Uplink

THE FORUM FOR COMPUTER-ASSISTED REPORTING

November 1992
Volume 3, Number 11

Pursuing the paper trail

Washington state reporters use data tapes to reveal government negligence and misconduct

By Jerry Uhrhammer
The Morning News Tribune
Tacoma, Wash.

While covering the case of a Washington state ferry boat captain accused of sexual harassment, *News Tribune* courts reporter Cynthia Flash began hearing stories of other lawsuits pending against the state. She wondered how many.

With the help of a computer, she found out. There were thousands, many thousands, not just for sexual harassment but also for illegal rectal probes of prison inmates and for people killed and maimed because of wrongful conduct and neglect by state employees.

An analysis of computerized state records led to a three-day series that revealed for the first time that state taxpayers were paying a heavy price for the negligence and misconduct of state employees — more than $76 million in settlements in less than a decade. The annual average was $9 million.

Although ranking only 18th in population, the state was paying more per capita in liability damage awards than any other state in the contiguous United States.

The series was the first major computer-assisted project at *The* (Tacoma, Wash.) *Morning News Tribune* and, as this piece is written, its findings are reverberating in the campaign to elect a new attorney general.

Aside from that impact, its success underscored the fact that time-tested techniques of investigative reporting — specifically, following the paper trail — should be an integral part of computer-assisted reporting.

Following the paper trail and using your noggin to figure out where documentation can be found are vital. The big difference is that, in these times, the indexes to the records, if not the records themselves, are kept on computer. In this project, the computer data led the way to the paper trail.

While there is a tendency to measure computer-assisted projects by the number of megabytes in the database you're crunching, don't be afraid to think small. In Washington state, many obscure public agencies have their own little fiefdoms. Instead of storing data on a state mega-computer, they tend to keep their records on PCs. The whole database you're after may fit on a few diskettes.

Something I learned in two decade of following paper trails: Read the applicable state laws before you get too far into your investigation. Become knowledgeable about how the system is supposed to operate.

That's what I did when I was asked to do the computer work for Flash's investigation. The risk management statutes yielded a quick dividend; I found that the risk management agency was required, by law, to maintain a computerized tracking system of all claims filed against the state.

The trouble was this computerized tracking system was, by law, confidential. However, the same law held that claims filed against the state were public records, available for inspection. Our strategy: Press for release of the data that was public, even though it was on a "confidential" database.

When Flash made her first visit to the agency, she was offered a voluminous print-out containing the key information of the public claims data kept on the computer. Altogether, there were some 25,000 cases.

The printed data was so diffused, so overwhelming, that it was virtually unusable. We needed the computer data in order to make sense of anything.

I made a public records request for the computer data, specifically the portion from the hard copy claims that state law said was public record.

Nothing like that had happened before at the risk management agency. The director sent the request to the attorney general's office for a legal opinion. It took six weeks to get an

answer.

Finally we got word that our request had been approved; we would be given the computer data for all closed cases, some 20,000 of them, together with the type of case — civil rights, wrongful death, personal injury, etc. (Later, we got an additional 5,000 records of open cases.)

Using XDB, I began breaking down the claims data for each department, year by year.

It was a case of using the computer trail to help us explore the paper trail, and its importance cannot be overstated.

The information was used in an information graphic with the number of cases per department and total dollars paid in settlements. I also tracked, year by year, the number of cases of each particular type. Civil rights complaints (including sexual harassment) turned out to be the fastest-growing type of claim.

This data was inserted into a QuattroPro spreadsheet that was invaluable for figuring percentages and making other calculations.

By manipulating the data, I was able to pinpoint which specific departments had the worst records.

But the most important use of the computer data was linking the worst cases, those with the biggest settlements, to claim numbers. This enabled Flash to go back to the agency and inspect the hard copy claims.

It was a case of using the computer trail to help us explore the paper trail, and its importance cannot be overstated. The hard copy claims were a rich source of detail, including names, addresses and telephone numbers, helping us to contact and interview the victims and survivors.

The series found that while total claims against the state were declining slightly, civil rights violations were skyrocketing — and so were settlements. The state ferry system, with the worst record, finally instituted a training program for employees.

But the series also showed that some state agencies simply kept doing things the same as ever, despite the fact that their mistakes were costing the taxpayer millions of dollars.

