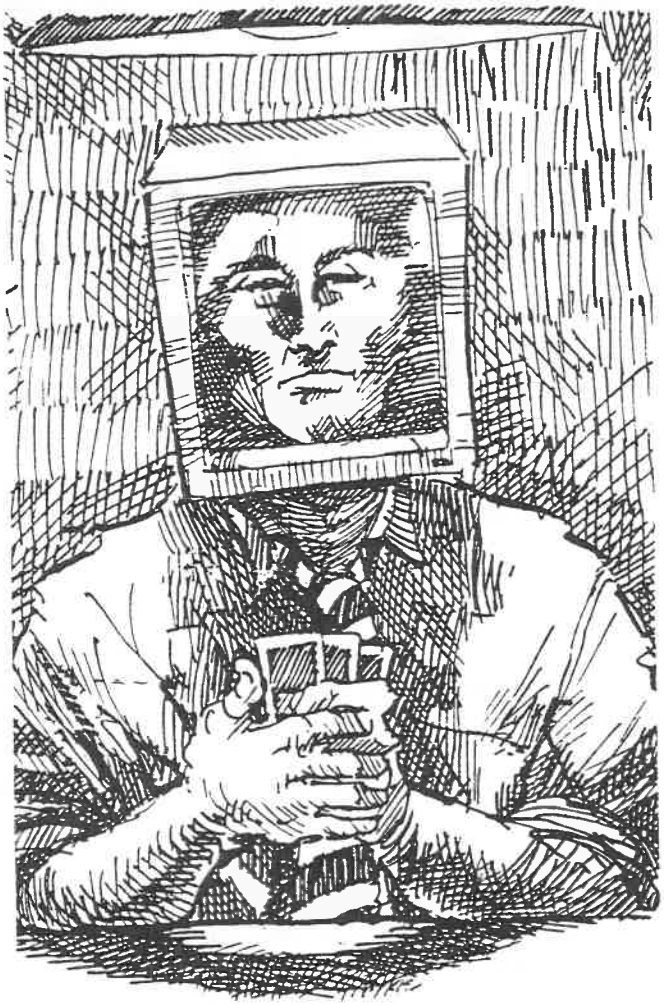
Who's holding the cards?



Reporters from top CAR programs place their bets on the "Big Three" database software packages: Foxpro, XDB and Paradox.

