Uplink

THE FORUM FOR COMPUTER-ASSISTED REPORTING

October 1992
Volume 3, Number 10

Quick hits!

Writing a computer-assisted story in less than a week

Van Denton, a beat reporter for the Raleigh (N.C.) News & Observer, was doing a routine assignment. He didn't need to look for a faster way. But he did, and he ended up with much more than a routine story. All it took was a newly formed database and a floppy disk.

Every year, Denton's paper does a story on the amount of money lobbyists spend wining and dining legislators. The information was only available on hard copy, which required searching through voluminous documentation, making calculations and coming out with a story in a couple of weeks.

Denton started early, looked around a little and found person in the secretary of state's office who was building a database with all the needed information. A couple of weeks later, when all the data was in, he took a floppy disk down to the secretary and had the file copied, at no cost to the newspaper.

His story was ready two days later and included not only how much lobbyists spent, in total and per legislator, but also was broken down by interest group, such as medicine or insurance. He got a better, more complex story, in less time.

There are hundreds of examples of reporters who have done in-depth investigations over months, or even years, using computer-assisted reporting. However, reporters all over the country, from North Carolina to Washington state, are turning over stories in a week or less with the help of a database and some creative thinking. These reporters are putting out faster, harder-hitting stories long before their competition.

Here's a simple rule to follow: If you're going through volumes of documents, ask if there's a tape somewhere. Then find someone who knows what to do with it.

Pat Stith has been The News & Observer's computer expert for the past two years. He said the paper put out 31 articles in the past year using computer-assisted reporting. Also, a 20-hour training program was set up to remove the mystique for beat reporters.

Denton took the course, which he said made him aware of possibilities of this technique. On the lobby story, he was faced with 866 reports to be analyzed over the course of a couple of weeks. He looked for another way. Then he found out about the database at the secretary of state's office.

On a Wednesday, Denton got a copy of the database and brought it back to Stith in the newsroom. Stith said the "little bitty database," a flat file, took "five seconds" to load into Foxpro. Reports were ready for Denton to take home by the end of the day.

Denton had requested reports on many possible combinations of the information. When he went through the reports that Wednesday night, he tagged all the special interest groups. The next morning, Stith created a new field for groups, such as oil, insurance and medicine, and printed more reports calculating amount spent by each group.

Denton did follow-up interviews and had the story ready on Friday. But it ran on Sunday's front page, because, after all, they were weeks ahead of their competition.

"I'm more interested in something I can use every day," said Denton. "It's just a tool to get a story like any other tool."

Penny Loeb of New York Newsday made it through 500 pages of a 1,000-page document before she thought, "This has got to be on a tape somewhere." Sure enough, it was.

She and Kevin Flynn were investigating allegations that attorneys were making contributions to Attorney General Robert Abrams' campaign for the U.S. Senate while they had business pending with the his office. The reporters discovered that Abrams had received more than $120,000 for his campaign from attorneys who do business with his office, including

Here's a simple rule to follow: If you're going through volumes of documents, ask if there's a tape somewhere. Then find someone who knows what to do with it.

$33,000 from lawyers who had matters pending before him at the time of the contributions.

After six hours of searching through documents, Loeb looked for a tape. She got a tape from the attorney general's office, but a record layout, telling her how to load it, was not included. She counted some fields and made her own record layout. She loaded the tape in about two hours.

With the tape, Loeb could immediately go into the database and see names of attorneys who had pending business.

Loeb printed two sets of reports, one from the attorney general's list of pending business and one from the Federal Election Committee's list of campaign contributions. Since neither list identified law firms, she had to do a little creative thinking.

Loeb asked herself what distinguishing characteristics are in law firms' names. Of course – ampersands! She printed reports of all occupations with ampersands in the company name. Then she went through both lists by hand, matching up the names.

Loeb loaded the tape on Friday, Aug. 28. By the following day, she had identified 12 instances of lawyers making contributions while they had matters pending. However, she wasn't sure she had the right firms.

On Tuesday, Sept. 1, she and Flynn got on the phones to the law firms to interview them and confirm the data. Loeb said this took three days and was the most time-consuming part of the story.

