

# ASIDIC newsletter

No. 53, Fall 1986

---

ASSOCIATION OF INFORMATION AND DISSEMINATION CENTERS.

## Postprocessing Discussed at Fall Meeting Marjorie Hlava Re-elected ASIDIC President

The Fall, 1986 ASIDIC meeting was held at the Doral Inn in New York City on September 14-16, 1986. The technical program, entitled "Beyond Online Retrieval: Value-Added Postprocessing," was under the direction of Cathy Ferrere, ASIDIC Program Committee Chair. It featured 12 presentations on the subject; following the custom started at the Albuquerque meeting, an exhibit area offered attendees the opportunity of viewing several postprocessing systems. A summary of the technical presentations appears in this Newsletter.

Attendance was about 60; although it was not as high as some recent meetings, the Fall meeting was still successful. As customary at Fall meetings, elections were held (see below). Local host was Rita Lerner (AIP); thanks to her arrangements, attendees enjoyed a delicious Chinese banquet at the Hee Seung Fung restaurant.

### Election Results

The Nominating Committee (David Grooms, Lois Granick, and Dan Wilde) submitted the names of Marjorie Hlava for re-election as ASIDIC President and Dennis Auld for Member-at-Large. Both candidates were unopposed; since there were no additional nominations at the Fall meeting, both were declared unanimously elected and were congratulated at the meeting.

### Committee Reports

*Finance:* Members are urged to submit their dues for 1986 as soon as possible; they are running behind schedule. Recent meetings have been larger than average; revenues and expenses have therefore increased. Scott Kostenbauder and the Finance Committee have completed an audit of ASIDIC's finances; they are in good condition. A Certificate of Deposit for nearly \$15,000 will be examined to see if a more favorable interest rate (currently 7.35%) can be obtained.

*Membership:* Three new members have joined ASIDIC since the last meeting; they are listed below.

*Executive:* Because of the many financial decisions that are made at Executive Committee meetings, the Chair of the Finance Committee will attend meetings of the Executive Committee in the future. Improved procedures for transmitting funds from the Secretariat to the Treasurer were established. By-Laws changes were recently made; no further changes are planned. The publicity brochure for

ASIDIC is progressing well and should be distributed soon. ASIDIC's status as a tax exempt organization was recently reaffirmed.

### **Spring, 1987 Meeting**

The Spring, 1987 ASIDIC meeting will be held at the Royal Sonesta Hotel in New Orleans, LA, on April 4-6, 1987. Its topic will be "What Makes a Database Profitable: The Need for Profitability in the Industry is Dramatic." Program Chair for the meeting will be Morris Goldstein (Information Access Corp.) assisted by Walter Beveridge (IBM). Further information will be distributed later; mark your calendars now and plan to attend!

### **Future Meetings**

The Fall, 1987 meeting will be at the Viking Hotel, Newport, RI on September 20-22, 1987. The Spring, 1988 meeting will be in San Antonio, TX in March, 1988. Current plans are for CISTI, the Canadian information center, to host the Fall, 1988 meeting and advise the Executive Committee on a suitable Canadian location.

### **New Telephone Number**

The telephone number for the ASIDIC Secretariat has been changed to (404)-542-6820.

### **ARIST Call for Papers**

The chapter authors for Volume 22 of the *Annual Review of Information Science and Technology* (ARIST) have been chosen. If you have written any papers relevant to their subjects which you would like to submit for review, please send a copy to the author. The chapter subjects are:

- Retrieval Techniques
- Information Systems and Services in Agriculture
- Image Information
- End-User Searching of Bibliographic Databases
- Micros and Library Automation
- Natural Language Processing
- Marketing of Information Products
- Information Policy in Canada
- The Electronic Office
- Optical Digital Storage

- Artificial Intelligence

Contact the Newsletter Editor to obtain the name and address of the author of any chapter you are interested in.