Flash found a 1978 case in which an auto plunged off an icy bridge, killing the driver. A Department of Transportation engineer recommended installing a $22,000 guard rail. Nothing was done.

Five years later, another driver plunged off the bridge during a rainstorm, falling 40 feet to the gully below. The victim suffered a head injury that caused paralysis, memory loss and emotional problems.

Less than a year later, a third car slammed into a bridge girder during a heavy rainstorm. The driver suffered a broken back.

The last two victims sued the state and won nearly $1.8 million.

And the bridge? It still isn't fixed.

The Department of Corrections, which operates the state's penal institutions, turned out to be a major trouble spot.

One series of cases involved a convicted rapist who was released from the sexual psychopath program at a state mental hospi. Over the next three years before he was caught, the rapist brutalized 26 Tacoma-area women.

So far the state has paid some $640,000 for its mistake in releasing the rapist, settling lawsuits brought by 18 of the victims. And the total is climbing as more victims sue.

Flash also unearthed the existence of "Dr. Bigfinger," the practice at state prisons of performing body cavity searches on all inmates admitted to the maximum security units.

Prison officials maintained the body cavity search policy was necessary to keep the prisons free from drugs, knives and rapes. The inmates, however, charged that the rectal searches amounted to rape.

In 1987, a federal judge ruled the policy unconstitutional, finding that prison officials used the policy for punitive, rather than safety reasons.

The state stopped the "Dr. Bigfinger" practice when the first lawsuits were filed. But the issue of whether damages should be awarded to the inmates is being appealed by the state.

So far, the illegal searches have cost the state $194,528. The state has paid 65 inmates $200 to $500 for each search. Some cases a still pending.

Flash emerged from the project as a believer in computer-assisted reporting. "The computer pointed the way," she said.

Six successes in computer-assisted election coverage

By Marie-Susanne Langille
MICAR
Missouri School of Journalism

This is a roundup of computer-assisted election stories from around the country on everything from campaign contributions to voting patterns.

"Looking into the deepest pockets"

By Rick Mendosa and Sarah Acosta
Hispanic Business (October 1992)

This computer-assisted story shows how Hispanic contributions to congressional and presidential candidates, national party funds and political action committees more than tripled over the past three presidential elections.

Interestingly, a disproportionate amount of the $3.16 million Hispanics contributed in this election cycle as of March 31 is coming from Florida and is going to Republicans, according to the report. Mendosa and Acosta also isolated the largest Hispanic political giver.

The National Library on Money & Politics, a Washington-based non-partisan, non-profit research organization, searched FEC records from 1979 to 1992 for 3,000 Hispanic surnames supplied by *Hispanic Business* and isolated 30,000 contributions. Mendoza and Acosta took these records, in comma delimited form, and organized them in dBASE by year, party, state and type of contribution.

"Senate race ante up to $1 million"

By Dan Popkey
The Idaho Statesman (Oct. 11, 1992)

Reporter Dan Popkey used FEC data to see just how seriously Republican and U.S. Senate hopeful Dirk Kempthorne took his own campaign call to "take our government back from the politics as usual crowd."

Popkey's analysis showed Kempthorne raised twice as much money from political action committees as his opponent.

Using the same data, Popkey showed Kempthorne's financial support came from timber and mining interests while his opponent's came from labor unions and agriculture interests. *The Statesman* also documented how much of each candidate's support came from out-of-state.

The reporting took three to five days, Popkey said. *The Statesman* got the FEC data from the National Library on Money & Politics and used Lotus spreadsheet to sort contributors by address, occupation and name.

"Bush holds big lead in corporate donations"

By Lance Williams
San Francisco Examiner (Oct. 13, 1992)

President Bush fell behind Democratic candidate Bill Clinton in California popularity polls in this year's presidential campaign, but the GOP was ahead of the Democrats almost three to one in state soft money donations, according to the *Examiner*.

Computer-assisted reporter Lance Williams looked at FEC records on state corporate contributions to four Democratic and four Republican committees from the beginning of the election cycle through June 1st to show how the GOP had raised $5.5 million in Californian "soft-money" — beefy company donations exempt from election finance reform laws that are meant to lessen the influence of special interest money in presidential campaigns.