By Friday, Sept. 4, the writing was done. The story was scheduled for Sunday the 6th, but ran on Tuesday the 8th. All together, Loeb and Flynn churned out the hard-hitting story in eight days.

Another story that took about a week came from Brant Houston at The Hartford (Conn.) Courant. A breaking story on a possible serial murderer had been unfolding over the course of a couple weeks. Houston built his own database in XDB, using clips on unsolved murders. He included dates, descriptions of the female victims, time and place the body was found, demographic information and race.

This took a few hours.

Next, Houston went to the medical examiner's office and made a separate database, to which he added significantly more information. After a couple of days, he had a clean printout with a couple of hundred records.

It took about an hour of database searches to see a pattern.

After two to three days of interviewing and updating information, the first story came out. Houston had established a possible pattern and enough connections that the state created an investigative task force. Additional stories were published and Houston's stories acted as a "tipster" for law enforcement officials.

"Once you have a database in-house, you seldom use it just once," Houston said.

Bob Wodnik of The Everett (Wash.) Herald agreed. His XDB database on building permit records had been used for numerous stories. Wodnik, Scott North and Jim Muhlst[illegible] did a story in three days on increased development which led to more stormwater runoff into an already flood-prone area.

With this story, there was no problem of building, finding or copying a database. It was there, and the reporters had become used to it, which made the quick story even easier.

George Landau, at the St. Louis Post-Dispatch, said his paper had done quick stories with in-house databases, too. Three examples came from computerized Missouri death certificates:

- Deaths from breast cancer were occurring at a higher rate (in certain age ranges) in the city of St. Louis than in the county or in the state as a whole.
- No decline was shown in the number of drowning deaths among children for the last 12 years in the St. Louis area or the state as a whole.
- The rate of asthma deaths increased in Missouri over the last 12 years.

Stith said it's a realistic goal for the News & Observer to run a computer-assisted story every week.

"You could sit there in one database and do ten stories," he said. But, he added, the objective is to find one good story and get it in the paper.

Alison Boggs is a graduate student at the University of Missouri School of Journalism.

When TRI is not enough, try state hazardous waste manifests

When it comes to environmental issues, particularly the treatment and disposal of hazardous material, Uplink often touts the EPA's Toxic Release Inventory. TRI is a database with untold uses, but it contains generalized data and provides little detail with regard to the fate of hazardous materials.

TRI records scarcely go beyond "Off-site Transfer" when describing the fate of the millions of tons of hazardous materials that were stored, treated and/or shipped offsite by the thousands of large-quantity waste generators in the United States.

TRI serves as excellent summary material and a good initial source for information concerning manufacturers and their toxic chemical discharges. The challenge for computer-assisted reporters is to find new databases on local or state levels and use them to develop more detailed applications in the environmental arena to supplement the federal perspective. At Electronic Public Information Consultants (EPIC), we tap into data that helps make this possible: state hazardous waste manifests.

Hazardous waste manifests are one of the many consequences of the 1976 Resource Conservation and Recovery Act. In theory, RCRA is supposed to help environmental enforcement agencies track every ounce of hazardous material "from cradle to grave." In short, the act compels manufacturers that create and ship hazardous materials to complete Uniform Hazardous Waste Manifest forms every time waste leaves their facilities.

Despite the name, there is no such thing as a standardized UHWM form. Some manufacturers actually generate their own forms. However, RCRA does specify reporting requirements. The manifest must not only describe hazardous material in superb detail, but also must specify who is transporting the material, where it is going (even if out of state), when the material is shipped, and just about anything else you would want to know about the material's transport, delivery and disposal.

Using hazardous waste manifests, reporters can track the transport of hazardous materials through their state. In all likelihood, they can also identify the source and destination of the most menacing material, cross-reference the data with TRI and state enforcement data to determine if discharge figures are consistent, and even identify illegal transport activities. Needless to say, reporters can thereby add a tremendous amount of detail to TRI's "Off-site Transfer" records.

