

Spring Meeting Discusses Implications of New Technologies

The Spring, 1988 ASIDIC meeting was held in San Antonio, TX, at La Mansion del Rio Hotel. Attendees from the northeastern portion of the country delighted in the warm days, leaves appearing on the trees, and a few flowers in bloom. They also enjoyed exploring the Alamo and related sites in the old city of San Antonio.

The technical program was arranged and directed by Bill Bartenbach (Foundation Center), Dave Grooms (Patent & Trademark Office); local arrangements were under the able hand of Harry Allcock (IFI/Plenum Data Corp.). The usual reception and buffet preceded the meeting; the Monday night program was a dinner and dance. The technical presentations are summarized below.

Committee Reports

Finance: Randy Marcinko (Dynamic Information) presented the budget and balance sheet, which show that ASIDIC continues in healthy financial state. Studies are underway to see if there is a simpler way to report financial data to the Membership.

Executive: Results of the questionnaire recently distributed to ASIDIC members have been received. There are no surprises; members like the meeting format, the dates, and the days of the week that meetings are held. There was a desire to strengthen the topics for meetings and to provide more support to meeting planners. The new Executive Committee will continue the work already started—preparation of an ASIDIC handbook to document the responsibilities of Committee Chairs and to describe their functions.

Publication: The Executive Committee likes the present format of the Newsletter and is working on ways to provide more support for the Editorial and Production processes.

Future Meetings

Fall, 1988: The Fall meeting will be held at the Claremont Hotel, Oakland, A. The Claremont, in a beautiful setting, is located at the foot of the Berkeley hills, and has lovely gardens, full sports facilities, and other amenities. It is a long-time fixture of the area and has been extensively remodeled and updated. It will be a fitting setting for an ASIDIC meeting. The meeting will take place September 18-20, 1988; its topic will continue the emphasis on new technology. An effort is being made to recruit major players in the information industry as

speakers. Because 1988 is the 20th anniversary of ASIDIC, a special observance, coordinated by Taissa Kusma (American Mathematical Society), is planned. Be sure and save the dates for this significant event.

Future meetings: Sites for future meetings were decided by the Executive Committee, as follows: Spring 1989 in Charleston, SC; Fall 1989 in Chicago; Spring 1990 in Miami; and Fall 1990 in Denver.

President's Column

by Dennis Auld

"Remember the Alamo"—or in specific ASIDIC terms, "Remember the price." Hot topic, great meeting. Bill Bartenbach, David Grooms, and Harry Allcock put on a meeting which not only created a forum for discussion, but drew on ASIDIC's ability to be flexible and allow the meeting's topic to be as fully plumbed as possible. Casual discussions with several attendees in the following weeks reinforced the assumption that the meeting had true impact. As a result, the topic of products and pricing in new technologies will be expanded on at the fall meeting in Oakland.

Behind the scenes, important issues relating to the operation of the Association were also taking place. As indicated in the Membership's returned questionnaires, the two meetings a year are the most important function of the Association—a place and time to meet and discuss issues relating to the flow of information. To that end, all current committees are embarking on a process aimed at establishing guidelines to better effect their tasks. The Executive Committee is spearheading this project and plans to develop an "Association Handbook" over the next two years.

To give you an example, the Executive Committee meets at each conference. One item on the agenda is to select future sites and dates for the meetings. Past practice has been to attempt geographical balance by moving the meetings between east and west, and to select dates by picking the usual March-September time slots. The process has usually worked well, but scheduling conflicts have appeared now and then. The geographic selection process, given the parameters in the survey, will largely remain the same. The date selection will now be improved through the process of monitoring events scheduled by other organizations that ASIDIC members attend. Not only will we factor in those dates for possible conflict, but the future ASIDIC conference dates and locations will also be placed in those publications and databases.

As you know, 1988 is ASIDIC's 20th year as an Association. Plans are currently under way to save a portion of the program in Oakland to recognize this fact and present to the Membership some history and accomplishments since its founding. With the festivities surrounding the anniversary and the program topic of marketing value and new technologies, presented by what is shaping up to be an outstanding list of speakers, it certainly will be a meeting you can't miss.

