

ASIDIC newsletter

No. 34, SUMMER 1979

ASSOCIATION OF INFORMATION AND DISSEMINATION CENTERS

BOSTON MEETING SET FOR SEPTEMBER 16-18 AT PARKER HOUSE

Several significant changes have been made to the program for the upcoming Boston meeting. The final program is set now, and will take place according to the following schedule:

Monday, September 17:

Overview of Nonbibliographic Databases

- Comparison of Bibliographic and Nonbibliographic Databases -
Judith Wanger, Cuadra Associates, Inc.
- Domestic Information Displayed for Decision Making -
John J. Quann, Goddard Space Flight Center
- Using Databases in an Information Center -
Donald Hawkins, Bell Laboratories
- Studying Societies: The Roper Center and Secondary Analysis of
Social Data - Everett C. Ladd, Jr., Roper Center and Social Science
Data Center, University of Connecticut
- Online with GE -
Carol Herrick, General Electric Information Services Co.
- The Government-Industry Exchange Program (GIDEP) -
Edwin Richards, Naval Fleet Analysis Center

Luncheon

Social Science Databases

Half-day session chaired and organized by John C. Beresford and
Rebecca E. Hutchins, Dualabs

Lobster Bake - New England Sea Aquarium, Central Wharf, Boston Harbor

Tuesday, September 18:

Scientific and Technical Databases

- Problems in Data Retrieval -
Gesina C. Carter, National Academy of Sciences

(continued)

(continued from page 1)

Physical and Chemical Data - David R. Lide Jr.
National Bureau of Standards
Copyrighting Numerical Data - Viktor Hampel,
Lawrence Livermore Laboratory
Chemical Registry and Structure Files - David
W. Weisgerber, Chemical Abstracts Service
Cartographic Information Systems - Gary W.
North, US Geological Survey

Late registrations for the Boston meeting are still very welcome. Anyone interested should call Jeanette Webb at the ASIDIC Secretariat (404) 542-3106 or Jean Carter at NERAC, the host organization (203) 486-4533. Hotel reservations should be made directly with the Parker House (617) 227-8600.

ASIDIC FINANCIAL CONDITION SOUND; BUDGET PRACTICES, POSITIVE CONTROLS PLAY MAJOR ROLE IN GOOD CONDITION

ASIDIC Secretary/Treasurer Daniel U. Wilde gave a very bright financial report to the ASIDIC business meeting during the recent meeting in Ottawa. The association currently has about \$16,000 in liquid assets, including both checking and interest bearing savings accounts.

ASIDIC operates on a calendar year financial budget. Wilde reported that expenses

are in line with the budget projections for 1980. Income is slightly ahead of projection at this point. This means that ASIDIC should enjoy another positive year under the direction of President J. Ron Smith, Wilde and the Executive Committee.

SPRING '80 MEETING SET FOR WASHINGTON; ATLANTA, NEW ORLEANS SITES TO FOLLOW

As reported in the last issue of the ASIDIC Newsletter, the Spring 1980 meeting of ASIDIC will take place at the Sheraton National Hotel in Arlington, Virginia March 4-6. The March 5 meeting sessions will overlap with the NFAIS annual meeting being held at the same location. The theme of both meetings will be "Transition to the Use of New Technology."

September 21-23, 1980 ASIDIC will meet in Atlanta, Georgia at the Marriott Hotel. The spring, 1981 meeting will be held March 29-31 at the Monteleone Hotel in New Orleans.

ASIDIC members and other interested persons with ideas for topics and speakers for any future meeting should contact Rita Lerner, ASIDIC Program Committee head. Rita can be reached at the American Institute of Physics in New York (212) 661-9404 ext. 603.

ASIDIC NEWSLETTER CORRECTIONS

The following corrections should be noted for the previous ASIDIC Newsletter (No. 33) published last month:

- p.8 The headline announcing ASIDIC/NFAIS joint meeting sessions should not have said "First Ever ASIDIC/NFAIS Joint Sessions . . ." Other joint ASIDIC/NFAIS meetings have been held, the last in 1973.
- p.9 Misspellings of two names of speakers for the Boston meeting slipped by the editor. Rich DeCicco was presented as Rich DeCico and David Weisgerber was listed as David Weiserberger. The editor regrets the errors and apologizes for any inconvenience caused by this commission.