**In Memoriam**  
*James Carman, 1926-1986*

Dan Wilde read the following statement at the Fall meeting:

"It is with deep regret that I must announce to the membership of ASIDIC that Dr. James Carman passed away on Saturday, July 26, 1986 at the age of 60. For those of you that did not know Jim, he was a founding member of ASIDIC. He was a major participant in those early meetings with Chemical Abstracts Service on how mechanized bibliographic information could possibly be useful. Jim was Director of the University of Georgia Computer Center and managed their Information Retrieval Center, a very early tape spinner. At his death, he was Assistant to the President for Computing and also Assistant Vice Chancellor for Computing Systems.

Not only was Jim a founder of ASIDIC, he was an early pioneer in information retrieval research. With a Ph.D. in statistics, he was well founded to undertake the very early user studies on how mechanized bibliographic information might benefit end users. His early papers contrasting *Chem Titles* and *Chemical Condensates* are such examples.

Jim worked hard and he also played hard. He helped establish the tradition that ASIDIC meetings should be fun. In particular, I remember those early meetings in Atlanta, Jim's home territory. I also remember the first Chicago meeting and the Playboy Club. Those of you who knew Jim remember your own good times with him.

But most of all, I remember Jim Carman as a friend and a true gentleman. To those of us who were new to the business, he was very patient and always willing to pass on his experiences. Those of you were fortunate to know Jim know you have lost a friend. Those of you who did not know him are probably here today because of his early contributions to our field."

The Executive Committee is investigating the possibility of a Distinguished Lecture to be held in Jim's memory.

**Government Information Act of 1986**

This Act was introduced into the House of Representatives on August 13, 1986 as H.R. 5412 by Rep. George E. Brown, Jr. (D-Calif.). It would establish a Government Information Agency (GIA), mandate that all information produced as a result of a Government contract be provided to the GIA in machine-readable form, and require that all such information be placed in one database. Since this bill is of great interest to ASIDIC members, the item from the *Congressional Record* containing Rep. Brown's speech introducing the bill is attached to this Newsletter. ASIDIC committees will be studying the Act in the coming months.

**President's Column**  
*by Marjorie Hlava*

Success comes to those who succeed! And succeed we did at the New York meeting in September. The technical program chaired by Cathy Ferrere of IEEE brought out several new points for producers and vendors to consider now that downloading is here and, when searching with a microcomputer, necessary. The demonstration exhibits drew great response, and, I suspect, at least one new business arrangement.

Of course, it wouldn't be an ASIDIC meeting without social arrangements designed to facilitate discussions among the membership. Rita Lerner of AIP put together an excellent local arrangements package from the coffee breaks to the Chinese dinner with the whole two-story restaurant to ourselves. Thanks Cathy and Rita--a job well done!

The next meeting will be in New Orleans on April 4-6, 1987, at the Royal Sonesta Hotel. Taissa Kusma of AMS is taking care of local arrangements, and Morry Goldstein of IAC has agreed to chair the meeting.

On other fronts, ASIDIC committees are reviewing the new Government Information Act and would welcome your comments on it. The text is reprinted in this Newsletter.

Our Association is in solid financial condition and is growing at a markedly increased pace.

See y'all in New Orleans!

**Book Reviews**  
*by Helen Citron Wiltse*

**Pricing Policies for Parallel Publishing**, by Peter H. Muller and Roy Wilson. Oxford, England, Elsevier, 1985. (EIB report series, no. 6). 74p.

This interesting analysis of print vs. online pricing did not provide new ideas; it substantiates those already known. The fact that online services are priced and handled by publishers as a by-product of the print services has always been assumed on the part of the users. This analysis proves that many costs of producing both the online and print are completely assigned to the print products. Increases in journal prices are traced side-by-side with the online price increases. The journal increases are substantially more.

The different approach of this publication is in looking at the marketing perspective of attracting additional users. It forecasts a redefinition of online pricing as scholars, scientists, and individual small corporations become the end users. The technology of providing end user searching is discussed only briefly.

The fact that all statistics and data base producers cited are British or the British counterpart of American ones is only a slight deterrent to the discussion. Muller and Wilson have stopped short of suggesting pricing policies. They have, however, indicated production costs which should be shared by both formats and not attached to just one.

The authors have attempted in an impartial manner to provide the publisher's view of the pricing dilemma. They have also succeeded in providing many end user concerns about the costs of online services.

This publication is well worth reading. The numerous charts of comparison data require some study to obtain the information. Both end users and publishers will gain from this work.