Williams' analysis showed the GOP has received particularly large donations from energy firms worried about laws restricting oil drilling, agribusiness interests worried about losing federal subsidies and real estate interests that would profit from a Bush plan to free up wetlands for development. The *Examiner* also listed the three top donors in each of six sectors, from energy to entertainment.

During this one-month, one-person project, Williams looked at more than 7,000 contributors and used dBASE to sort data according to interest group.

"Paying attention to political influences"

By Thomas Huang and Alex Marshall, with computer analysis by Tom Boyer
The Virginian Pilot and the Ledger-Star (Oct. 18, 1992)

A *Virginian Pilot/Ledger Star* survey of congressional records and a computer analysis of campaign contributions illustrated a political maxim that is easy to complain about but difficult to document: money buys influence.

Computer-assisted reporter Tom Boyer used an FEC database supplied by the National Library on Money & Politics to show an alignment between PAC campaign contributions and the political positions of three congressional incumbents. Boyer also isolated each candidate's top five donors and the amounts, as well as the top 10 donors overall.

Boyer recommends dealing with the Library instead of the FEC. "FEC tapes are expensive and cumbersome to work with, and to make sense of PAC contributions it's a huge help to have the detailed industry codes the National Library folks have added."

The best way to keep up to speed with contribution data, Boyer says, is have National Library data plus a link to the FEC's on-line system. The best way to use PAC-donation data, Boyer suggests, is in routine congress coverage.

"We've plowed campaign finance data into our day-to-day coverage of the Virginia General Assembly," he says. "Legislators hate it, and lobbyists hate it even more.... Some lawmakers become timid about carrying a PAC's water if they know their vote will end up beside their contributions in the next day's paper."

"Bush slipped in suburbs"

By Ted Mellnik and Allen Norwood
The Charlotte Observer (Nov. 11, 1992)

One reason for computerized analysis:

"Some lawmakers become timid about carrying a PAC's water if they know their vote will end up beside their contributions in the next day's paper."

— Tom Boyer, Virginian Pilot/ Ledger Star

A post-election story by the *Observer* matched precinct results with 1990 census data to show how the swing away from George Bush in North Carolina was more pronounced in cities and suburbs.

Bush support in state urban and suburban areas dropped since 1988 by more than 15 percentage points, compared with a 13.6 point drop across rural areas, according to the study.

The *Observer* also used census data to analyze election returns for four groups of precincts, each group dominated by people with similar education, occupations and household income.

The study showed Bush lost support among upper-income voters, mostly executives and professionals, and that most of those votes went to Perot. The votes Bush lost among middle-class voters were more likely to go to Clinton.

For this precinct analysis story, the *Observer* started with block-group level figures on income and education, then used Atlas GIS to make precinct-level estimates and SPSSPC to search for data patterns and to create precinct groups. The paper used dBASE to relate demographic class with precinct vote results.

"The politics of reform"

By Patti Epler, with computer analysis by Jerry Uhrhammer

The (Tacoma, Wash.) *Morning News Tribune* (Oct. 1992)

The *News Tribune* used computers to analyze the impact of a Washington state campaign financing initiative before the votes were cast.

Critics of Initiative-134, aimed at limiting the money going into political campaigns, said it was a political maneuver to boost the influence of business at the expense of labor interests. The criticism focused on a clause prohibiting the use of some union dues for political purposes.

The *News Tribune*'s analysis of 1990 campaign records from the state Public Disclosure Commission bore out the critics' claims by showing that I-134 would have cost Democratic candidates $1.2 million dollars in labor backing. Overall, I-134 would have cost the Democrats $400,000 more than the Republicans in campaign contributions for that election cycle.

New bill promises access to federal databases

By Suzanne Langlois
MICAR
Missouri School of Journalism

Pending legislation could make electronically stored government data more accessible to the public.

During the last session of Congress, the House Administration Committee approved the Government Printing Office (GPO) Electronic Information Access Enhancement Act of 1992. The bill would require the GPO to provide on-line public access to the Federal Register, the Congressional Record and a directory of federal information. The bill would also provide for on-line access to some publications under the control of the superintendent of documents and other departments.

The act (HR 5983) awaits a floor vote in the House and a vote in the Senate when Congress reconvenes.

