority of librarians that owned titles that were available as separates or as journal issues indicated they they owned the title in journal issue form. This supports the IUGS policy of publishing in journals wherever possible, particularly if the journal is a well-known and widely distributed title (e.g. Tectonophysics, Canadian Journal of Earth Sciences).

Conclusions

Librarians who responded to the questionnaires (51% of all librarians polled), indicated in general that IUGS publications were a valuable addition to their collections. Highest ownership rates of these publications were shown by university libraries, with company, governmental and public libraries having significantly lower rates of ownership. Highest ownership rates for all libraries were shown for conferences, symposia and proceedings volumes. Particularly high ownership rate for the Lexique Stratigraphique International in company libraries was indicated, and company librarians stated that they placed great value on this publication. Comments from many librarians pointed to a need for better advertising practices for IUGS publications. Many librarians were simply not aware of certain IUGS publications that they would otherwise have purchased.

References

Directory of Geoscience Departments. United States and Canada 1976-1977. American Geological Institute, Falls Church, Virginia, 1976.

Bichteler, Julie. Publications of the International Union of Geological Sciences: Their Influence on U.S. Geoscientists. Geoscience Information Society, Proceedings, v. 7, 1977. pp. 1-16.

Walker, Richard D. and Diane Parker (comp.) Directory of Geoscience Libraries, U.S. and Canada, Second edition, Geoscience Information Society, Washington, D. C. 1974.

GEOSCIENCE INFORMATION SOCIETY

May 1977

Dear Librarian:

At the suggestion of the National Academy of Sciences the Geoscience Information Society is investigating the impact on U.S. geoscientists of publications of the International Union of Geological Sciences (IUGS). We are conducting a survey of holdings of IUGS publications in U.S. geoscience libraries and we are also interested in any comments that you can make concerning the usage and importance of IUGS publications in your library.

Please help us by taking the time to answer the attached questionnaire and return it in the enclosed addressed envelope. Thank you.

QUESTIONNAIRE

Please check those items owned by your library.

Conferences, Symposia, and Proceedings:

Antarctic Geology, Cape Town, 16-21 September 1963. Raymond J. Adie, ed. Amsterdam: North-Holland, 1964.

Antarctic Geology and Solid Earth Geophysics, Oslo, 6-15 August 1970. Raymond J. Adie, ed. (IUGS Series B, Number 1). Oslo: Universitetsforlaget, 1972.

Deep-Seated Foundations of Geological Phenomena, Prague, 1969. F. Delany and C. H. Smith, eds. Amsterdam: Elsevier, 1969. (also published in *Tectonophysics,* v. 7, no. 5-6, June 1969).

Genesis of Precambrian Iron and Manganese Deposits, Kiev, 20-25 August 1970. Paris: UNESCO, 1973.

Geochronology of Precambrian Stratified Rocks, Edmonton, 1967. Ronald A. Burwash and R. D. Morton, eds. *Canadian Journal of Earth Sciences,* v, 5, no. 3, June 1968.

Geology of Saline Deposits, Hanover, 15-21 May 1968. G. Richter-Bernburg, ed. Paris: UNESCO, 1972.

Gondwana Stratigraphy, Buenos Aires, 1-15 October 1967. Paris: UNESCO, 1969.

International Clay Conference, Stockholm, 12-16 August 1963. I. Th. Rosenqvist and P. Graff-Petersen, eds. New York: Pergamon, 1963.

International Clay Conference, Jerusalem, 20-24 June 1966. L. Heller and A. Weiss, eds. Jerusalem: Israel Program for Scientific Translations, 1966.

International Union of Geological Sciences. Commission on Stratigraphy. Committee on Mediterranean Neogene Stratigraphy. *Proceedings of the Third Session in Berne,* 8-13 June 1964. C. W. Drooger, Z. Reiss, R. F. Rutsch, and P. Marks, eds. Leiden: E. J. Brill, 1966.

Ores in Sediments, Heidelberg, 31 August-3 September 1971. G. C. Amstutz and A. J. Bernard, eds. (IUGS Series A, Number 3). Heidelberg: Springer-Verlag, 1973.

Origin and Distribution of the Elements, Paris, 1967. L. H. Ahrens, ed. Oxford: Pergamon, 1968.

Problems of Hydrothermal Ore Deposition, St. Andrews, Scotland, 1967. Zdenek Pouba and Miroslav Stemprok, eds. (IUGS Series A, Number 2). Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung, 1970.