#### **New Members**

ASIDIC welcomes the following new members:

American Society of Hospital Pharmacists  
4630 Montgomery Ave.  
Bethesda, MD 20814  
Mr. Dwight Tousignaut

I.S. Grupe, Inc.  
948 Springer Drive  
Lombard, IL 60148  
Mr. Peter Schipma

OCLC  
P. O. Box 7777  
Dublin, OH 43017  
Ms. Mary Ellen Jacob (returning member)

#### **Papers from Previous Meetings**

As one of the benefits of ASIDIC membership, reprints of three papers presented at the Spring 1985 meeting in Clearwater are enclosed.

## FALL MEETING SUMMARY

### **Strategic Information in the Technical Professional Office,** *Scott Kostenbauder, IBM.*

Right decisions are critical to the success of any enterprise; information helps technical management not to make wrong decisions. History shows us the value of information; recognizing this, many people today are selling information services. The computer makes information retrieval possible to all, but we must know what is the best information available. Most of the information retrieved today is in raw form, which has limited value; synthesizing it helps lead the user to the desired result. In many environments, decisions are made at the lowest levels possible; therefore searching systems are now found everywhere.

The IBM Technical Information Retrieval Center (ITIRC) provides a window on the world. They have found that many technical people are not inclined to go to a library; often "Charlie down the hall" is consulted. The main sources of information in the office are peers, the telephone, printed sources, and the computer. Computerized searching allows technical professionals to do complete literature surveys; productivity in the office can be enhanced.

At ITIRC, reaching the data is easy, fast, and reliable through high speed circuits. If access is too slow (such as 1200 baud dialup), people won't use the information. The most popular databases are brought in-house; this is expensive until many users use the system. Internal information is integrated on the same system, providing a single source for most needs. Users should not need background information to overcome system mechanics; menu selections are brief, complete, and can be bypassed by experienced users. Users also need to recognize when they need professional help; an in-house hot line is provided for that purpose. Major needs are immediate access and SDI service; timeliness is important. Coverage must be comprehensive and broad--ideas can be anywhere. An information system is the most useful when users get the right information easily; it will be used when scientists can be convinced that the system receives the information before their colleagues hear about it or before it arrives in the library.

### **Postprocessing and the Scholar's Workstation** *Victor Rosenberg, Personal Bibliographic Software*

What do we do with information after retrieving it? Large databases were developed for large mainframes. Now, however, we have small computers, and not everyone wants to drink from the fire hydrant! "Retail" databases and systems (boutiques) are growing, but simply mechanizing what exists is a misapplication of technology. The intelligent terminal, popular a few years ago, is also a misapplication because microcomputers are more powerful.

End users have not been involved much in information because it hasn't been easy to deal with. Regardless of how much information is available, the dominant factor in using it is the ease of use.

A postprocessing system should have the following functions:

- Combining information in to boutique systems distilled from large databases.
- Formatting and reformatting the information.
- Indexing, both automatic and manual.
- Eliminating duplicate items.
- Typesetting and printing using systems like the Apple Macintosh.
- Generation of information products (document orders, etc.)
- Addition of notes to documents.
- Automatic production of bibliographies.
- Catalog card production.
- Merging several databases into one.
- Use of the data as input to a large mainframe-based system (uploading).

Just giving the user information is no longer enough; we must also give the tools to manipulate it or make it amenable to manipulation. We must link systems so that it is possible to go directly from retrieved items to a publication without intervention if desired. The essence of postprocessing is to make the fire hydrant into a water cooler.

**Competitive Marketing of Information Files: Blending In-House and Commercial Data**

*Pamela Danziger*, User Technology Service, Inc.

A competitive intelligence database has the following features:

- It contains information collected from scanning the competitive environment.
- It is used to identify trends affecting business operations by strategic planners, public relations people, etc.
- It is owned, operated, selected, and processed by a single organization.

Competitive intelligence databases are important to organizations planning their future and trying to get an edge on the market. Therefore, such databases require focused timely information and must present diverse resources in a single system. To build a competitive database, one must first identify the need and the information sources, then the system specifications and resource requirements, before getting management approval to proceed.

