

ASIDIC newsletter

No. 48, Spring 1984

ASSOCIATION OF INFORMATION AND DISSEMINATION CENTERS.

Spring Meeting Held in New Orleans Fall Meeting on "Planning for the Future" to be in Washington, DC ASIDIC Members to Consider Revisions to By-Laws

The Spring, 1984 meeting of ASIDIC was held in New Orleans, LA on March 19-20 at the Monteleone Hotel. Its topic was the economics of database production. The meeting, chaired by Edith Crockett of BIOSIS, featured seven talks on the subject, followed by discussion group sessions. A summary of the presentations appears in this Newsletter. About 45 persons attended. Local arrangements were by Marjorie Hlava of Access Innovations, Albuquerque, NM.

The Monday-evening event was a reception with a band; following it, attendees dispersed to sample the excellent cuisine of several noted New Orleans restaurants.

Future Meetings

The Fall meeting will be held in Washington, DC, at the Marbury House in Georgetown, on September 17-18, 1984. It promises to be a most interesting meeting, with the topic "Planning for the 21st Century: 1985 and Beyond." All members and interested persons are urged to attend. The Spring, 1985 meeting will be in Tampa, FL, and the meeting for Fall, 1985 will return to Boston (and another clambake hosted by Dan Wilde!). Consideration is being given to the southwestern part of the country for the Spring, 1986 meeting.

Review of ASIDIC Constitution and By-Laws

Some changes to the Constitution and By-Laws of ASIDIC are being proposed by the Executive Committee. They will be sent to the membership 90 days in advance of a vote, as required. Some other changes were discussed and referred to the Planning Committee.

Committee Reports

Betty Unruh of Dialog Information Systems and Linda Legarde of NTIS, have joined the Planning Committee. Cathy Ferrere announced that five new members, listed below, have joined ASIDIC since the last meeting. David Grooms, President, announced that the Internal Revenue Service has audited ASIDIC and has not changed its non-profit status.

New Publications, Etc.

- Gale Research Co. has announced the first edition of their *Abstracting and Indexing Services Directory*. It lists more than 2,000 English-language abstracting journals, etc. Price is \$140 from Gale Research Co., Book Tower, Detroit, MI 48226.
- John Wiley's catalog will be searchable on the Dialog* system beginning April 1, 1984. It will be the first major book and journal catalog to be available online, according to Wiley.
- The newsletter *Software Protection* will begin monthly publication in June 1984 and will expand its coverage to include international developments and software licensing law. Further information is available from the publisher, Law & Technology Press, 1112 Ocean Drive Suite 201, Manhattan Beach, CA 90266, phone (213)-372,1678.
- Three major database producers, Predicasts, Disclosure, and Data Courier, have joined forces to offer a single training seminar on business database searching. The first offering of this seminar was March 29 in Chicago; if successful, it will be offered in other cities around the country. Representatives from each database producer spoke as well as someone from Dialog. Attendance is free of charge, but registration is required. Contact Teresa Gamble at (800)-626-2823.
- The Special Libraries Association will hold its 75th anniversary meeting in New York on June 9-14. The theme of the meeting will be "Information in the Electronic Revolution."

New Members

ASIDIC welcomes the following new members:

H.W. Wilson Co.
950 University Ave.
Bronx, NY 10452
Bill Bartenbach

Elsevier Science Publishing Co.
52 Vanderbilt Ave.
New York, NY 10017
Dr. George C. Krajcsik

* Dialog is a trademark of the Lockheed Corporation.

Pergamon International Information Corp.
1340 Old Chain Bridge Rd.
McLean, VA 22101
P. James Terragno

Department of Energy and Natural Resources
325 W. Adams, Room 300
Springfield, IL 62706
Peter Ioes

Ministry of Energy & Infrastructure
P. O. Box 1442
Jerusalem, Israel
(representative not yet appointed)

Reference Technology, Inc.
2070 Chain Bridge Road, Suite 500
Vienna, VA 22180
Bettie Steiger

Spring Meeting Summary

It's Your Nickel: Variables in Database Production Costs.
Marjorie Hlava, Access Innovations, Inc.

Although database production costs vary from \$3 to \$60 per unit in the U.S., most fall between \$13 and \$27, with costs for government databases at the upper end of this range. Some factors influencing costs are marketing, data tagging, abstract production, indexing, coding, and the amount of data. Database production is labor intensive, especially tagging and preparing the data for input. OCR has been widely discussed, but not many use it. It is especially important to set the editorial guidelines and then abide by them. Changes are expensive; for example, a recent database changed two items in a 120,000 record file; the associated cost was about \$35,000. Abstracter and indexer burnout is also a problem; according to the January 1981 issue of *VITA*, the newsletter of the Suicide Prevention Center (1041 Menlo, Los Angeles, CA 90006), the suicide rate for abstracters and indexers is 18 times the national average.