Recent Crustal Movements and Associated Seismicity, Wellington, 10-18 February, 1970. (IUGS Series B, Number 2). *Royal Society of New Zealand Bulletin,* 9, 1971.

Second Gondwana Symposium, South Africa, July to August 1970. Pretoria: Council for Scientific and Industrial Research, 1970.

Sexual Dimorphism in Fossil Metazoa and Taxonomic Implications, Prague 1968. G. E. G. Westermann, ed. (IUGS Series A, Number 1). Stuttgart: E. Schweizerbart'sche Verlagsbuchhandlung, 1969.

Spilites and Spilitic Rocks, Prague, 1968. G. C. Amstutz, ed. (IUGS Series A, Number 4). Heidelberg: Springer-Verlag, 1974.

Trace Fossils, Liverpool, 6-8 January 1970. T. P. Crimes and J. C. Harper, eds. Liverpool: Seel House, 1970. (also published in *Geological Journal Special Issue*, no. 3, 1970).

The Upper Mantle Symposium, New Delhi, 1964. Charles H. Smith and Theodor Sorgenfrei, eds. Copenhagen: Berlingske Bogtrykkeri, 1965.

Reviews:

Almeida, F. F. M. de, O. H. Leonardos, Jr., and J. Valenca. *Granitic Rocks of Northeast South America.* Prepared for IUGS/UNESCO Symposium in Recife, Brazil, October 1967. 41p.

Didier, J. *Les Enclaves des Granites dans la Litterature Geologique.* Paris: B.R.G.M., 1964. 18p.

Dylik, J., and G.C. Maarleveld, *Frost Fissures, Frost Cracks and Related Polygones; A Summary of the Literature of the Past Decade.* Maastricht: E. van Aelst, 1967. 15p.

Evans, I. S. *Salt Crystallization and Rock Weathering: A Review. Revue de Geomorphologie Dynamique*, no. 4, 1970.

Maarleveld, G. C. *Frost Mounds: A Summary of the Literature of the Past Decade.* Maastricht: E. van Aelst, 1965. 16p.

Other Publications:

Hruska, J. and C. F. Burk, Jr. *Computer-Based Storage and Retrieval of Geoscience Information: Bibliography 1946-69.* (Geological Survey of Canada Paper 71-40). Ottawa: GSC, 1971.

International Union of Geological Sciences. Commission on Stratigraphy. Subcommission for the Stratigraphical Lexicon. *Lexique Stratigraphique International.* Paris: CNRS, 1956- .

Please comment on the usage and value of these IUGS publications in your library.

APPENDIX B

COMMENTS FROM RETURNED QUESTIONNAIRES

Governmental Libraries

"usage is marginal"

"very useful and frequently cited - hence they are essential in any geoscience library. The Lexique is a very useful reference."

"have been used occasionally. They have provided valuable information when used. We have also borrowed these publications marked on interlibrary loan."

"the publications are not heavily used, however they are important additions to the collection insofar as the few people who utilize the publications are concerned."

"We do not have any of the above publications in the library at the Ohio Geological Survey. Our library is mostly publications about Ohio. Our geologists use the Dept. of Geology at Ohio State University, where all our exchange publications are on deposit."

"none of these is among our holdings, which are primarily USBM, USGS and other publications specifically about Oregon and the Northwest"

"The USGS Library in Reston, Va. has all the publications noted on your list, with one exception. We could not locate a listing for Didier, J., Enclaves des granites dans la litterature geologique. We have requested it on exchange. IUGS publications are considered important primary publications, and essential to any wellrounded earth science library collection. The Survey Library attempts to maintain a comprehensive, indepth file of such literature."

Company Libraries

"not too important - general circulation pattern seems to be that they circulate a bit when we announce them on our accessions list, then they just sit on the shelf"

"as indicated we have rather few IUGS publications, mainly because they are academic in nature, and this company tends to do little in the way of academic study. However, when a special project comes up we would very possibly need some of these publications, so please notify us of new ones."

"checked references were ordered for the most part due to requests from research personnel for use in their work"

"We have found them to be well-prepared presentations. They continue to be used on occasions where they are useful reference tools."