Conventional bibliographic online databases, while somewhat useful in this context, suffer because they are too slow, too general, too hard to use, and often too academic. They do not reflect a business organization's view of the world.

Building a competitive intelligence database is costly; scanning the literature or converting existing files costs \$8-20/record, and 10,000 to 30,000 records are necessary for the database to be useful. An alternative is to blend in-house data with commercially available databases. The external data can be leased or downloaded; in-house data can be entered directly into the database. Databases to be used should be those covering an industry completely; their data may need to be customized to meet the user's needs.

Input issues include varying data quality, preprocessing and reformatting requirements, global algorithmic changes followed by manual scanning and editing of records, disk storage requirements, and merging differing databases. Building such a database provides opportunities for an information center to increase its contribution to an organization and provide professional growth for its staff. Information center members can become members of the strategic decision team.

#### **The DOD Gateway Information System**

*Marjorie Powell*, Defense Technical Information Center

The DTIC gateway is used by both intermediaries and end users to access databases; it provides powerful postprocessing features. Functions include merging output from several databases, eliminating duplicates, reformatting, graphical trend analyses, and online review. Citations can be algorithmically transformed to a common format; files can be merged to produce a unified bibliography. Statistical analysis on the retrieved items is a unique feature of the DOD gateway. The system can count the number of occurrences of a term in a field, then provide cross correlations between fields (author-descriptor, author-author, etc.), and present the results graphically. References can be evaluated for redundancy, and indexes can be made. Reformatting features allow citations to be reviewed, fields to be dropped, citation styles to be unified, etc.

Most of these gateway features are still in the prototype stage; a PC implementation will follow. The gateway runs under the UNIX system; original development work was done at the Lawrence Livermore National Laboratory.

#### **Automating Drug Information: From Print to Database to Print**

*Dwight R. Tousignaut*, American Society of Hospital Pharmacists (AHSP)

AHSP is changing to automated production of its products, but it is still in a print mentality; 2/3 of its income is from publications. All publication activity is self-supporting. The first automation was in 1969 with the production of *International Pharmaceutical Abstracts* (IPA). In 1973 IPA became part of Toxline, and in 1978

it went up on Dialog.

Postprocessing has been done using the online database. Spinoffs from the electronic publication were a subject index and subsets of the database with their own indexes. These began as batch printouts of items having specified vocabulary terms; now the database is searched online using Boolean logic and used to produce the subsets. This service has been sold to other organizations; AHSP makes the indexes for *Drug Information Journal* and other publications. They are now producing a tradename list with generic names indexed against tradenames and CAS Registry Numbers. In 1983, production of *Drug Information Fulltext* was automated, and the file was mounted on Dialog; in 1984, the *Consumer Drug Digest* went online. A floppy disk product with menu-driven interaction has also been prepared.

All this automation has been done without automating the in-house operations; ASHP still uses a manual production system which operates by moving paper around. ASHP therefore views postprocessing as a by-product of its main tape production function.

**TEX Software: From Online Output to Typeset Mathematics**  
*Taissa Kusma*, American Mathematical Society (AMS)

In the 1970's, the computerization of *Mathematical Reviews* was begun, culminating in 1981 with the introduction of the Mathfile database on Dialog and BRS. A major problem was the plethora of non-ASCII characters and complex mathematical equations in the databases. The following three propositions for dealing with the problem were proposed:

- Put only the bibliographic information online, eliminating the abstracts, where most of the problems lay. This would have severely limited the usefulness of the database.
- Eliminate the equations and substitute a string of asterisks for them. This looked bad and was rejected.
- Put the typesetting information online. This was not acceptable to mathematicians who were used to good looking equations.

A program was therefore developed to translate objectionable typesetting symbols to mnemonics and linearize the equations. When a version of the file with linearized equations was put online; expected protests from mathematicians did not materialize. Although the file was used and well received, there was no way to convert the linearized equations back into typeset ones when offline prints were generated. Then the TEX typesetting software was developed, not as an online postprocessor, but for producing high quality books. With a little modification, TEX was able to postprocess offline prints and produce output with the high quality equations mathematician had become used to. The AMS therefore decided to produce *Mathematical Reviews* with TEX and, in 1985, asked Dialog to produce offline prints from Mathfile with it.

