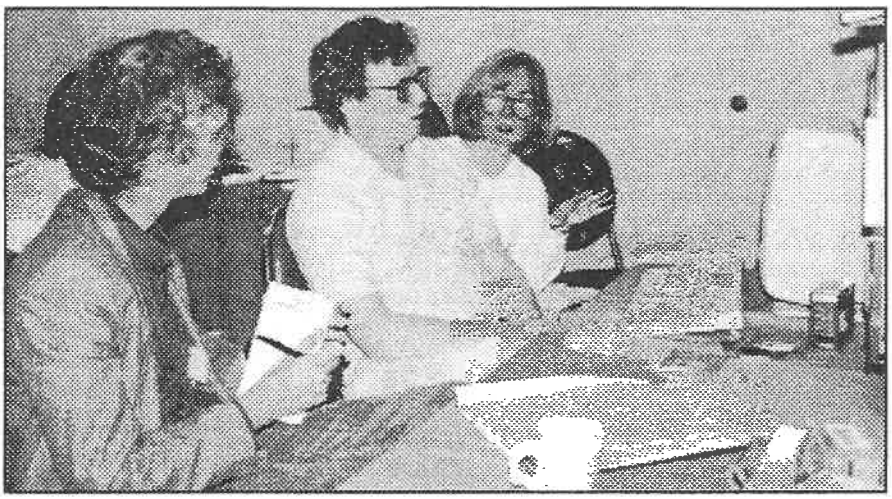
Download it . . . and they will come.



John Freed of The New York Times demonstrates mapping software to Mary Lou Gallagher of Planning Magazine and Eric Scigliano of Seattle Weekly at the Investigative Reporters and Editors National Conference in Portland.

More than 400 reporters from all corners of the United States flocked to the Investigative Reporters and Editors National Conference in Portland to get a glimpse at Journalism's new "field of dreams."

Inside

The forum for computer-assisted reporting

July 1992 Vol.3, no. 7

Politics, Power and PCs

Orange County is the home of Super-rich moderate Republicans who like to make political contributions and middle-class conservatives who can't afford to. Guess which group dominates Republican primaries in California's GOP stronghold?

Ronald Campbell
Orange County Register

If you bet on the monied moderates, you bet wrong. But so did several wealthy developers who have thrown their chips behind a succession of moderate Republicans, only to see them buried at the polls by hard-right conservatives.

The Orange County Register documented the split between the county's money and its votes through a computer analysis of nearly 150,000 campaign donations.

The database was the starting point for several pre-election analyses challenging widely held beliefs, such as the omnipotence of money in politics.

For example, we showed that local Democrats, usually portrayed as inept sideliners, in fact were powerhouse givers. Orange County raised $5 million for Democratic candidates for the legislature and Congress between 1987 and 1991. But most of this money went to candidates outside the county.

We also established that 1991 was a lousy year for fund-raising, much worse than would be expected in a normal off-year. The falloff couldn't be blamed on a lack of competition. Using a simple test for a candidates seriousness—could the candidate raise $10,000 in an off year?—we documented that political competition was measurably hotter in '91 than in '89.

And when one Republican congressman threatened to challenge another GOP incumbent for a newly created seat, we quickly showed the race would be a mismatch, at least in fund-raising. (Despite our analysis, the stronger candidate pulled out, unwilling to get into a fight with a rival who publicly said he had nothing to lose.)

Computers were made for stories like these. They are the perfect tool for a political junkie. But there are a few obstacles to watch for before diving in:

*First obstacle: The records are old. Candidates file their end-of-year reports on Jan. 31. Transcribing the reports from paper to computer tape easily can take an additional on to two months. There are only two ways around this obstacle—hire your own people to enter just the records you want or pray that the Federal Elections Commission begins requiring candidates to file reports on computer.

The Register did not get its state and federal records for 1991 installed until late April. You might be able to shorten the time by subscribing to a database service, but even a little searching on-line will quickly outweigh the cost of maintaining your own database in the newsroom.

The built-in delay means you can pretty well give up the idea of tracing, say, all donations from trash haulers to legislative candidates until months after the fact. But for spotting long-term trends, nothing beats a computer.

*Second obstacle: The records are filled with errors. Not just the deliberate obscuring of a donor's name or employment. That sort of error I can understand; in a weird sort of way, I can respect it.

But why would anyone report a donation on "2/29/90," much less several dozen donations on this non-existent date? My database program refused to load 10,000-record tables when there was as much as one misdated contribution. It took me hours to figure out what was wrong.