I will keep you informed about the progress of the Committees. The energy and abilities displayed by the members is truly inspiring, and with Jeannette Webb continuing to be the "glue that holds the effort together," progress will be assured.

See you in Oakland!

FALL MEETING SUMMARY

Product Applications and Pricing Implications of New Technology

Pricing New Products as a Result of New Technology.

Harry Allcock

IFI/Plenum Data Corp.

IFI recently developed a chemical patent database, and, while technology helped its development, cost considerations caused many problems, both for the producer and the users. IFI first conducted some market research, surveying 16 chemical information managers to get their opinions on how the database should be structured. Features they requested included substructure searching, link operators, role indicators, and polymer indexing. When IFI discovered that DuPont had already developed a similar database, they purchased it and spent a year merging their database with DuPont's before marketing the product. Of the 16 managers advising IFI, only 12 bought the system. IFI then marketed the database publicly, which raised costs. The Chairperson of the advisory committee suggested dropping the fragmentation codes to cut costs, but others wanted the system enhanced, causing another increase in costs and prices.

Costs continued to rise for the users, primarily because of high prices charged by their in-house computer centers, so IFI agreed to load the database on a commercial online system. Advantages of this step were immediate response and interactive searching; disadvantages lay mostly in the area of online printing. Some users balked at paying for prints when they already had the database in-house, not realizing that IFI had to pay the vendor to carry the database, and this cost had to be borne by them. The database was originally available only to subscribers, but for a time, it was opened to all at a price of \$150/hour (subscribers paid \$300/hour). After three months, surveys showed that most subscribers continued to use the private version of the database, even though it was more expensive. A few new users liked the file, but when IFI asked them to become subscribers, they declined. The database will be publicly available until July, 1988, then users must be subscribers to continue using it. Conclusions of this experiment are:

- Many people are prefer to pay a high online price rather than become subscribers.
- IFI needs the subscription revenues to be able to maintain the database. Online fees do not bring in enough revenue, even if they are high.
- If non-subscribers are allowed to use the file, subscribers will protest and will not renew their subscriptions, causing producer's revenues to decrease.
- Database usage must produce enough revenue to support production costs.

New Pricing Strategy—A Result of Technological Trends

Jim Seals

Chemical Abstracts Service (CAS)

The connect hour has long been a major basis for pricing in the online searching industry, but advancing technology has recently brought it into question. CAS observed that, although usage was growing, connect hours had plateaued; therefore, they concluded that the connect hour is not a proper measure of the database's value. Technologies causing revenues to decrease are:

- higher baud rates,
- ability to download information rapidly to microcomputers,
- front end software to prepare queries offline and then rapidly upload them,
- extraction commands allowing data to be taken but not displayed, and

— the ability to use the same query in multiple databases.

The economics of the online retrieval business are also changing. Print subscriptions are declining; CAS now receives less than 50% of its revenue from subscriptions to the printed *Chemical Abstracts*. Online revenues are therefore becoming increasingly important in database production. Current online prices are unrealistically low; because of budgetary pressures on libraries, there is a limit to price increases. CAS does not see a large increase in the online market, but more revenue is needed; even a large growth in usage would not make up for the revenue loss. Finally, "traditional" pricing methods exacerbate these problems.

CAS feels that database pricing should:

- Reflect the value of the database, including the amount and kinds of data available,
- Reflect the value of the information retrieved and the amount displayed, and
- Be independent of the technology used for retrieval, whether it is online or offline, high speed or low speed, etc.

New use fees were therefore introduced in 1988. These include charges for search terms used (10¢/text term, \$20/substructure), charges for the amount of information displayed, and a charge for the information "extracted." Although CAS began to levy these fees upon its licensees, the licensees were not required to pass them on to their users. The new fees are initially low but can be expected to increase. Although the new prices reflect a move away from connect time as a charging algorithm, connect time rates were not lowered in 1988; users can expect that they will be lowered in the future.

When CAS announced their new fees, users voiced many concerns. They were concerned that the new price structure was complex and would make predicting search costs difficult. They need to predict search costs for budgeting; CAS replied that it is not easy to predict connect time anyway. Users are concerned that the new pricing algorithm will motivate searchers to use fewer search terms and therefore miss information; CAS has seen no evidence of this.