The good news for computer-assisted reporters is that about half the states in the United States are using computers to compile UHWMs. They are: Alaska, Arkansas, California, Connecticut, Delaware, Illinois, Indiana, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, Nevada, New Hampshire, New York, New Jersey, Oregon, Rhode Island, Texas, Vermont, Washington and Wisconsin. As of last year, Alabama, West Virginia, and Maine were planning to automate.

The other states, as you might imagine, are complaining about volume problems and their relative inability to handle the information. Ohio, for example, supplements its current system (or lack thereof) by purchasing manifest databases from bordering states – Pennsylvania, Michigan, and Indiana – all of which have automated systems. You can take a similar approach if your state is not using computers. At this point, it is difficult to be specific about fees for data. EPIC paid $86 for Texas manifests, Illinois manifests are free of charge, and New Jersey charges $4.50 per second of CPU time with a $500 maximum fee.

Requesting data should not be difficult. A data processor at the Department of Natural Resources in New Jersey boasted that he fills hundreds of requests for manifest databases each month (don't worry, he couldn't recall doing much work for journalists). Whether he was exaggerating or not, he did make it clear that obtaining data was no problem.

State agencies store manifests anywhere from three years to indefinitely. Unfortunately, every state will have a different agency responsible for compiling the manifest data, and each will have different policies concerning format and maintenance.

One way to get started would be to call the agency in your state responsible for enforcing environmental permits (e.g. Department of Natural Resources, state Environmental Protection Agency, or Department of Ecology). If that fails, call your state Department of Transportation. EPIC has started to compile as many of these hazardous waste databases as possible. Feel free to call us.

For more information regarding hazardous waste manifests, there is an excellent, free resource manual called "A State Guide to the Uniform Hazardous Waste Manifest" published by the National Governors Association (444 North Capitol Street, Washington, D.C., 20001-1572). In addition to describing the manifest process in detail, the guide identifies appropriate agencies and contacts in every state.

Adam Berliant is a founder of EPIC, a computer-assisted research firm in Seattle for journalists and other professionals. Berliant can be reached at (206) 448-0454.

The manifest not only describes hazardous material in superb detail but also specifies who is transporting the material, where it is going, when the material is shipped and just about anything else you would want to know about the material's transport, delivery and disposal.

The National Library on Money & Politics searches databases for y

Think about what the civics and history books tell young citizens about democracy and voting, how the members of Congress represent The People in their district, about the Power of One Person in the Voting Booth.

And now, think of this, from the Real World. Every member of Congress has two constituencies. The first has the most people — the citizens of that district: the few who voted for the winner, the few who voted against and the majority who didn't vote at all.

The other constituency, the one not in the schoolbooks, has the most power. The congressperson knows every member. They gave the money for the expensive campaign. And the high cost of modern campaigning means that the congressperson spends a great deal of time with this limited group of people who have the money and motivation — business, political, ideological — to give to campaigns.

There are lots of stories in tracing what money does to politics. Most of the information is on record. It affects just about every aspect of the lives of the people who read your paper or listen to your broadcasts. Since you're a reporter, you should be figuring out what the data means and then telling the stories to your readers and viewers.

So what does it take to get started? You can go to the Federal Election Commission with your FOIA request, pay the fees, load the tapes, convert the data to a form your software can import and start figuring out how to write the queries.

But the National Library on Money & Politics has already done this. You can get more information, in less time, with less trouble. And in many cases it will cost your news organization less than the mileage to the state high school football championship.

The Library gives you information that's not in the FEC records — a coding system that identifies the political/ideological/business interest of the Political Action Committees. The FEC has the PAC's full official name and the sponsor, but many PAC names give you no clue about what their interest is.

The Library already bought all the tapes, with all the data, all the way back to 1979. It can give you information that covers just the states, districts, officials and years you're interested in. You can get the data for just one congressperson now, and after those stories are done, decide to order all the data about the two senators in your state.

The FEC's job is to maintain the records and make them available to the public. The Library's job is help reporters figure out what data they need and get it to them quickly. If reporters need help deciding how to approach the data, or finding the stories in the data, the Library staff can help. They know campaign finance and they've been reporters, too.