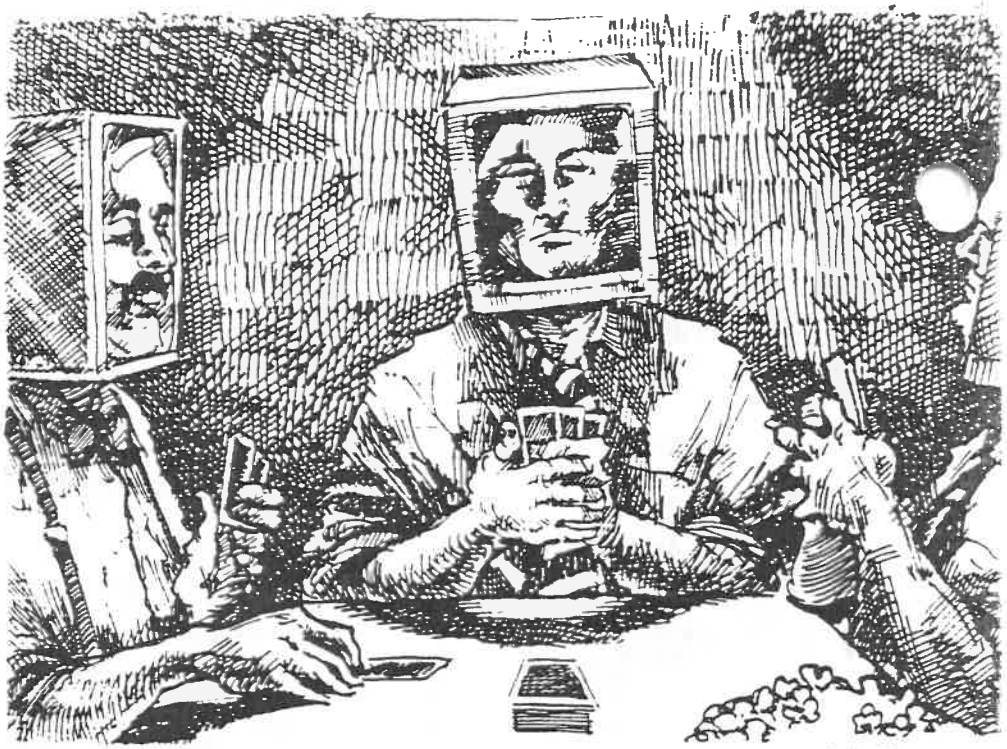
Inside

The forum for computer-assisted reporting

June 1992 Vol. 3, No. 6

FoxPro, Paradox or XDB

by Bob Jackson
MICAR



John Owens

Which should you choose?

When you're in the middle of indexing more than 250,000 records you need to be using a database software that can accomplish the task in a quick, efficient and easy manner.

More importantly, if you're sorting the same 250,000 records, you need software you can trust.

In the world of computer-assisted reporting, there are plenty of database management programs to choose from. Foxpro, XDB and Paradox are the most widely used by CAR programs. Other programs like PC SAS and DataEase are also popular in some newsrooms.

Which is the best?

That depends on who you talk to.

If you were to ask Brant Houston, who manages the computer-assisted reporting program for the Hartford (Conn.) Courant, the answer would be simple:

"You make the software you have work for you."

"A lot of it has to do with personal taste, and what works best for you," Houston said. "And part of it has to do with the task you are trying to accomplish."

Houston primarily utilizes XDB, but also depends on Paradox and PC SAS to analyze data. He is also in the process of taking a look at Foxpro's capabilities, and may make the switch from XDB in the future.

Currently, he explained, XDB's performance is getting the job done.

"We're using the new version of XDB (version 2.41) and although we are still encountering bugs, the software is very much improved over earlier versions."

Houston said that XDB is easy for reporters to learn, and is also great for smaller databases.

"But the larger the database the slower XDB runs" he explained.

Many reporters started their programs with XDB, but have since made the switch to other products because of different software problems (see April, 1992 Uplink). Reporters have cited problems indexing and sorting, but according to XDB management, those problems have been ironed out.

Indexing with the earlier versions of XDB, for instance, became a problem when the software encountered more than 31,000 unique indexes.

Although XDB has taken a bad rap because of past software problems, Elliot Jaspin, director of the Missouri Institute for Computer-Assisted Reporting, still stands by the software.

Jaspin lists speed, flexibility and ease of use as some of the product's stronger points. He also says that XDB has a simple interface and has good implementation of structured query language. Jaspin uses XDB in MICAR's quarterly seminars, which attract reporters from across the country.

George Landau, of the St. Louis Post-Dispatch, was one of the computer-assisted reporting project managers to

"You make the software you have wo for you." --Brant Houston

dump XDB.

"I started with XDB for a year and found that my tables were getting too large for the software to handle quickly, and I began having problems," Landau said. "Other bugs in the software made me lose confidence in the product and I was forced to switch to something I could trust."

Landau now uses Foxpro version 2.0 and is extremely pleased with the product, calling it "fast, full featured and bug free."

"Foxpro is very fast and I've found that it does a better job of chewing through large volumes of data," he said. "It has the analytical power of SQL but runs much faster."

Landau is also impressed with its report writing function. "It's wonderful to work with. It provides you with easy to read reports quickly."

"I've found that Foxpro can analyze greater amounts of data quicker and it has greater analytical power than XDB," Landau added. "It is the nicest, easiest software I have found."

Bob Paynter, of the Akron Beacon-Journal, switched from XDB to Paradox because of major problems encountered with the software. He made the switch, he explained, after consulting some colleagues.

"I said 'I just can't trust this (XDB)," Paynter said. "And I talked to other database managers, and nobody had ever heard of XDB."

Paradox is a menu driven program that doesn't require the user to know procedural language (such as SQL), making it easier to use. The drawback to using Paradox is that, like XDB, the program has a hard time managing large volumes of data.

Pat Stith, who manages the CAR program at the Raleigh (N.C.) News and Observer, is one of a growing number of Foxpro users."We stumbled into Foxpro and are extremely happy with it," he said.