(continued)

CONTENTS Issue No. 34

	page(s)
Boston Meeting Sept. 16-18 Schedule	1-2
ASIDIC Financial Report	2
Future Meeting Sites Chosen	2
Call For Papers, ARIST Vol. 15	3
NSF Grant Given <u>Chemical Abstracts</u>	4
Spring Meeting Speakers	4-8

(continued from page 2)

p.9 No valid, recognized English dictionary uses the editors spelling of "acquarium." The Monday evening lobster bake at the Boston meeting will take place at the New England Sea Aquarium.

ARIST CHAPTER AUTHORS CALL FOR CANDIDATE PAPERS; ELEVEN AREAS SET FOR VOLUME FIFTEEN OF WORK

The individual chapter authors of volume 15 of the Annual Review of Information Science and Technology (ARIST) have issued a call for candidate papers through Martha Williams, ARIST Editor.

Those interested should contact authors of the chapters that might be applicable to their published works. Papers can be sent directly to any author.

The chapter titles and the respective authors include:

Information Communication Technology -
CAWKELL, Anthony E.
17 The Drive
Ickenham, Oxbridge
Middlesex, England

Techniques and Measures for Information Storage and Retrieval -
GRIFFITH, Belver C.
Drexel University
Graduate School of Library Science
Philadelphia, PA 19104

Numeric Databases -
HAMPEL, Viktor
Mailstop 316
University of California
Lawrence Livermore Lab
Livermore, CA 94550

Copyright -
KEPLINGER, Michael
Special Legal Assistant to the Register
Library of Congress
Copyright Office
Washington, DC 20559

Information Systems in Developing Countries -
KEREN, Carl
National Center of Scientific and Technological Information
National Council for Research and Development
Ministry of Energy and Infrastructure
84 Hachashmonaim Street
Tel Aviv 61 200
Israel

Legal Information Systems and Services -
LASS, Jack
National Criminal Justice Reference Service
1015 20th Street, NW
Suite 400
Washington, DC 20036

Artificial Intelligence Applications for Information Retrieval -
SMITH, Linda
Graduate School of Library Science
University of Illinois
Champaign, IL 61820

Information Analysis Centers -
TALMI, Bonnie
P.O. Box 141
Oak Ridge, TN 37830

Computers in Publishing -
TERRANT, Seldon
American Chemical Society
1155 16th Street, NW
Washington, DC 20036

Library Automation -
VENEZIANO, Velma
Northwestern University Library
Northwestern University
Evanston, IL 60201

Emerging Careers in Information Science and Technology -
WELT, Isaac
Graduate School of Library Science
American University
Washington, DC 20016

The views, reports, and opinions published in this or in any issue of the ASIDIC Newsletter do not necessarily reflect the official views or policies of the ASIDIC Executive Committee or the full membership of ASIDIC.

NSF GRANT GIVEN TO CHEM ABS FOR SPELLING ERROR DETECTION, CORRECTION STUDY; PROJECT TO LAST 30 MONTHS

The American Chemical Society's Chemical Abstracts Service has received a grant of \$153,160 from the National Science Foundation for a 30-month study of automatic spelling error detection and correction in large scientific databases. Dr. Joseph J. Pollock, a senior associate information scientist at CAS, is principal investigator for the study.

Pollock and his coworkers will evaluate the effectiveness of various algorithmic approaches to detecting and correcting misspellings, including procedures based on human error patterns, word classification, letter co-occurrence patterns and computer stored dictionaries that automatically recognize inflectional variation. They also will analyze the relative frequencies, types and causes of spelling errors in large samples of text recorded in machine-readable form for publication in Chemical Abstracts, Information Science Abstracts, several American Chemical Society journals, and a number of other scientific and technical publications. The study will draw upon computer procedures developed by CAS over the past several years for alerting its editors to possible misspellings in the computer database from which Chemical Abstracts and other CAS publications and services are produced.

While spelling error detection and correction by computer have been studied for more than two decades, previous investigations have dealt with relatively small bodies of text or limited or controlled vocabularies, and the resulting techniques have not proved effective for large and heterogeneous bodies of text such as Chemical Abstracts, which contains more than a million words each week. The results of the CAS study are expected to be widely applicable because of the large volume and diversity of the text involved.