Computerization of the production process can be done in the following ways:

- ⊕ **OCR.** The machines are expensive (over \$100,000), and the scanning cost is about \$.50/page.
- ⊕ **Combination of existing files.** One can receive machine-readable output from an online search service. Programming time then must be invested to write software to process the resulting tape. Typical programming times for one person can range from a few hours to two weeks. Machine time in a data processing shop ranges from \$5 to \$10/hour, plus a character reading charge (about \$.002/character).
- ⊕ **Downloading.** The data are transmitted over the telephone lines from one host to another. Communication protocols must be the same for both machines. Line noise can introduce errors which may be undetected. One method of checking the data is to measure the length of the character streams between carriage returns and detect abnormally long lines. Cost factors include the telephone connection, standby labor to ensure that the transmission is uninterrupted, and programming to process the resulting data (particularly if the fields are not tagged).
- ⊕ **Transfer on Floppy Disk.** The following must be considered: disk size, blocking, sector size, operating system, and density. Processing costs \$25-35 per disk plus read charges.
- ⊕ **Photocomposition Tape.** Database production from a photocomposition tape is common. Print codes must be stripped out. There are many of these, and new ones are always cropping up.

⊗ **Manual Input.** This is very expensive and time-consuming. Abstracters and indexers cannot work more than four hours at a stretch before losing accuracy. Typical costs are \$16 per hour; output is 2-5 papers/hour. Costs for digitizing or keying the data range from \$.75/1000 characters in India to \$6.50 on the U.S. east coast. Proofreading may or may not be included in these costs. It is important to insist on over 99% accuracy in a large database.

Make, Buy, Customize, or Wait?

Dorothy Ringer, Logicon, Inc.

Tradeoffs in database production were discussed. The options are buying or leasing an existing system, making a system to fit the user's needs, customizing an existing system, or waiting until a suitable system appears on the market. Buying an existing system is probably the least expensive and most rapid option, but it usually doesn't fit a user's needs exactly. Some considerations in database production are as follows:

- ⊗ **Database requirements.** How much data do you have? How often must it be updated? Is it formatted or textual? Is there a controlled vocabulary?
- ⊗ **Publishing support requirements.** Can you drive more than one device (for proof and final copy production)? Will there be more than one product produced from the database? Can the database be produced with only minimal human intervention?
- ⊗ **User interface requirements.** Is ongoing training available? Is the database integrity protected? Will the database be used for a variety of tasks (cataloging, searching, editing, etc.)?
- ⊗ **Operational requirements.** Will the system support unusual interfaces? Is it run on a dedicated or shared computer? Is the computer in-house or at a service bureau? Will existing staff be absorbed into the operation? Can service be provided in parallel during the transition from the old to the new system?
- ⊗ **Transition support** Will conversion be from an existing database? Can parallel operation be provided? Is there total systems analysis?

Logicon's preferred approach to database creation is to customize an existing system. They start with a sound system, then customize through table lookup and parameter specification, add new special capabilities, and provide personal support through the whole process.

Off-Site Indexing: A Cottage Industry

Catherine Fay, Management Contents

Management Contents (MC) is a small company with 15 full-time and 2 part-time employees. They employ 65 outside indexers to produce two

databases covering about 1,700 titles. They advertise locally, in the Chicago papers, or on library joblines for people. Response is usually 20-30 per ad. New indexers are trained at the MC office by the Database Administrator. They sign a contract for one year as an independent contractor and are paid \$5 to \$10/hour. They can process 2-7 articles/hour. Each month, indexers are ranked by productivity and by their cost to MC per article. They receive feedback from the Database Editor; the main problem is consistency among the indexers. The scheduling is rigid (one week for most journals); most indexers average about 15 to 20 hours work per week.

The typical indexer for the MC database is female, in her 30's, with two children, has an MBA or MLS, and is married to an executive. She is doing the work to keep up with her training and not for the money (which is why the indexers agree to the relatively low pay). In contrast, the typical indexer for The Computer Database is male, in his 30's, with a full-time job, and is doing indexing to keep up with his field of interest. Many Computer Database indexers are from the academic world. All the indexers like the work and the flexible hours, and are concerned to produce good quality work.