A second type of error is guaranteed to foul up any database analysis. I refer to inconsistent names. "Lyon, William M" and "Lyon, Gen Wm" and "Lyon, William" all refer to one of the nation's largest homebuilders, a member of the Forbes 400 whose hobbies are collecting antique cars and bankrolling Republicans. But to my computer there are three Lyons, not one.

A few steps will solve this problem. First cast a wide net for donors. If you're looking for everyone who gave more than $10,000, look beyond that set to everyone who gave more than, say, $2,500. If you want to get the 10 largest donors, get the 20 largest. This will probably turn up a few donors with inconsistent names. Next run a wildcard search for every last name on your list of top donors; beware of "Mr., "Mrs." and middle initials. Finally, select a standard name for each donor and use the "update" command to change non-conforming records.

*Third and worst obstacle: Lack of imagination. The day you get campaign contributions loaded on the newsroom PC, the political reporters at your paper will clamor for the names of top donors and for total donations by city or by date. After that, ideas come few and far between.

Breaking out of the obvious-question mold is hard. One possibility is to collect conventional wisdom from political writers, then figure out how to prove or disprove it. The conventional wisdom was that Orange County, a stronghold of conservative voters, spent its money the same way; the computer showed **otherwise.**■

Bytes, Bits and Nibbles

Edited by
David Washburn

"Sin" study wins award

Greensboro, N.C. News & Record staff writer Andrew Barron received a Wilber Award in April from the Religious Public Relations Council for "Sin: The Triad has a lot to say about it," the News & Record's first computer-assisted reporting project. (Published in the November, 1991 Uplink)

The package, published Oct. 13, 1991, examined the views of News & Record readers on what acts constitute sin, the relative importance of the Ten Commandments and the seven deadly sins, and other topics.

More than 1,100 readers returned the published survey, database reporter, Lex Alexander used the newsroom computer and XDB to tabulate the results, cross-referencing them by major religious denominations in the Greensboro area. He also and compared respondents' demographics with Census Summary Tape File 1A data for the coverage area to see how they differed from the population at large.

Although the tabulated responses were interesting, Barron's thoughtful interviews with clergy and lay people alike (including an entire Sunday school class) made the package worth reading. Way to go guys!

Making claims in Missouri

George Landau of the St. Louis Post-Dispatch combined computerized campaign contributions and computer records of the Second Injury Fund, an obscure workers compensation fund, to show that appointees and political supporters of Missouri's Attorney General (and candidate for governor) William L. Webster win up to three times as much in settlements and fees from the fund than attorneys without those connections. The total value of excess settlements was above average at $12 million over 3.5 years.

The Post- Dispatch separated lawyers who file claims against the state's Second Injury Fund into three groups, each with its own degree of success in winning money from the fund. In almost all cases, the lawyers kept 25 percent of each settlement as their fee.

Average settlements paid from the fund:
****Lawyers with no political connections— $3,548**
****Lawyers who gave to Webster's campaign— $5,307**
****Lawyers Webster has hired to defend the fund $11,808.**

Buyer Beware

As journalism enters a new information age, reporters who use computers to assist them in reporting must not only negotiate the price of a reporters' notebook with a Wal-Mart clerk, but also learn to negotiate the price of a tape drive with a computer hardware salesperson. The negotiation process, not to mention the price, is quite different.

Adam Berliant, co-owner of Electronic Public Information Consultants, a company geared towards obtaining, analyzing and reporting on electronically stored government records, will be the first to say that government bureaucrats are not the only obstacles to a database reporter.

In October of 1991 E.P.I.C. founders Berliant, Tom Braden and David Hinchman began operations in Seattle. A top priority was to buy a tape drive. The fledgling company wanted a Cipher drive, known for its high quality and reliability. However, a new Cipher drive can cost up to $14,000. E.P.I.C. could not afford a new Cipher, so they decided to buy one used for $6,000 from Contech Computer Corporation of Wilmington, N.C..

Berliant negotiated the financing, which included a $2,000 down payment. He says he was told by Contech's president Robert Conti that E.P.I.C. would be spinning tapes within a week after ordering the drive. But after a month of waiting they were still without a drive. Berliant was tired of waiting and decided to cancel the order.

A few weeks later he received just over $600, less than a third of his down payment. He was told by Conti that all of Contech's written purchase agreements include a restocking charge if an order is canceled. The charge, which covers the reconditioning of a drive when the customer sends it back to the warehouse, amounts to 20% of the purchase price.