Dialog declined because of the substantial financial investment which could not be justified for only one file, so AMS wrote a program to covert typesetting codes in Dialog and BRS records and process them using TEX. They now market their own version of TEX for this purpose and have found high interest, especially in a version for the Apple Macintosh. TEX has changed AMS's marketing strategy and is a good example of what can be achieved with a postprocessing product.

#### **Possibilities and Potentials**

*Robert F. Jack, NASA Technology Transfer Center*

Postprocessing is closely linked to downloading and uploading; it should pose no threat to database producers. The NASA Technology Transfer Center uses the Lawrence Livermore gateway and a NASA program called SORTAID to add category codes, delete or add information to records, move items around, and rank them. Ranking is done by counting word frequencies, then ranking the occurrences of words to compute a value of relevance. SORTAID gives the user the ability to control the use to which information is put; it does not give access. It allows sorting the retrieved information by the easiest way to retrieve the original documents (which is not the "last in first out" order one usually obtains by default on online retrieval systems).

Postprocessing adds value to an information product. It creates something useful for end users. Answers solve problems, bibliographies don't, but bibliographies can be used to find the answers. Intermediaries are experts at creating useful products for end users, and they have long been using and marketing online databases to their users. The current focus is on the end users; producers must not forget the intermediaries and should listen to and value their views. There is an adversarial position between end users and database producers; the end user wants to spend as little as possible, and the producer wants to sell as much as possible. Intermediaries are in the middle and can be a powerful force to help the database producers.

#### **Solution in Search of a Problem: Why We Can Survive Without Postprocessing**

*Ellen Shedlarz, McKinsey & Co.*

McKinsey spends about \$1/2 million in searches on Dialog but does little or no postprocessing. They are an international consulting firm; their clients use them to draw on their problem solving expertise, build capabilities, obtain independent third-party opinions, or as a catalyst for change. McKinsey uses online databases because of their convenience, efficiency, accessibility, speed, flexibility, coverage, and currency, but *not* because of their accuracy. They see no need to do postprocessing if they cannot depend on database accuracy. They do not use the databases most useful for postprocessing such as Disclosure, Predicasts Forecasts, and the Bureau of Labor Statistics. Factors in their environment include:

- Multiple use of the data by client teams.
- Separation of internal and external data because of confidentiality considerations.
- An internal production department providing client support.
- Heavy reliance on textual data in contrast to numeric.
- Individual value added benefits with clients doing their own evaluation and postprocessing.

Problems affecting the use of online databases are:

- Data sources are not specified. Some databases (D&B, DRI) have no dates; others do not indicate their sources (Disclosure).
- Terms are not always clearly defined; the user must depend on the definitions of others.
- Confidence is eroded by poor consistency, inaccurate data, and errors in the data. Databases are therefore regarded as a tool, not the final source.
- There are often contractual or legal restrictions on the use of the data.

For these reasons, McKinsey does little postprocessing. To change this situation, downloading restrictions would have to be eased, and it would have to become easier to manipulate the data and merging it with data obtained from internal sources.

**Overview and Insights, *Joseph Bremner*, Database Development.**

It is important to realize that most postprocessing is in the area of numbers, such as financial or econometric data. There is a strong move to decentralize this processing, transferring it from the vendor's mainframe to the user's PC using appropriate software. Many companies now have proprietary packages allowing users to download data, place it into a cell on a spreadsheet, scan multiple databases, and then analyze it. These systems are difficult to use; most of them can be found in the offices of portfolio analysts, corporate financial experts, and so on. These people are just discovering online databases. They *expect* to download and reuse data and are unhappy that they cannot use the powerful and sophisticated products they are accustomed to. For example, lawyers would like to store information on cases and statutes on their PC's, but cannot download information without Mead's proprietary software. Database producers should recognize that people are used to downloading in other environments and base their market strategy on those needs. These end users will expect to download data and use it in their PC; they will need application tools for it. Advancements and improvements will come from specialization and niche products; Dialog cannot serve everyone!