CAS feels that the new fees are predictable; one just counts the number of search terms used to predict cost. They are more fair because there is no more benefit to fast typists or to those who upload their searches after preparing them offline at no cost. And they have met the three features that CAS feels are necessary in database pricing.

Gateways: EasyNet's New Pricing Attempt

Dick Kollin

Telebase Systems, Inc.

Connect time pricing originally began in 1969 with the Dialog-Pandex contract because database vendors did not want too many simultaneous users on the system. It made prices as predictable as possible and appealed to the untrained person or occasional user who did only a few searches a year. Occasional users are a large market; EasyNet's typical user might be a 16 year old in a rural area searching at night from home. Home and corporate users did 550,000 searches in 1987; less than 1% of EasyNet's usage is from libraries because professional searchers want fast response and are not willing to put up with menu-driven systems.

After a few searches, end users begin to make demands on the system, wanting faster response. So EasyNet developed a system based on the ISO standard's common command language. The new system was liked by professionals, but its flat-fee pricing structure has resulted in increased prices for libraries used to searching inexpensive databases. The libraries wanted to be able to offer their customers public searching with a way to easily charge the customer for the search. EasyNet's answer is a "battle-hardened" cover over standard searching equipment: a PC, color monitor, laser printer, etc. The user simply puts a credit card in a reader, and the search costs are charged directly to them. Libraries can therefore offer online searching to the public without any stress on their budget. Alternate pricing arrangements are available if the library wishes to subsidize any part of the search costs.

The "Answer Machine" was tested at the University of Pennsylvania and received an overwhelmingly positive response. EasyNet expects to produce over \$10 million in revenue for database vendors and producers in 1988.

Price Implications of Implementing Word and Phrase Hit Charges

Geoff Sharp

Dialog Information Services

Pricing online services is a complex issue. There are many different market segments and many different kinds of databases and applications. Online accounting systems are complicated and burdensome; they represent a significant overhead to the vendor. Pricing therefore needs to be kept as simple as possible, both for users and vendors.

There are now many different kinds of databases in addition to the original bibliographic ones, and also many different applications such as text searching, report generation, full text, CD-ROM, and multi-file searching. Many different market clusters and strategies exist as well; database providers and vendors must therefore be pragmatic and willing to make compromises.

Information has a different value to different people. A company name field, for example, may be worth only 10¢ to a direct mail firm, but it may be worth \$10 to a salesman on a call, or \$100 to someone doing a credit check, and \$1,000 to someone searching for a merger or acquisitions target. How can we price information fairly for all users?

Accounting systems impose considerable overhead on a search service. In Dialog's case, there are 5,000 possible connect time prices, 700 SDI prices, over 100,000 possible printing prices, as well as other miscellaneous prices. The total is over 1½ million cells in the price matrix, all of which must be tracked by the accounting software. Adding prices for specific data elements would add another 5 million cells. As another measure of accounting overhead, Dialog spends 188 hours/month doing basic accounting, 65 hours generating various usage reports, and 48 hours printing invoices. These 301 hours each month fill 800 tapes with data and use 20 large boxes of paper to report 86.5 million transactions. Adding data element pricing would increase the number of transactions to over 350 million, increasing the vendor overhead and, ultimately, the prices to the consumer.

Revenues in the online industry must be grown through products and marketing, not by changes in pricing methods. There needs to be an outreach to more market segments so that the price-demand curve is pushed out, not merely slightly adjusted. Pricing should reflect daily usage like utilities.

The Image System at the Patent & Trademark Office

William Maykrantz

U.S. Patent & Trademark Office (PTO)

The PTO has over 20 million patent documents in its files on over 140 million pages of paper, generating a massive filing and retrieval problem. In addition, only one patent examiner at a time can use the documents. An automation effort has therefore been underway for some time.

The system architecture is a NAS 9080 mainframe with magnetic disk storage and images on optical disks. The optical disks are mounted in a "jukebox" for rapid retrieval. The examiners' workstations are linked together by an Ethernet LAN through a PBX digital switch; they also incorporate a gateway to commercial searching systems.