Some common requests for data: Send the tables of all the PAC and individual contributions to a congressperson over a certain period. How much money did Senator Z get from energy interests? From the transportation industry? Some questions can be answered with just a fax or a return phone call: How much money have cable industry PACs given since 1980?

You may want to start out with everything for your state since 1979. That would include all the individual contributors in your state who gave to PACs, political parties and presidential candidates; and all the contributions from PACs and individuals to Congressional candidates in your state.

The data is put on floppies in the format you choose: CSV format, a DBF file for Foxpro and dBase, or a Paradox file. We can send it to you Federal Express or, even better and faster, to your Compuserve E-mail.

The National Library on Money & Politics is affiliated with the Center for Responsive Politics. The Library is a non-profit, nonpartisan public education service. The offices are at 1320 19th St. NW, Washington, DC 20036. The Library's phone number is (202) 857-0318.

Common requests for data:

- **Tables of all the PAC and individual contributions to a congressperson over a certain period**
- **How much money did Senator Z get from energy interests? From transportation?**
- **How much money have cable industry PACs given since 1980?**

Richard Mullins, researcher and syste administrator for the Library, is curren working with Elliot Jaspin and Bill Dedman on *Power Reporting*, a guide to computer-assisted journalism.

Tech Tips:

FoxPro adds a new dimension to SQL

FoxPro's Point and Shoot SQL Interface can smooth the way for beginners.

The great thing about Structured Query Language is that it is simple, powerful and for the most part, easy. Beginners can get started quickly, and pros can go as fast as they need to. It's no surprise that SQL has been adopted by reporters enthusiastically.

For the first stage of a program of computer-assisted reporting, SQL releases the power of the highly-motivated individuals who are driving the effort. But after a small cadre of reporters has hit a few home runs, knowledge must spread throughout the newsroom in order to avoid a bottleneck. That's why training eventually becomes one of the top, if not the top priority to ensure the success of a program of computer-assisted reporting.

For those news organizations using FoxPro Version 2.0, the Relational Query By Example interface provides a tremendous tool for teaching SQL to beginners. No matter how easy SQL may seem to the initiated, beginners need, and should have, as much help as possible.

What the Relational Query By Example interface does is provide a point and shoot menu interface for SQL. For each portion of the SQL statement, the user chooses from a menu instead of typing the command. For the seasoned SQL programmer, this may seem awkward at first. I know it did for me, but now I write almost all of my SQL statements using RQBE.

The benefit for beginners is that they don't have to sit there facing a SELECT statement, wondering what to type next. With the RQBE interface, they can play around and use menus to build the statement they want as follows.

To enter the RQBE interface, a FoxPro user creates a new query file by choosing **New** from the file menu and specifying the file type as **Query**. This displays the interface screen depicted in the diagram.

When FoxPro creates the query, it assumes that whichever database is open will be the subject of the query. If this isn't the correct database, then it can be cleared from the screen, and another database can be added from a menu of all available databases. If a database named "DB1" were open, then that database would be used for the query and its fields would be displayed in the **Select Fields** window.

The user can change which fields appear in the query, or to choose various functions, such as **SUM**, **AVG**, **COUNT**, etc.