Berliant contends that E.P.I.C. should not have to pay a restocking charge because they never received the drive and had only paid 30% of the $6,000 purchase price. He has contacted the FBI and the Better Business Bureau but has not received his money back. ■

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:
120Neff Hall
University of Missouri
Columbia, Mo. 65211
(314) 882-0684

The Greensboro News &Record has North Carolina dairies crying over spilt milk.

Lex Alexander
Greensboro
News & Record

Federal and state investigators believe that between 1982 and 1989, North Carolina dairies rigged bids for school milk contracts and got as much as $8 million of illicit revenue, 2 cents at a time, from the state's school children.

The News & Record first considered looking into the bidding in July 1991, just after we got our newsroom computer and about two months after National Public Radio reported that a North Carolina-based dairy had been implicated in school milk bid-rigging in Georgia.

At first, we had very little to go on. We assumed that the U.S. Department of Agriculture or the state Department of Public Instruction would have bidding information for school milk contracts in a computer file in one central location. They didn't. The records were kept on paper at 133 separate school system offices.

But in August, after another dairy admitted having rigged bids in a school system in an adjacent county, we jumped in anyway. We requested (via form letter) information from all 133 systems, back to 1981, on which dairies had bid on a contract, what kind of contract it was (whether it kept a price firm for the entire school year or allowed the price to vary with changes in the cost of raw milk, which was set by the state dairy commission), and how much per half-pint of milk they were charging.

What we got back created a stack of paper 2 feet high. The only help our tape drive offered was to give me a place on which to set the stack.

Because schools were required to keep such records only for three years and three months, many systems had records only as far back as 1988. And, with school about to start, many child nutrition directors felt like they had more important things to do than help us.

But other systems fell all over themselves to give us everything they had. And some of them had a lot: About 40 systems had usable records all the way back to 1981.

We also got an unexpected break. Early on, one school system's child nutrition director told us that federal marshals had told her that her system definitely had not been touched by the bid-rigging. We were able to use bids in that system as a benchmark for comparing bids in other school districts; in almost every case, the differences were dramatic.

From mid-August until Christmas, I divided my time between typing in the information we had received and calling school systems to beg, plead, cajole, and just generally make enough of a pest of myself to get them to give us what we'd asked for, or fill in the blanks where information we got was incomplete. It wouldn't have been cost-effective for us to sue the holdouts, and ultimately we got at least some usable information from 112 systems.

Meanwhile, I ran up a huge bill in our library, using VuText, DataTimes and other databases to check publications from New Jersey to Texas for valuable information on school milk bid-rigging in other states.

I talked to dairy industry experts who told us what kind of bidding picture we could expect if bidding had been honest—and what we could expect if it hadn't. Then, during Christmas week, I sat down to crunch the numbers.

Comparing bid prices statewide with bid prices in the system where no bids had been rigged lent credence to investigators' claims that the dairies had received about $8 million in illicit revenue from milk sales between 1982 and 1989.

The analyses and other information showed that:

** North Carolina dairies frequently enjoyed exclusive, long-term relationships with school systems in the towns or counties where they were located.
** In any given year, prices bid by a particular dairy generally varied by 15 percent or more. One dairy's bid prices varied 60 percent. Our experts considered those swings significa... an industry in which, because of ... s perishability, profit margins seldom exceed 2 percent.

Those swings also were consistent with investigators' suspicions that dairies had bid intentionally high prices in systems where they had no intention of competing for the contract. Such bidding created the appearance of competitive bidding where there really was none.

I also traced the movement of milk prices against the dairy commission's changes in the cost of raw milk that the processors had to pay. (The commission had been disbanded in 1990 and its records destroyed, but a source in state government was able to fax me copies of commission records dating to 1981.) I found that the prices bid by dairies tended to go up more quickly, and some came down more slowly, than the cost of raw milk, the dairies's biggest expense. That meant, if all other expenses were equal, the dairies were making more money. In fact, between 1982 and 1986, their second biggest expense, the price of gasoline, was falling, meaning even more profitability.
** Comparing bid prices statewide with bid prices in the system where no bids had been rigged lent credence to investigators' claims that the dairies had received about $8 million in illicit revenue from milk sales between 1982 and 1989. This analysis suggested that the dairies got that &8 million the hard way—2 cents at a time.

The analyses did not conclusively prove bid-rig... (I... fact, despite subpoenas for dairy company travel, expense and telephone records, the feds couldn't prove bid-rigging here unti... two dairy executives secretly began cooperating with them in the

spring of 1991.) But in records from 1982 to 1989, when word of the federal investigation in other states spread to North Carolina, the analyses did show a strong similarity between the bidding patterns of the three dairies that had admitted rigging bids in North Carolina and the four other large dairies that had not. (Three of those have been implicated in other states.)