The indexers can choose the journals to work from. They come to the MC offices to receive a batch of journals and take them home to index. The Database Administrator manages the workflow and monitors productivity but has little direct contact with the indexers. The Editor gives feedback to the indexers and coordinates quality control.

MC sees many advantages to the "cottage industry" arrangement for database production. The indexers are independent contractors, so their salaries are low. Benefits do not have to be paid for them, and they do not take up expensive office space. They do all indexing manually, so there are no equipment failures to worry about. There are few managerial problems; production is maintained by salary incentives. And the indexers form a talented pool of people available for special projects.

The chief disadvantage with "cottage" indexing is the high turnover. One-quarter of the indexers last less than a year. They feel cut off because of the irregular communications with the editor. (The lack of communications is being remedied by holding in-house workshops for the indexers.) Quality control can be a problem with so many people preparing input to the database. One possible way to help quality control and improve turnaround would be to supply indexers with terminals at home so they can transmit their work electronically.

Multiple Use and Other Mysteries: Creating the Database as a Byproduct of Composition.

Bruce Foreman, American Institute of Physics

The American Institute of Physics (AIP) provides services and publishes journals and a database (SPIN) for nine member societies. Two input systems are used to produce the journals: the ATEX system

which covers most journals, and the UNIX* system which is used to produce journals for the American Physical Society. Indexing is done in-house and as a "cottage" process and costs about \$5/article. A photocomposition tape is produced from article headings; it becomes input for the SPIN database. The photocomposition codes are stripped from the tape to produce the online database. Total cost per heading is \$12-15, which includes input, indexing, and composition. SPIN indexes 17,000 to 20,000 articles per year.

Economic Concerns of Producing the NTIS Database

Marvin Wilson, NTIS

This talk traced the history of user fees at NTIS. Before 1972, all costs were from appropriated funds; NTIS was not required to recover the costs of compiling the database. Since 1976, all activity has been on a cost recovery basis. The goal has been to recover costs through user fees, with the objective of minimizing costs and increasing available products. Costs include those for acquisition and collection, cataloging, indexing, abstracting, keyboarding, and proofreading.

In the effort to minimize costs, interchange formats with NASA, the Department of Energy, and the Department of Defense were developed. A minicomputer keyboarding system was introduced, and improvements in the workflow were made. New products, such as annual indexes, weekly abstract newsletters, an SDI program using microfiche (SRIM), published searches, and a title index on microfiche were introduced. The SDI service contributes about 45% of the revenue.

Projections for the future indicate that the contribution of the paper products to the revenue stream will drop, and income from published searches will slowly increase. The database is now being leased to subscribers who process it on their in-house mainframe computers; this activity has probably peaked. The SDI service has large potential, but it is influenced by the growth of microcomputers. There is a need for a new low-priced paper product and increased usage of the database in the published search program. NTIS has a goal of holding or lowering the current price of the SDI service. Products that use microcomputer technology will increase the use of the database.

An assessment of the cost recovery program reveals that NTIS has met its cost recovery goals during the last eight years. They have reduced the staff by 40% and have controlled costs, but have still been able to develop and implement needed products. The future is bright and challenging; it will depend on microcomputer technology.

Impact of Technology on CAS Database Production

David Weisgerber, Chemical Abstracts Service

* UNIX is a trademark of AT&T Bell Laboratories.

Chemical Abstracts Service (CAS) has developed a highly automated system for database production over the past few years. It has had a significant impact on the costs and efficiencies of the operation. Between 1973 and 1983, the volume of literature grew about 40%, the number of entries in the volume subject indexes grew 65%. The professional staff increased only 11% over the same time period, and the number of production staff *decreased* 5%. The timeliness of the database has been improved by technology. In 1973, it took 130 days to get an abstract into CA; in 1983 it took 80 days. In 1973, 12 months were needed to compile the volume indexes; this was down to 115 days in the first half of 1983 and was further reduced to 100 days in the second half of the year.

The editing process is as follows:

1. An analyst reviews and dictates entries for the database.
2. A keyboarder transcribes the data.
3. Programmed edits are applied.
4. The analyst reviews and edits the data.

There are no more hard copy proofsheets and manual proofreading processes. Online editing has evolved into a distributed system using a local area network and terminal concentrator for load balancing.