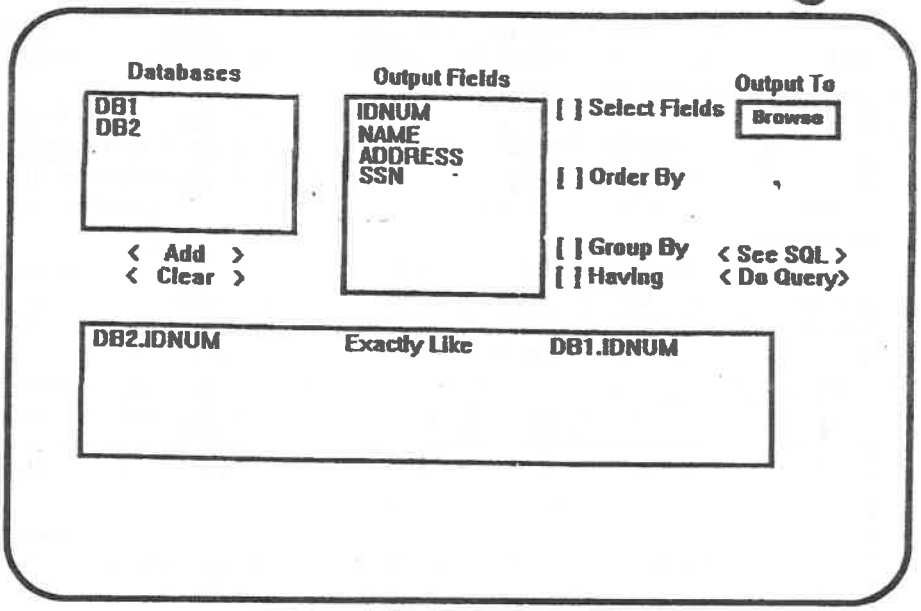
When a user adds a second database to the query, say "DB2" FoxPro pops up a menu-driven interface to the **WHERE** clause. Instead of typing in **WHERE DB2.IDNUM = DB1.IDNUM**, the **WHERE** interface allows a user to point and shoot. First a menu of all the fields in the DB2 database is presented so the user can select the appropriate field to join the databases together. Then the user selects the comparison operator, (**LIKE** or =, **EXACTLY LIKE** or ==, etc.), and finally a menu of fields from the DB1 database appears.

After the **WHERE** menu has been filled out, the where clause will appear in the large box at the bottom left of the screen. Also, all of the fields in the DB2 database will appear when the **Select Fields** menu is chosen. The same style interface is available for the **ORDER BY**, **GROUP BY** and **HAVING** clauses.

The menus are nice for queries that use several databases, because the job of placing the alias tags that identify which field belongs to which database is handled automatically.

When I first started to use this interface, I balked because I thought that it was a replacement for SQL, not a way to learn it. As soon as I clicked on the **See SQL** box, I realized I was wrong. Clicking on this box displays the SQL query built by the interface. A beginner can point and shoot for a while, then look at the SQL that he or she has created.

Once the query has been completed, then it can be executed. The output can be directed to a temporary database for browsing, a permanent database for later use, or a report for printing.



The mouse-driven interface of FoxPro's Point and Shoot makes using SQL easy, even for beginners.

Dan Woods is a reporter at The Raleigh, (N.C.) News & Observer.

Bits, bytes and nibbles

■ An electron... forum on computers in journalism debuts

The University of Louisville in Kentucky has set up an online forum for working journalists and journalism educators, news librarians and researchers.

Computer-Assisted Research and Reporting List, or CARR-L, as this corner of cyberspace is called, focuses on the use of computers in journalism.

Topics range from text processing and graphics to online database searching, computer communications and investigative reporting, according to Elliot Parker, one of CARR-L's curators.

The nature of the exchange depends on the users, but here are a few topics which have passed through CARR-L recently: Where do we get state-level information in electronic form and how do we know what is available?; is there a net source for backgrounding information on hurricanes?; software reviews, job announcements.

For more information on CARR-L, contact Elliot Parker, Department of Journalism, Central Michigan University, Mount Pleasant, Mich., 48859, (517) 774-3196.

• • • • •

■ More errors surface in HMDA data

More problems have surfaced in the 1990 Home Mortgage Disclosure Act (HMDA) data.

Last month Anchorage Daily News reporter Bruce Melzer told Uplink that the Feds used the state abbreviation "AK" for Arkansas as well as Alaska in the bank address file.

This didn't seem like a problem at the time since the loan applications, listed in a separate file, used a numeric code to identify states.

However, Melzer found that "02",the code for Alaska, was also used for Arizona part of the time. Of course, this means Alaska records could be hidden under some other state code.

When asked about this, Cornelius Driggins at the Federal Reserve said the problem falls within their accepted 3 percent error rate.

Unfortunately for Melzer, the error rate in Alaska runs closer to 35 percent, with 2,077 Arizona loan applications recorded under "02".

• • • • •

MISSOURI INSTITUTE FOR
COMPUTER-ASSISTED
REPORTING
120 Neff Hall
University of
Missouri
Columbia, Mo.
65211