The text system will contain over 1.3 million U.S. patents issued since 1970, plus 36 million pages of images from all U.S. patents since 1790. When fully loaded, the database will have 30 terabytes of data and may be the largest in the world. Today, 1 million patents since 1975 have been digitized and entered into the system, and all 1,400 examiners have been trained to use it routinely. Images from 151,000 patents are also available. Resolution of the images is 300 dots/inch; they are stored in compressed form.

Image retrieval and display works like the paper files; response is typically about 1 second/document. Each image workstation has two 19 inch screens and 300 megabytes of local storage. The examiner can make a hard copy of both documents and images on a local laser printer. They can also rotate the images or zoom in on parts of them as desired.

Ergonomics of the workstations are a concern. When 20 stations were put in a room, there was a heavy load on the power and air conditioning systems. Glare was a problem; new lighting systems had to be installed. Ergonomic furniture is being designed so that examiners will be comfortable spending their days in front of these workstations.

The images from the 36 million patent pages were captured at a cost of 17¢ each. Digitizing was done by a contractor; production began in 1987 and should be completed in the fall of 1988. About 1,300 images can be stored on a reel of tape; they are then transferred to optical disks. Quality has been excellent; out of a sample of 1,000 patents, less than 0.5% of the images were unacceptable. Most problems were caused by the poor quality of some of the source documents. The major problem with the optical disks was corruption by toner dust from the laser printers. The decompression software must be robust; reversal of only one bit will cause the image to fail to decompress.

Conversion Technology: New Opportunities and Problems

Don Kyser

Computer Microfilm Corporation (CMC)

CMC produces images as part of the data conversion process. They manage the ERIC database, microfilm data for Citibank, etc. and are the vendor producing the image database for the PTO. Besides image conversion, they produce microfilm, COM, and raster images. Few conversion systems are available today, but there is a huge market potential, and more players will be entering the marketplace.

We need digital imagery because the clerical workforce is inundated with paper. The imaging business is marked by slow growth, few standards, and expensive systems. The legality of image databases has not been determined; so far, people are reluctant to destroy backup paper files.

The PTO job is a major one for CMC; it will produce 40 million pages at 300 dots/inch. Document control and workflow is the most important part of the process. Raster images produce about 8.4 million pixels/page, so the PTO database will total over 1 megabyte/page. Even compressed at a 17:1 ratio, each page occupies about 60,000 bytes.

The conversion process involves PC's to generate document headers and optical scanners to transfer the documents to tape. The images are printed on laser printers to check their quality; the printers run 24 hours a day 5 days a week and produce over 1½ million images a month.

Explosive growth is on the horizon for imaging companies because of the pent up need for data conversion. Prices are going down, and new applications are appearing. Large companies will soon be entering the marketplace; the field is ripe for a leader.

PC Information Management Software

Harold Kinne

I-Track Corporation

Recent advances in hardware are now beginning to affect software development. It is hard to realize that only 11 years have passed since the first PC became available, and only 10 years since floppy disks were introduced. Other milestones of the past were Visicalc (9 years ago) and hard disks (8 years ago). The IBM PC (introduced 7 years ago) changed the world; it allowed the MS-DOS operating system to become dominant and permitted the free exchange of software. Now with the 80386 machines, PC's are getting powerful enough to be worthwhile! But the hardware is still ahead of the software.

In the spreadsheet area, Lotus 1-2-3, long dominant, is coming under attack because it has not kept up with hardware advances. Borland's Quattro is an excellent program. A new program,

Lucid 3-D from PC Supply Group allows continuous outlining programs to be combined with a spreadsheet and is available for under \$100! Text utilities will handle the large amounts of text associated with optical disk databases; by 1990, many PC's will be sold with a built-in CD-ROM slot.

Personal information managers are becoming popular; they help the user improve productivity. Some of these are:

- Inside Track builds a personal database which can be searched with Boolean logic. It sells for \$99.
- MemoryMate (Borland) allows the user to build an unstructured database, assigns keywords, and makes an index for the user to search.
- Act! (Conductor Software) has pop-up menus, keeps a record of incoming phone calls and maintains statistics on phone calls. It is useful for users who want to keep data on contacts, etc.
- Agenda (Lotus Development) builds a memory-resident unstructured database with automatic indexing on every entry. It has some AI features, including natural language input. It will sell for \$400. Although it has not been released yet, Lotus claims to have 400,000 orders for it already!