The legislation is a boon for reporters. On-line access would make it easier and less expensive to retrieve and use government information.

"The bill would particularly help small news organizations by making available some of the things that can only be accessed through tapes now," said Penny Loeb, an investigative reporter for *New York Newsday* and a member of the board of directors of Investigative Reporters and Editors. "It would make it possible, for example, to go on-line cheaply and get background on a specific issue, like hazardous waste. It opens up a whole avenue of opportunity," she said. Loeb testified on the bill at the hearings.

The bill is the brainchild of the Taxpayer Assets Project and the American Library Association, and the result of the efforts of numerous organizations including the American Society of Journalists and Authors, Project Censored and the Coalition of Networked Information.

James Love, director of the Taxpayer Assets Project, anticipates congressional approval, but not of the legislation he originally sought.

"It's more a matter of how good a bill we'll get, not if we'll get a bill," Love said.

Rep. Charlie Rose (D-N.C.) and Vice President-elect Albert Gore sponsored the original legislation, commonly called the GPO Gateway/WINDOWS bills. But somewhere between their initial proposals and the revised House-approved version, lobbyists for commercial data vendors convinced Republican members of the House Administration Committee to narrow the scope of the legislation.

The most significant exclusion in the revised bill is funding, Love said. The original bills suggested appropriating $3 million this year and $10 million next year for development of on-line databases. The revised bill doesn't offer any appropriations.

The new bill also limits the range of on-line information. The Gateway/WINDOWS bills sought an array of on-line databases, including the federal LEGIS and JURIS systems, the SEC's EDGAR system, MEDLINE and the FDA Bulletin Board. The revised bill only mandates that the Federal Register, the Congressional Record and the federal information directory go on-line.

The original bills provided for on-line access through mainframe networks like Internet and through federal depository libraries. The revised bill abandoned mandatory network access in favor of a study of Internet access.

The new bill still allows for free access for 1,400 federal depository libraries and will charge all other users cost of dissemination only. Fees should be significantly less expensive than those charged by private data vendors.

Love said that Rep. Rose and other members of Congress have promised to give the bill priority next session.

Despite the limitations of the revised bill, the legislation would provide the public with government information for which taxpayers continually pay but can rarely access.

"We should all push to make it better," Loeb said. "Even though it's a small step, it's something."

Tech tips

Unraveling those mysterious string functions

By George Landau
St. Louis Post-Dispatch

String functions are the E. Coli of computer-assisted reporting. Digesting a database can be impossible without them, but who likes to talk about it?

I, for one, don't like even to admit knowing such stuff. But this being a forum for geekspeak, I swallow my pride and cough up the following.

All about string functions:

String functions are the features of a database manager or spreadsheet that allow you to alter text in a field through truncation or substitution.

Say you want to match two databases in which names and addresses are the only identifiers. One of the databases is a list of private lawyers who occasionally do work for the state; the other is a list of contributors to the campaign of the state's attorney general, who hires the private lawyers without competitive bidding.

The database of lawyers has records like this:

LAWYER NAME	ADDRESS
Dew E. Cheatham	2424-A Greenback Way
Hope U. Cansue II	One Mercantile Center

And the contributors database looks like this:

NAME	ADDRESS
Cheatham, Dew E. Jr.	2424 Greenback
Cansue, Hope U. II	#1Mercantile Ctr.

To minimize the amount of manual comparison you'll have to do between the databases, you'd want to standardize the formats of names and addresses. It would help to split each name field into four new fields: **LASTNAME, FIRSTNAME, MIDDLEINIT** and **JRSR.** You could split the address fields into **NUMBER, STREETNAME** and **STREETSUFFIX.**

I don't have the space to go through each step, but I'll explain how you could use string functions to extract each piece of information for the new fields. (Note: I'll be referring to string functions available in FoxPro 2.0; I believe that similar functions are available in dBASE and Paradox, although XDB users might find their selections of string functions to be more limited.)

FoxPro has a handy function called **AT( )**, which returns the location of a character in a string. For example, in the **NAME** field with the value "Cheatham, Dew E. Jr.," the function **AT(",",name)** returns the number 9 -- the position of the comma from the start of the **NAME** field (the capital "C" is in position 1). If there were two commas in the **NAME** field and we wanted everything to the left of the second one, the **AT( )** function could help us: **AT(",",name,2)** will give the position of the second comma. **AT( )** can have two or three items between its parentheses: the first item is the string to search for (it can be more than one character), the second is the field to search, and the third, optional item denotes which occurrence of the string to search for.