"light use"

"used for petroleum exploration - very useful when needed"

"the publications which we have are very useful - especially the Lexique"

"in most cases we have more than one copy of the references checked. Some on the list we did not know about and will probably try to obtain. With IUGS publications being published by such a variety of publishers, it's not always easy to keep up with IUGS programs. Those in our areas of interest have been quite useful."

Public Libraries, Museum Libraries etc.

"these publications of the IUGS are highly relevant to the research interests of at least three members of our Geology Dept., and have been requested by others on the scientific staff as well."

"not much demand"

"as the primary thrust of the museum's Geology Dept. is in paleontology, most of the materials are outside our interests. I did see several items listed which may be of interest to staff members, and I would find it useful to receive a publications list with prices and order information."

University Libraries

"Geosciences are among the most prestigious and important areas on this campus - hence IUGS publications have been an important part of the library's buying program for years. Usage and overusage are a problem due to high rip-off rate. The fact that all IUGS publications are not available from one office can make buying a problem."

"these publications are used extensively and are quite valuable to our library"

"IUGS publications are used heavily by faculty, staff and students as well as by off-campus patrons."

"We have a graduate program in geology and try to support research in the field."

"some of these publications have been more popular with us than others. With limited funds we will be buying fewer books in the future. Ease of identifying a title for purchase will be a factor, also, some indication that the book will have solid interest for our people. I think that the usage given pretty well reflects the current interests of the Dept. IUGS publications seem to span the spectrum of geology and the ones we have are very important to us. In the future, however, we cannot afford a high cost for a title we think is too specialized and will not at least get moderate use from graduates and faculty."

"all of the IUGS publications checked are heavily used, with the exception of the International Clay Conference, Stockholm. The most used volume of all is Trace Fossils. Of course not all IUGS publications are relevant to the interests of the earth scientists in this institution, but I can see other titles of possible interest and it would be helpful to receive the same list with prices and vendor addresses for those titles still available."

"our holdings are very selective, but those which we do have are used quite heavily."

"not used much, but we are glad to have them"

"the use of these titles varies with the changes in faculty and student interests, although the two Gondwana symposia received a great deal of attention. Most of these publications remain useful, that is they continue to be consulted. In one case, Ahrens, Origin and Distribution of the Elements, the book has recently been replaced. Value seems to me to be a complement of use. In a somewhat different sense, it is important for a library like ours to provide access to these volumes regardless almost of the value placed on them by the scientific community."

"We do not have all the publications that you listed, and we also have other IUGS publications such as Geological Newsletter, maps etc. We also have other issues in several of the series indicated. Because of the nature of our collection we would never feel that we could do without the publications of such a key organization as IUGS. We feel that our collection is not only to fulfil the needs of staff and students of the university, but that it is also a major national resource. I do not keep track of the frequency of use of every item in our collection because it would be an insurmountable task. All of the items you cite are used with varying frequency depending on the research of staff members and graduate students."

"most of the publications were ordered for the library from specific requests. The majority are used frequently."

"heavily used - very valuable to the library"

"all of these are valuable publications and a complete set should be in our library. Many of the items should receive major use for many years to come."

"IUGS publications are difficult to acquire; they are placed in numbered series so often, and are published by so many varied vendors. Some publications in the University of Kentucky Library are used quite often i.e. Gondwana Stratigraphy, International Clay Conference and Ores in Sediments, others are not used quite as often."

"Your publications receive high usage in our library and contribute greatly to the need for geologic information to be made available to the public."

"As you can see we own the publications appearing in scholarly journals and some books published by commercial publishers. It would help the librarian if an annual bibliography was published. Our marine geologists expect to have most of these publications available in the library."

"These books are indeed used by our patrons. Note, however, that we have not ordered those publications not issued by a publisher that regularly sends us blurbs - the only way we keep up with what is available. Now that I know of them I may order. Do you have a mailing list?"

"I'm sure that someone has used some of these publications but I've never noticed it and no enquiries have been made to me."

"The Geology Dept. of this university feels that these publications are an integral part of our library."

"Our library has only those publications indicated. I feel that the others are worthwhile and that our library ought to carry these other publications and that we do not only because we have been negligent in ordering them. We would appreciate receiving these."

"Although this labortory's interest in geology has increased in the past few years, our collection concentrates on a few specific topics. It is, therefore, rather limited in scope. We do borrow many publications that are no longer available for purchase, such as volumes of previous International Geological Congresses and other conference proceedings in geology."

"Excellent sources of information - used heavily."