Substance name editing at CAS improved 17% in the first year of the system; improvement has now grown to 44%. Abstract and subject index editing improvement was 23% in the first year; it is now 47%. About 5.5 million sheets of paper per year have been eliminated, and abstract production time has been shortened by a week. Two specific projects which improved productivity are the Author Index Manufacturing System (AIMS), which produces consistent author names from a master database, and the pre-1965 registration project, which is using a Kurzweil data entry machine for data input, followed by name match, nomenclature translation, and problem resolution.

The impacts of technology at CAS have resulted in increased staff production, lower costs, improved timeliness, and improved quality.

Converting the H. W. Wilson Indexes to an Automated System: A Functional Analysis.

John Regazzi, H. W. Wilson Co.

The H. W. Wilson Co. began in 1898 and now publishes 8 book indexes, 16 periodical indexes, two monthly magazines, and several general reference works. It has a staff of 450; its output is about 80,000 pages/year. It is a vertically integrated company, with everything from authorship and compilation to printing, shipping, and marketing in the organization. Logicon Corp. began development of an online system, called Wilsonline, in 1979, using the Medlars-II software.

About 80% of the page production has been converted to the online system.

Wilson's goals were to design a computer system, acquire and develop staff, acquire hardware and software, convert to the computer system, and develop Wilsonline for public use. Most of these goals have been realized; Wilsonline is scheduled for release in the second half of 1984. It will have 24 databases; the user will be able to search several or all of them simultaneously.

Wilson sees the benefits of their system as moving a labor-intensive project to the computer, better cost-effectiveness, economies of scale, and maintenance of editorial control and integrity.

Discussion Groups

The four groups were asked to consider the following issues: effects of technology on database production costs, in-house production versus contracting, cost recovery--which user should pay, effect of blurring of producer-vendor-user distinctions, and abstractor burnout. The results of the discussions, as reported by the group leaders were:

Group 1

The spread in cost per record as reported in the first talk is probably due to the fact that all costs are not included in the studies. Cost is influenced by difficulty of the material, kind of keyboarding, and value added to the items (indexing, etc.).

Group 2

The shift from printed to electronic products is causing pressures on database producers. Tape leasing is decreasing, and the primary product is becoming electronic. New products must be made from existing databases, and a complex pricing system will evolve. Information on users will need to be gleaned from surveys. One often hears the idea of charging is what the market will bear--how is that evaluated?

Group 3

Technology lowers costs and allows depreciation of machines, economies of scale, and staff reductions. Storage costs are decreasing, personnel costs and taxes are increasing. In-house abstracting has the advantage of allowing the producer to keep highly skilled people, but outside abstracting has lower costs. Older or handicapped people have good potential because they have good skills and generally take pride in their work.

Group 4

The "cottage" method is advantageous, especially if the work can be delivered in machine-readable form; however, programming time must be invested to process the data. Another cost factor not covered by the speakers is the increase in telecommunications costs due to the AT&T divestiture. Charging for telecommunications is likely to change; in the future, it may be based on the number of characters transmitted. Two areas where it is important *not* to compromise in database production are, first, timeliness, and second, accuracy.

Charleston Meeting Update: Videotapes for Training *Dora Moneyhun, Oak Ridge National Laboratory*

The Oak Ridge Office of Scientific and Technical Information is developing a series of videotapes to train users of the DOE/RECON system. There are six tapes in this series: (1) overview and promotion, (2-3) basic commands, (4-5) how to search, and (6) database contents. The first tape and portions of the next two were shown at the meeting. Local employees were used as actors; the script was written in-house. Copies of the tapes will be available after June 15, 1984 for loan without charge to interested organizations. Contact:

Barbara Goad
Office of Scientific and Technical Information
Technical Information Division
Oak Ridge National Laboratory
P. O. Box 62
Oak Ridge, TN 37831

Acknowledgement

Thanks are due to Betty Unruh of Dialog Information Services for sharing her notes of the meeting with me.

National Online Meeting

The Fifth National Online Meeting was held in New York on April 10-12, 1984. Most of the presentations appear in the proceedings (available from Learned Information); the keynote address by Roger Summit and the following panel discussion do not and will be of interest to ASIDIC members.

Trends in the Online Industry

Roger Summit, Dialog Information Service, Inc.