The immediate objective of the work is to explore means for eliminating as much as possible of the human proofreading and correction required in processing scientific

text by computer and reducing the number of misspellings that escape detection. The study also is expected to provide some new insights into the nature and structure of scientific text that may contribute to the design of more effective information storage and retrieval systems.

MINISIS IS MINI-BASED SYSTEM FOR CREATING, MANAGING INFORMATION FILES; USE GROWING IN CANADA, WORLD-WIDE

In 1973 the International Development Research Centre (IDRC) began using the Integrated Set of Information Systems (ISIS) computer system developed at the International Labour Office in Switzerland. The system was put up on an IBM-based service bureau computer in Ottawa, Canada, the headquarter site of IDRC.

IDRC was established to support development research in Third World countries in the fields of: Agriculture; Food and Nutrition; Health Sciences; Social Sciences; and Information Science. During the time that ISIS was used by IDRC, data entry and retrieval were conducted online. The production of indexes, lists, purchase orders - any printed output except search results printed on a hard copy terminal - and the updating and management of files were batch operations carried out at the service bureau from job control cards submitted by the IDRC staff.

Anne Edwards and Kate Wild of the IDRC described the history of IDRC, the operation of ISIS, and the development and application of a new mini-based MINISIS by IDRC at the recent ASIDIC meeting in Ottawa, Canada.

IDRC was searching existing databases as well as creating databases for searching. "During the first three years (1973-1975) that we had ISIS in operation, we created our own database of some 15,000 records and provided search services on this as well as on the databases of ILO, FAO, and UNIDO. We also built a small database relating to health care delivery systems in developing

(continued)

(continued from page 4)

countries," related Edwards.

"ISIS was costing between \$120,000 and \$140,000 a year to operate through the service bureau. To avoid the great costs of a large computer and the constant inconvenience of working with a service bureau, the decision was made to investigate the possibility of acquiring a minicomputer to run the library's operations to maintain the specialized databases we were already making available, and to develop a relatively inexpensive and reliable package of both hardware and software that could be transferred to developing country institutions," said Edwards.

After an evaluation of mini manufacturers and institutions where mini-based systems were being developed and used, the Hewlett-Packard 3000 Series II was selected. IDRC then turned to the several alternatives for developing their system. Recoding ISIS for the HP-3000 system and adapting the HP database management package, IMAGE, for the new system were considered and discarded as the best path for IDRC.

Edwards said that, "We decided to design a new system, but one that would perform all the functions of ISIS and that would enable us to continue to exchange information with ISIS users." The system was designed to be general purpose so that it could process more than straight bibliographic information. The system would be modular and applications programs would be independent from the operating system making multi-user environments more manageable.

For the users, the system was designed to handle data in a number of physical forms. Simplicity of operation, allowing the user maximum control of the functioning of the system, and a wide variety of outputs were designed into the new system.

The management considerations of cost effectiveness, systems compatibility and compatibility with international standards for information transfer were important elements of the new mini-based system.

System development started in June of 1976.

About the same time a decision was made to put all databases into accordance with the UNISIST/ICSU-AB tape exchange format.

MINISIS performs the same functions as ISIS, and more. It handles all the same library and information management duties plus it allows users to handle data from the system in particular ways. Acquisitions and cataloguers work with much of the same data but have different fields and parts of records to affect directly. Each gets what is needed from the system without having to manage the full portion of each record.

The database building, information retrieval aspects of MINISIS were described in detail by Edwards and Wild.

MINISIS is compatible with the newer HP-3000 Series III and the Series 33. Different configurations for the system make giving a flat cost difficult. A typical Series 33 configuration would run about \$135,000. A Series III would cost about \$180,000.

MINISIS is being applied at the ILO Office in Geneva, at the Agricultural University at Wageningen in the Netherlands, and at sites in Tunisia, the USSR, and in Canada.

In conclusion, Edwards said, "... the main reason we embarked on the MINISIS program was to encourage the recording and sharing of valuable unpublished literature. Our decision appears to have been justified by the results."

MINI-BASED SABINE SYSTEM SHOWS BIBLIO S & R ONLINE PRACTICAL, FAST

"A minicomputer can be the base of a very large, economic and powerful bibliographic data retrieval system as long as some simple operating options are respected," said Gratien Rousseau, director of information systems for Informatech of Montreal, as he spoke to the ASIDIC meeting attendees in June in Ottawa.