"Foxpro is excellent because it is much faster than XDB," Stith said. "We can go through 1 million records in a very short period of time. You just don't sit around with Foxpro."

Stith explained that Foxpro is simple to use, which makes it easy to teach to reporters. For Stith that's important because he is responsible for conducting the News and Observer's two-week training seminars for reporters. The seminars, which are run periodically, have trained about 30 of the newspaper's reporters.

"I like Foxpro for three reasons," Stith said. "It's easy to understand; I believe it is the fastest database handler on the market; and I have had zero problems with accuracy."

The Boston Herald's Dave Armstrong agrees. ``Queries that would take overnight with XDB are taking a matter of minutes with FoxPro," Armstrong said. ``We will never go back to XDB." ■

FoxPro Paradox XDB FoxPro Paradox XD

Tech Tips

SUBJECT: building ranges in SQL

PROBLEM: CBS Evening News obtained a database of approximately 15,000 children who took a test to measure the lead levels in their bodies. While the database had individual scores ranging from 0 to 45, CBS wanted a report showing the number of children with lead levels of:

10 to 15
16 to 20
21 to 25
over 25

There was also the possibility that the producer would want further analysis of the children within one or several of the ranges.

SOLUTION: One approach would be to do a count based on the values in the test field:

```
select count (*)
from leadtest
where test >= 10
and test < 16
```

This approach requires running the same query, with the test ranges modified, again and again for any further analysis. Also, there would be several pages of reports to sift through. Not a major problem, but not a very elegant solution either.

A better approach is to create a new field called "range." This is a character field 8 bytes long. The next step is to run the following update command:

```
update lead test
 set range = "10 to 15"
where test >= 10
and test < 16
```

After this command is run four times for each range, the following query will produce a one page report with all the necessary information.

```
select range, count (range)
 from leadtest
 group by 1
```

Any subsequent analysis would use the range field to group the children presumably saving a substantial amount of time.

If the table were larger, space could be saved by making the range field a one byte code and using a second table called "Ranges" with a code field and the text such as "10 to 15." A join of the two tables would then produce the same results.

What does it take to run a successful computer-assisted reporting program?

The answers vary widely. But as more and more newsrooms across the nation begin to use computers and data analysis on a daily basis, reporters and editors are looking for the key ingredients that separate the successes from the failures.

For some newspapers, providing reporters with the ability to read data off government computer tapes is an investment that gives a quick return.

For others, however, the pay-off is not so swift. Starting a computer-assisted program can be a long, complicated, expensive process that offers limited results.

Uplink interviewed principal reporters in several computer-assisted programs to find out why some are flying high and others are having trouble getting off the ground.

By
Trebor
Banstetter
and
Lou Grieco
MICAR

Still on the ground

Journalists who have experimented with computer-sisted reporting with few results give a number of reasons why they have had trouble starting a successful program: unenthusiastic reporters and editors; rarely used equipment that makes it is hard to find help when problems occur; and the biggest issue, time. It is often difficult to balance other reporting duties with training and working on computer tapes.

"It's working out, but very, very slowly," says Dave Davis, who runs the program at the Cleveland Plain Dealer. Most editors, he says " go crazy if you suggest they give you someone for two days."

"There're so many stories that have to be done for the next day's paper," says Rae Davidson, who runs the program at the Toledo Blade. Davidson, who received training two years ago, says the Blade's program is "definitely in its infancy." Only four stories have come out of the program. All of the reporters whom she originally trained have since left the newspaper, though she is currently working with another reporter.

Steven Bennish, a reporter for the Columbia (Mo.) Daily Tribune, trained at the MICAR two years ago. He says that even with the full support of the managing editor, he has produced no stories with what he learned.

"I don't think there's really been the time to do it," Bennish says. The Tribune employs eight full-time reporters.

The amount of money invested varies at the different newspapers. For the Tribune, the investment was a few hundred dollars for Bennish's training, because the paper hoped to use MICAR's equipment at the University of Missouri-Columbia. Davidson says the Blade paid $10,000 for equipment, and about $500 for XDB.

The Roanoake (Va.) Times and World News paid $10,000 for its equipment. However, leadership became an issue when the person in charge of the program resigned only weeks after attending a MICAR conference.