This was a highly technical subject. A spreadsheet program would have helped tremendously during the analyses and in generating graphics that would have made the stories clearer to the reader, but we don't yet have one.

There weren't a lot of real people in the package with anything meaningful to say. They only helpful source I could find in the state's dairy industry stayed strictly off the record. And it's not at all clear yet what impact the package hashad.

Since publication on March 1, seven months after our investigation began, no new indictments have been handed up or guilty pleas announced. The investigation continues at this writing.

The project gave me some practice in setting up relational databases working from paper records, and it helped demonstrate to other staffers the value of computer-assisted reporting. ■

Tech Tips Tech Tips Tech Tips

By Elliot Jaspin

Subject: Extracting information from a tape.

Problem: The Toxic Release Inventory uses what is called either a "submission number" in its 1987 documentation or a "document control number: in 1989. The number, which is 15 bytes long, occurs at position 5 in each of the different record types that make up the TRI file. The information in this field is crucial for a number of reasons.

The document control number is used to identify a specific source of pollution such as a plant, and is included in the various documents that are filed by the polluter. Using this number you can tie together all related bits of information on one site that are scattered throughout the TRI file.

The number also contains the state where the the plant is located. This information occurs as the last two bytes of the number.

Most people who use TRI are interested in information on polluters in one state or at most a small group of states. How then do you extract only those records for a specific state while keeping the all important document number intact during the transfer.

Solution: It is possible using NineTrack Express to define a field in two different ways. Thus, you can first define a field called "DocNo" that begins at position 5 for a length of 15. When you transfer information from the TRI file, you would make sure to import this field.

At the same time, you can also define the last two bytes of the document control number as the field. This field begins at position 18 for a length of 2.

Assuming that you only want those records where state equals "KY", there is no need to import the field. Astate field that only contains"KY" would only take upvaluable space on the hard disk.

However, you can still filter on this field even though you will not be importing. As each record is read from the tape, NineTrack Express will check to see if the state field contains "KY" and import the rest of the fields including the 15 byte document control number. ■

Journalists from around the country "log on" at the Portland IRE conference.

Jon Schmid Sotomayor
MICAR

Question: Why on earth would 400 reporters get up at 8 a.m.. Answer: A whole day of lectures and demonstrations at the largest conference ever on computer-assisted reporting.

Investigative Reporters and Editors sponsored the program on computer-assisted reporting as part of its national conference held in Portland, Oregon June 11-14.

But judging from the turnout, most conference attendees considered the day essential, rather than optional.

"I have been going to these conferences for 10 years and I have never seen an 8 a.m. conference with standing room only," said IRE president Mark Middlebrook.

In addition to the seminars and panels, a computer practice room staffed by volunteers gave everybody the opportunity to practice downloading data from 9-track tape. Participants tested a variety database management programs, including XDB, Paradox and Foxpro. Experienced journalists gave demonstrations of different mapping programs and anybody could try tapping into research libraries through on-line services.

The National Library on Money and Politics gave away more than 200 free diskettes with FEC information listed by state. And IRE sold, for $4, about 100 shareware diskettes each containing a sampler of 10 utilities programs.

The nine seminars and panels Thursday ranged from recognized experts discussing the direction of computer- assisted reporting to nuts- and- bolts demonstrations on how it is actually done.

Statistics sage Phil Meyer talked about the past, present and future of the craft and called for more sophisticated analysis instead of the usual "quick-hit" approach. However, Elliot Jaspin assured the audience that basic counting and sorting can still lead to great stories.

Other seminars covered more practical problems for journalists eager to start computer sleuthing, such as "Getting Started: Computer-Assisted Journalism on a Budget in Your Newsroom."

Accomplished reporters shared war stories and tips on building databases, prying tapes from unwilling government bureaucrats and mining huge databases, such as the census.

In "Deciphering the Gibberish: A Short Course in Government Databases" the audience watched the computer screen, projected overhead, as reporters George Landau, of the St. Louis Post -Dispatch, Penny Loeb of New York Newsday, and Bob Port of the St. Petersburg Times typed through the methodology used for some of their best stories.

"Computer-assisted journalism has always been pushed by reporters,'' said Bill Dedman, IRE boardmember and Pulitzer Prize winning journalist. "This (conference) means the seeds have been planted at several hundred more newspapers." ■