Rousseau described in detail the design and functions of SABINE, the French acronym for "Automated System for Non-numerical Databases." Built upon a PDP 11/40 computer,

(continued)

(continued from page 5)

the system is written in MACRO-11, the powerful macro-assembler language from Digital Equipment (DEC).

SABINE is used to create, update and interrogate interactively bibliographic databases. SABINE is written in reentrant and "run anywhere" code. Memory is dynamically allocated between system functions, support of online users, and memory needed for processing, in particular the processing of boolean operations. A system pool of memory exists where every program can ask for memory and return to the system when finished with it, where it is immediately available.

SABINE allocates about 7K of active memory for each online user using the system. Rousseau described the optimization of CPU work loads in SABINE, which allows the smaller mini to perform search functions for the users in much less capacity than traditional large main frame computers.

The average record handled by SABINE is 750 bytes long, with about 40% of each record taken up in the abstract. An average of 30% of each record is inverted. For both the text and the inverted file, the system requires an average of 910 bytes of disk space, some 1.2 times the original text length.

The disk storage and access speed is the critical resource of SABINE, according to Rousseau. The system now uses eight 40 Megabyte RPO3 disk drives. With the 720,000 references currently in the system, four accesses are needed to locate proper keywords and a fifth access to store the answer. With this level of performance, up to 250 simultaneous users can be served with a response time of less than one second.

SABINE is available to anyone at a price of \$40,000. The hardware for the system would cost about \$230,000. The storage costs per reference for the 320,000 online PASCAL type references would be about \$0.84. Taken over a five year period, the cost comes down to \$0.17 per record.

SABINE is reliable, with less than 2% downtime in the first full year of operation and with the addition of more memory the downtime

should nearly disappear. A DATAPAC connection for SABINE is expected in the very near future, greatly broadening the access to the system.

EXXON COMBINES MINI NETWORK WITH MAINFRAMES IN LARGE-SCALE OPERATION

Janet Conner, of the Exxon headquarters business and economics information center in New York City, gave the attendees at the recent ASIDIC meeting in Ottawa an insight into how minicomputers serve the information and library needs of one of the largest companies in the US.

Conner emphasized that the traditional definition of minicomputers based on lower costs and sizes no longer held true. The network Exxon is developing emphasizes the joint use of computers and telecommunications. Distributed processing is really beginning to work with the development of microprocessors which allow more of the processing to be done locally and then move data onto the minicomputer for key processing functions. New storage devices allow increased flexibility and depth of data files for the network.

The impact of all of these advances is "that computing power has come out of the data processing center and into the office or library," said Conner. The minis in the Exxon network are located at the sites where data are generated and used. Modular growth of computing power is made possible with the additional minis as they are needed.

To achieve a true network of minis, Exxon is now implementing and installing Datapoint Attached Resource Computing system. This will be a new concept in mini systems; functionally dispersed, mutually coordinated set of processors, interconnected for universal access to common databases by high speed interprocessor bus. The system control is totally decentralized and does not rely on a host or "key" computer to govern or manage the network.

The new ARC network relies upon software-controlled specialization of processors by

(continued)

(continued from page 6)

function into execution of applications and management of data and hardware/software bus to handle high speed interchange of data.

Several advantages to this approach are seen by Exxon. The system architecture allows for almost unlimited growth, with more processors and disks being easily added to the system. The configuration is flexible and is user-defined. Common databases can be created and any portion of a database can be restricted. Failure of one processor will not cause the entire system to come down. All of these beneficial features allow the system to function in an economical fashion according to Conner.

The future configuration of the Exxon system, with hardware and software features were described in some detail by Conner. An IBM 370 will be used for batch processing jobs to serve the network.

Planning the functions best set up for the minicomputers and those that will stay on the large mainframe is important. The training and staffing of the mini network takes place at the sites participating in the net. The management responsibility is increased at these sites within the network. Control and access are also increased.

Conner said "a minicomputer network system can be a most powerful resource if it is used efficiently and realizes the economic benefits available."

PACKET BROADCASTING USES MICROS; SEEN AS MASS DATALINK FOR FUTURE

Robert T. Rouleau, of Gestion Cadence, Inc., of Montreal, gave a fast moving address to the recent ASIDIC meeting in Ottawa on the development, application and future prospects for packet broadcasting. The importance of this technology on database communications was emphasized.