Copy editor Jim Ellison took over, mainly because he had a PC at home and knew a little about computers. He considers the program "halfway there," though it hasn't produced any stories. Ellison says he has "had to work in between writing headlines.

"My biggest problem is I forget this stuff as I go along," Ellison says, adding that the only way to become proficient with computer equipment is to keep working on it.

Ellison hasn't been trained in NineTrack Express, so sorting out problems with tapes takes him longer as he tries to reinvent the wheel.

This problem also affects the program at the Plain Dealer. Davis says the newspaper invested $80,000 in equipment and has done some stories with it. But the Plain Dealer bought the UNISYS system, which Davis considers a mistake, because few programs use the system.

"It's a problem if you're out of the loop," he says. "Then you're always on new ground," when you need help with your system. "You spend a lot of time and energy doing things people have already done."

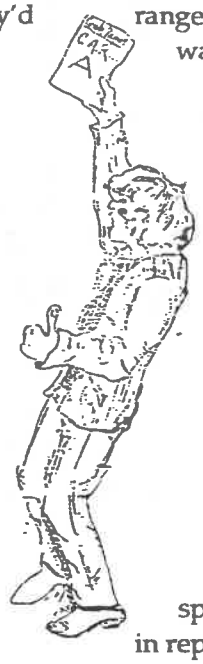
This adds to the problem of limited work time that reporters have at the Plain Dealer. Davis says that about a dozen of them can load tapes and do some basic queries. However, the newspaper is far from management's goal have the program "touc everybody's beat and everybody's job." ■

Flying High

Some C A R programs and their reporters have moved into the new information age with relative ease. What have been the keys to these successful transitions?

"Persistence," said St. Louis Post Dispatch reporter George Landau. "You run into so many problems working with government data. You need patience and determination to solve them."

Landau said the Post's program began when "they'd had some stories that could have benefitted from data analysis." A senior editor, Richard Weil, initiated the program in 1989. "I'd been working GA for a couple of years, and was pretty well versed in computers, so I was tapped for it."



Greg Reeves, reporter for the Kansas City Star who did the data analysis for the Pulitzer winning "Fialing The Grade" series, said that helping other reporters with ongoing stories has been one of the successful qualities of the Star's program. "I've started some stories, but the peripheral stories are the most important."

Brant Houston, of the Hartford Courant, agrees that helping other reporters is essential. "You've got to have a willingness to spend a lot of time with people answering questions. If you think you're the only one who should be doing this, you're going to choke your program to death."

Houston also attributes some of his success to a willingness to stick with it, despite the initial hours. "You have to realize that `full time' means anything from eight to 16 hours a day. It's as though for 10 hours you're a journalist, for four hours you're a technician, and for two hours you're a database administrator.

"There's no question that success is a two-edged sword."

Pat Stith, of the Raleigh News and Observer, credits "the newspaper's patience" with determining the success of their program.

He added that an open mind is helpful. "I learn something new every day. You never learn it all, you just get a little better as you go along.

"I've loaded 40 databases, more or less, and each one presents a different problem."

Implementing a computer-assisted reporting program is not cheap, and financial support from the publication is essential.

"We were able to wrangle about $12,000 for equipment when we started," said Landau. "The biggest investment was my salary, letting me do this full time."

Landau said that they were able to sell the idea to the Post's management as a "secret weapon" to gain a competitive edge over the now-defunct St. Louis Sun.

Most agree that to ensure success, a program should include more than one reporter. In-house training of reporters is essential to computer-assisted reporting. "This should be a routine tool in the newsroom used by a wide range of reporters," Stith said. "This is exactly where we want to go."

Stith has gone a long way to train his colleagues at the News and Observer and familiarize them with the technology, holding classes for reporters, editors and news researchers. "The class was successful in destroying the mystery of the computer. That's progress. You can't teach people computer-assisted reporting in 20 hours. It requires a lot of practice."

Stith said that about 15 reporters have produced one or more computer-assisted stories for the News and Observer in the last year. However, he said, "We don't have a situation where those reporters can do all their own work independent of anyone else."