Hypertext is a growing area; more programs for Apple's Hypercard have been written in the last three months than were written for the MacIntosh in the past three years!

The Dept. of Energy's Superconductivity Database

William Buchanan

U.S. Dept. of Energy (DOE)

President Reagan has committed the U.S. to take the initiative in superconductivity research; one of DOE's responses was to create the Superconductivity Information System (SIS). SIS is available to 100 DOE sites and contractors; its goal is to keep U.S. researchers current by giving them an interactive system to use. Requirements established for SIS include operation on a short timeframe and at minimum cost. The system had to be easy to use and available to any U.S. researcher. An initial trial is being funded by DOE; access is free to users.

SIS was established on existing VAX machines at DOE and is available on a dialup basis through existing public networks (Tymnet, Telenet, etc.) It uses bulletin board software; input, help, database, and search modules were developed by DOE.

Features installed in SIS are:

- Electronic mail for private conversations between users,
- A bulletin board for upcoming events, news, meetings, comments, etc.
- A "Work in Progress" database for posting general description of research underway. (This module has hardly been used; even in this active area it has only six entries!)
- A searchable database of preprints of forthcoming papers (corresponding to the electronic version of the *Ames High T_c News* with abstracts added).
- The most recent year of the Energy Database.

Future enhancements will be to evaluate the trial, improve the data comprehensiveness, expand access to others, provide data to the non-research community, and expand the communications capabilities.

The IBM Technical Information Retrieval Center (ITIRC) supplies information and bibliographic support to IBM staff members. Among its services are 28 in-house databases, an alerting service, document ordering, and a direct connection to online retrieval vendors. ITIRC's networking strategy revolves around productivity, quality, and cost. They keep in touch with their requesters' needs by surveys, analysis of usage statistics, and SDI profile additions. It is critical to know the characteristics of their user population, who want easy, fast, and convenient access to systems. They also want to be able to talk to their peers rather than intermediaries.

ITIRC's strategy includes getting people to read the right information wherever it may be found, enhancing their productivity by providing access to the world's published information, automating support mechanisms, and maintaining a professional awareness. They provide a window on the world of information. Most professional or technical personnel will not tolerate slow system responses, they want systems to be reliable and convenient to access, they do not view the library as a first source of information but prefer to ask their peers, and they want the full copy of an item delivered to them rapidly. When presented with an automated system, they want to test its capabilities first, and they expect specific answers from it. If they use the system, they usually forget the protocols between sessions because usage is infrequent, so it is important to indicate the next step on each screen.

The largest problem faced by information providers is "you can't get there from here." ITIRC has attempted to solve this problem by using 9600 baud leased lines to heavily used services, bulk contracts, single point billing, etc. They have worked with networking people, using SNA routing, to provide high speed, interactive, full screen access.

The librarian's job has evolved into that of a network and systems expert. They provide management of information resources, electronic retrieval, cataloging, ordering, help, etc. ITIRC is changing the way IBM staff works. Surveys show that it has saved millions of dollars, time, and has had a significant influence on new products. They survey their users regularly, at least every two years, and use the results to improve their service.

Panel Discussions

Following the formal presentations, all the speakers gathered for a panel discussion. Some of the points raised were:

- Raising prices will not drive users back to print products. They may complain, but they will go along. We need to look at pricing options, such as advance payments by subscribers, to accurately forecast revenue flows. People do not understand that information has value.
- New equipment is changing economics, especially in the area of raster scanning. It may even replace offshore keying. Errors are very visible online, so 99% accuracy is often necessary.
- Database producers risk copyright violations and piracy because of the power of the new microcomputers. Technological copy protection schemes worked until hard disks became common; now methods such as insistence on a "key disk" are not tolerated by users. So copy protection today is largely an artifact of the past. Even in the business world, three or four copies are made for every copy sold. We must trust American businesses that they will not pirate software.
- Information users are price-sensitive. We need to know *what* they did with the information, not that it was cheaper to get it. We cannot sell information by counting good and bad searches—the critical factor is what the user saved by having the information. Our dilemma is that information must be priced low enough to attract new users, but not so low that the producers cannot afford to produce it.