We are in the midst of a "magnificent information revolution." As recounted in Nesbitt's book, *Megatrends*, the industrial society is giving way to the information society. We are learning to become mass

producers of information; within this trend there are sub-trends and milestones which can be identified. The following are the historical trends of online searching, which began approximately 20 years ago:

- ⊕ Evolution from batch systems to online systems, which was made possible by third generation computers and the shift from tape technology to disks.
- ⊕ Development of proximity searching allowed searching of the full text online. Mead Data Central was one of the earliest services to offer full text searching.
- ⊕ The introduction of geographically diverse telecommunication networks allowed the world demand for information to be focused at a single site. The FCC deregulated telecommunications, and the communication cost became much less than the search cost.
- ⊕ Portable terminals allowed the searcher to go to the customer and allowed the vendors to give demonstrations at the customer's facilities.
- ⊕ The availability of well-formatted databases led to the concept of the "supermarket" and to specialty areas of information.
- ⊕ The library community was quick to recognize the value of online retrieval. They served as missionaries to end users and helped the market to expand.
- ⊕ Miniaturization of computer hardware kept costs down and saved storage space.

The sub-trends for the present can be divided into those on the supply side and those on the demand side. On the supply side are the following sub-trends:

- ⊕ Movement from a sampling of full text online to complete availability.
- ⊕ Movement from offline prints sent through the mail to the requester to direct delivery to the terminal (perhaps on a delayed basis).
- ⊕ New types of online information, such as graphics, chemical structures, or mixed text and graphics.
- ⊕ Prototypes of videodisks.
- ⊕ Increased evaluation of information. The user wants to find knowledge in information.

On the demand side:

- ⊗ Remote information services are being used to justify the purchase of personal computers.
- ⊗ There is an increased opportunity for information specialists as more and more end users sign up for online service. (In 1983, the ratio of end users to intermediaries signing up for Dialog was 5 to 1.)
- ⊗ More courses on information sources in non-library disciplines.
- ⊗ More effective packaged software for information retrieval. Roger predicted the day when a Visicalc-type system would be available for online retrieval and also noted that one user has characterized the Dialog system as his "favorite PC peripheral."

In the future, Roger predicts that customer growth will be healthy but not overwhelming. There will be an explosion on the demand side, and we will know that it has arrived or is soon to arrive when the following occur:

- ⊗ An online searching package for personal computers becomes as popular as Visicalc.
- ⊗ Modems become an integral part of personal computers.
- ⊗ More articles about online retrieval appear in the popular press.

Panel Discussion

The panel discussion featured seven leaders from the online industry, who each gave a short presentation of developments in their organizations.

Ronald Dunn, Chemical Abstracts Service.

CAS has become vertically integrated and is moving into the international area with the formation of STN International. In May, 1984, *Physics Briefs* will become available online; in mid-1985, the connection to the FIZ computer in Karlsruhe will occur. A Japanese contact is under investigation.

William Marovitz, BRS.

BRS is in the business of information delivery and is experimenting with videodisks, graphics, and a hierarchical network of personal computers. End users are rapidly signing up for BRS service, and in organizations where an intermediary has paved the way, the positive effect on revenues is particularly noticeable. Pricing is a problem; connect hour charges have been lowered, and charges for information units have been increased.

Cynthia Hull, SDC.

SDC is being restructured. System support will be increased, and there will be more sponsored file arrangements such as the Derwent files. A new contract has recently been obtained with the U.S. Patent and Trademark Office; an end user interface for trademark lawyers is under development. Digitized images will allow graphic display; phonetic and syllabic searching will be available.

Robert Donati, Dialog Information Services.

The new Dialog-II system will feature enhanced software, new products, and additional databases. New products will be electronic mail, an alternate form of delivery for offline prints and SDI output, and a report system giving the user more ability to process and manipulate data. Specialized training courses for the new system are being developed. In the area of databases, full text databases such as the ASAP files from Information Access are planned, as well as more health information and financial database (a contract with Moody's was just signed). The Derwent database will be available in October.

Anne Saunier, Mead Data Central

Mead changed its dedicated terminal policy last year so that a few "approved" terminals can also be used to access LEXIS and NEXIS; other terminals will soon be approved. Usage of the New York Times databases is growing rapidly. Pricing for NEXIS was changed in response to many requests from the library community; the pricing structure is now \$20 per hour plus a search charge that ranges from \$9 to \$18. A third mainframe has been added, and better training is being developed. An SDI service is also planned.

Eugene Garfield, ISI.

ISI offers its databases to all vendors, but some of them have found no takers. Users also may lose sight of the fact that SCISEARCH has more data on some subjects than discipline-oriented databases. The ISI Search Network has been discontinued because of bad response times, operating system difficulties, and problems educating users to new search concepts (such as research fronts). A new system containing an "atlas of science" and *Current Contents* online is planned. *Index Chemicus* and a chemical citation index may be mounted on the Questel system. The SCI-Mate system may be used to distribute ASCA profiles on floppy disks.