According to Rouleau, packet broadcasting is nothing more than the application of time-share techniques to a fixed resource, in this

case the radio spectrum. It is the Time Division Multiplexing of a radio channel.

The features (and benefits) of packet broadcasting were revealed in a hypothetical case study that progressively applied buffered line transmission, address initialization for sending and receiving points of data communication, automatic message acknowledgement and microprocessor repetition of unconfirmed message transmission to solve data communication problems.

After packet broadcasting is applied, "... it is discovered that where only two stations were using it (a broadcast channel) many stations could share it without even knowing that it was being shared," said Rouleau.

He described the development of ALOHA NET carried out in the early 1970's at the University of Hawaii. ALOHA NET proved the viability of the technique and recent advances in microprocessor technology have made it very cost effective, said Rouleau.

He said "Packet broadcasting is taken from the fact that data is carried in a package much like a parcel at the post office. It has an address, return address, the contents, and the error detection mechanism."

Satellite technology matured at just the right time as packet techniques make for very efficient use of satellite links. The error detection systems and self corrective nature of the system allows for signal to noise ratios which would be intolerable for other modes.

A small experimental packet radio system is being implemented in Montreal which is much more efficient than the ALOHA system. The development station costs about \$4,000. The micro controlled interface to the system costs about \$500. The terminal can consist of an encoded keyboard attached to a home television set in a similar fashion as a TV game. Mass produced, the costs per unit would be very reasonable. The radio portion could be a simple frequency modulated transceiver similar to one used on a small pleasure craft

(continued)

(continued from page 7)

(a CB radio). Using off the shelf parts, the entire system can be built today for about \$1000.

Some 150,000 users averaging one 1,000 bit packet sent per minute could use one 6 MHZ UHF TV channel. Since there are something like 69 discrete UHF channels, "it seems obvious that we could give up one or two per city without undue hardship," says Rouleau.

Packet radio systems portend a great development in electronic mail. Since all terminals talk to other terminals, no complex switching gear is needed. According to Rouleau, "it won't be long before greater demands are made on databases. Inexpensive terminals and a simple foolproof system will bring small business and even individuals into the database market. Our purchasing habits will be altered, and a mail order business using electronic mail resembles a database system more than just casually."

Public service broadcasting is experimenting with packet broadcasting and growth of the technique for emergency police, fire, and health purposes is certain to gain popularity rapidly.

Security looms as a problem that must be dealt with. The use of encryption will become commonplace, and database managers will have more concern for their security as well.

The demand for better communications by all sectors of the population will push for packet broadcasting. Of no small consideration are the price/cost breakthroughs in small computers and their peripherals. The growth of bi-directional cable systems favors packet techniques. The surplus capacity on the UHF channels favors wide applications of the techniques without causing disruptions to existing users or a race for most favored channels. The great compatibility of packet radio with satellites offers the prospect of a totally connective global network while simultaneously conserving vital broadcast spectrum.

Perhaps the greatest immediate benefit and appeal of packet broadcasting is that is is "cheap", according to Rouleau.

CANADIAN CAL SYSTEM TRULY MULTI-MEDIA; MINIS, MICROS COMBINE IN INTERACTIVE PROCESSES WITH STUDENTS, LESSONS, DATA

Computer Assisted Learning, delivering educational programs with the greatest interactivity of man/machine possible, has been developed by the Information Science Section of the National Research Council of Canada. The CAL system development and applications were explained and demonstrated by NRC's William C. Brown at the recent ASIDIC meeting in Ottawa.

The DAL system, now licensed through private companies as the TELEDON system, is built upon a large DEC PDP-10 system with distributed mini and micro processors and terminals at many sites. The system has been functioning and adding more features and services for the past ten years.

Brown explained that the components of the system include specially adapted CRT terminals housing micros, computer controlled audio and video disks and slide projectors, and audio response units. The CRT and the self contained slide projector have special touch sensitive overlays that use x,y coordinate tactal response sensitivity to interact with the system. These can be for selecting or responding to questions from the system.

A thorough demonstration of the system was presented to the ASIDIC participants from slides and a video tape program that showed all of the system components and how they are actually utilized in a variety of applications. Many universities are developing and placing courses on the system.

The software of the CAL system is called NATAL 74. Course authoring language, and the support software are executed by NATAL. Part of the system uses automatic indexing, classification and retrieval.