It is important that computers be used for broad spectrum of stories, Stith said. "We've used databases in reporting business stories and features," he said. "If we can get a database, load it in and crank out the data, why not use it for features?"

The Courant's Houston said being able to "crank" out the data is vital to a program. "The challenge now is to continue to integrate it into the newsroom without creating a lot of bottlenecks," Houston said. "We have to figure out how to make it more accessible to people on the fly."

Although Landau said he has been doing most of the Post's data analysis by himself, "I would like every reporter to be able to do their own analysis when they need to. They would need training and periodic refreshers."

Houston said it is vital to have more than one person possessing the skills. "You need a couple of people who are up on things," he said. "Otherwise, if you're sick or go on vacation, everything comes to a grinding halt."■

CORRECTION:
Last month, Uplink reported on the Kansas City Star's investigation of the USDA which won the Pulitzer. However we incorrectly reported that the same series won this year's IRE award. We were also in error in stating that reporter Greg Reeves used the software Ninetrack Express to download the USDA's records into his mainframe processor. Finally, the 'dirty grain' series which helped spark the Star's main investigation was written by then-reporter Mark Zieman and not "Failing the Grade" reporters Mike McGraw and Jeff Taylor. Uplink regrets the errors.

Putting a price tag on data

By Diana Borsi
MICAR

When you're up to your ears in alligators, it's hard to remember that the original idea was to drain the swamp.

That's probably what the head of Missouri's Department of Revenue thought when a proposed bill intended to bring in money instead drew the wrath of the press and questions on access to information.

Missouri Senate Bill No. 597 started years ago as a way to raise money by charging for driver's records. The most recent draft of the proposed bill divided public information users into five separate categories with "amounts charged by the department of revenue for providing copies of driver records ... based on categories of customers..."

Bill 597 ended, killed by the lack of passage by the Missouri Legislature, but not before it spawned a dialogue between the Missouri Press Association and the Department of Revenue on access to information. The debate also raised the issue of the wisdom of allowing states to use access to information as a revenue source. It also should send up a warning flag to the press in other states.

> "It's a fact that the state government never has enough money to do what they want to do and so people start casting about for ways to make money."
> -- Sandra Davidson Scott

"It was never intended to narrow the scope of information," said Ray Wagner, director of Missouri's Department of Revenue.

"Senate Bill 597 was proposed years ago to maintain funding for the Department of Revenue. We didn't think it would have any impact on the media for two reasons," he said. First, in the final version of the bill, the media's rates were no higher than the current fees. Second, the department can waive the fees under certain circumstances and often does in the case of media requests, Wagner said.

The Missouri Press Association took another view.

"We don't think the state should be selling public information for a profit," said Doug Crews, executive director of the MPA.

An ad hoc MPA public documents group met with Wagner and hammered out a four-point letter of recommendation. It included the MPA's position against the state selling information and recommended that the Department of Revenue charge all public information requestors the same fees which should comply with Missouri's open access statute. It also should "not exceed the actual cost of document search and duplication"; publish a list of fees for document search and duplication; and develop as soon as possible of on-line access to public records to "provide for citizens the easiest, widest and least expensive access to the Department's data."

Instead of a fight, MPA got support.

"I fully understand their concerns and I don't disagree with their concerns," Wagner said. According to him, the bill won't be back. Instead the Department will seek funds from other sources, make adjustments in personnel and reallocate resources, Wagner said.

But the issue of money and information remains.

"This is an example of a nationwide phenomenon," said Dr. Sandra Davidson Scott, an assistant professor of communication law at the University of Missouri School of Journalism. "It's a fact that the state government never has enough money to do what they want to do and so people start casting about for ways to make money." And sometimes that means state agencies trying to cash in on the often lucrative information brokering business, a move Scott opposes. "I don't want to see records accessibility as a money making activity." she said, which would create a potential for information restriction. "You start getting into a lot of problems," Scott said, "especially if you start dividing citizens into categories."

While the beast of information for money is dead in Missouri, the national economic crunch could bring it to life in other states. ■

MICAR
is interested in any information, ideas or stories related to computer-assisted reporting for future issues of UPLINK.
If you wish to contribute, please mail stories or ideas to:
120 Neff Hall
University of Missouri
Columbia, MO 65211
(314) 882-0684