## EXHIBITS

The following products were demonstrated at the exhibits:

- CD/Corporate Database (Datext, Inc.)
- TEX (American Mathematical Society)
- Sci-Mate (ISI)
- Pro-Cite (Professional Bibliographic Software)
- UMI Information Delivery Module (UMI)

Following the exhibition, brief presentations described three of these products.

### **UMI's Information Delivery Module**

*Pauline Smillie*, University Microfilms International

Secondary publishing is a form of postprocessing; new ways of access, such as optical disks, are appearing. In such an environment, easy access to the full text of documents will be important and it should be only a keystroke away. The journal is the primary mode of information distribution; their number doubles in 15 years, and in some fields, they are growing even faster. Even using the high storage densities available on optical media, it is difficult to store the full text of journals; one year's worth of the journals in Medline would fill 900 compact disks!

Tools available for accessing the full text of documents are online databases, providing quick identification of citations, near instantaneous transmission of orders, and quick processing times. But it still takes too long to get the right items to the users.

UMI is developing the Information Delivery Module (IDM), which will provide communication among library networks, online vendors, bibliographic utilities (OCLC, etc.), and document fulfillment vendors. The IDM has an attached CD player and includes database access software. It is a single point of access but requires much cooperation among vendors. The value added in the CD system is continuous selection, local searching and printing, currency, ease of use, and a linkage to document providers. A major problem is turnaround in disk production; if disks cannot be made fast enough, the product should not be put on disks.

### **CD/Corporate for Business and Finance Information**

*Stephen Cucchiaro*, Datext, Inc.

Datext serves the business and financial community using PC's and compact disks. After several months of market research, it designed its product--an integrated source of business information from several databases. People were doing without information because access was too difficult; the Datext system can be used without formal training. Their market study showed that people wanted one-stop access to information on a company, a group of companies, a line of business, an industry, or an executive.

Datext's product integrates information from Predicasts, ABI/Inform, Business Research Corp. (Investext), Media General, *Who's Who*, and Disclosure. They created one master database,

defined company, industry, etc., and made it simple to transfer information into a PC Software Package for further manipulation. The information is stored on four compact disks, according to industry sector (for conglomerates, the information is replicated on several disks as needed). Product development took over a year, then six focus groups tested it. The focus groups indicated that timeliness was important, so the updating interval was changed to monthly from the original quarterly.

The CD/Corporate product is targeted at Fortune 500 companies, financial institutions, and large service organizations. Early sales show that it is generating new demand, not taking from existing markets. It is increasing the awareness of the value of electronic information. Use of CD technology is viewed as complementary to online databases, not in competition with them. Datext's next product is CD/Corptech--a directory of 14,000 high technology companies indexed by product and market served. They will soon connect to the Dow-Jones News Retrieval Service, providing a convenient access to very recent information.

The response to Datext's products has been excellent. Its value is that it does more than just retrieve information; it organizes it and makes it easy to do competitive analysis.

#### **Customizing Search Results: The Sci-Mate Solution**

*Cathy Mundy, ISI*

Sci-Mate was introduced three years ago by ISI and is now in its fourth version. It is designed to help the user process search output. Originally two modules, a Universal Online Searcher and a Personal Data Manager, it was expanded to three this year with the addition of an editor.

The Universal Online Searcher sells twice as many copies as the other components; it is a front end searching five online systems using the same interface. It supports the common mechanical functions of automatic logon and offline strategy preparation, but does not provide help with database selection or search strategy assistance. Because it interfaces to five systems, the Searcher requires the most customer support. Its market is small and requires much searcher education. It is not meant for professional searchers but casual users.

The Personal Data Manager is Sci-Mate's most generic module; it provides local file management. Users can create files from downloaded data, manually entered information, or existing text files. Many users of the Data Manager do not use the Universal Online Searcher but search with other menu-driven packages (BRS/Colleague, etc.). Locally created databases are searchable; templates for reformatting are available. Applications include the delivery of searches on disks by librarians for further searching by the requesters, handling reprint collections, managing non-bibliographic files, and teaching online searching.

The editor was introduced in early 1986; it can reformat data using "style sheets" to conform to differing publisher's styles. It provides flexible output and can be integrated with the Searcher and Data Manager.