To extract the last name from the field containing "Cheatham, Dew E. Jr.," we want the first eight characters starting from the left. The function **LEFT(name,8)** would yield the string "Cheatham". To automatically extract all characters to the left of a comma, not matter where the comma appears, you can combine the **LEFT( )** function with the **AT( )** function: **LEFT(name,at(",",name)-1)** will do the trick.

Another function, **SUBSTR( )**, can pull a string of any length from any position in a text field. **SUSTR(name,3,6)** would return the characters "eatham": the function went to the third character in the **NAME** field and extracted the next six characters.

You could write a short program that scans through each record of the file, combining **SUBSTR( )** with the **AT( )** function to pull out text between the first and second spaces that occur in a name field with the format, "Firstname I. Lastname." Then, using the **LEN( )** function, you could copy that text to the **MIDDLEINIT** field if the text is only one or two characters long, or copy it to the **FIRSTNAME** field if it is longer.

Another string function that can prove mighty useful is **STRTRAN( )**. This one automatically searches a field for any string you specify and replaces it with whatever else you specify -- or deletes it altogether.

STRTRAN(name,"eat","chow") would replace "Cheatham, Dew E. Jr." With "Chchowham, Dew E. Jr." You can use this function to remove all punctuation -- **STRTRAN(name,".")** would strip the periods from the name field, for example -- or to standardize the spellings of small numbers in addresses or street names **(STRTRAN(address,"One ","1")** or **STRTRAN(address,"Tenth","10th").**

One last tip: A string function called **SOUNDEX( )** can be useful in matching phonetically similar words with minor spelling variations. The function returns a four-character code describing the sound of a word; "Diane," "Dyann" and "Deanne" all produce the same SOUNDEX code of D500. This function is useful in identifying possible matches that deserve closer inspection -- matches that the computer would otherwise miss.

If string functions seem a little complicated, don't worry. These things start to make sense only after you've experimented with them for a while. But they are worth learning, because they'll save you lots of time in the long run. If you have any questions, don't hesitate to call me at the Post-Dispatch: (314) 340-8296.

These things start to make sense only after you've experimented with them for a while. But they are worth learning, because they'll save you lots of time in the long run.

Bits, bytes and nibbles

■ OSHA DATA analyzes information for you

■ New M.E. furnishes a PC-friendly newsroom in Chicago

OSHA DATA, in Basking Ridge, N.J., offers analyses of Occupational Safety and Health Administration records.

Set up to do claims adjusting for insurance companies, OSHA DATA will do a one-time background check of a single company for $195.00.

Because this company performs many requests in any given month for insurance companies, the rates go down with multiple uses. If a company uses their services 30 times a month, for example, the cost per use is $105.

The OSHA violations database, which includes violations for the whole country from July 1972 to the present, is available from the Department of Labor. OSHA's phone number is (202) 219-7008.

OSHA DATA can do specialized searches such as which standard OSHA is emphasizing in a specific area, why some companies are never inspected or which safety and health officers are stricter when evaluating violations.

OSHA DATA can help its customers decide how to go about searching the database and explain the findings.

For information about OSHA DATA, write 28 Bradford Lane, Basking Ridge, N.J. 07920-1548 or call (908) 766-4551.

• • • • •

Resourceful reporters at the *Chicago Sun-Times*, accustomed to doing data analysis on PC's at home, got a morale booster with the appointment of their new managing editor.

Julia Wallace left *USA Today* in September to take the helm at the *Sun-Times*.

Wallace helped establish *USA Today*'s computer-assisted reporting program and was managing editor of special projects when she left.

"*USA Today* had a very structured program," explained Wallace. "We don't have the same resources here, but some people have already figured out how to use computers for the appropriate stories."

The *Sun-Times'* upcoming conversion to a PC-based network should help get data analysis out of reporters' living rooms and into the newsroom.

When *USA Today* started its computer-assisted reporting program, it was a luxury that gave the paper an edge over competitors, said Wallace. But she now considers such technology a necessity.

"Times have changed so much that it's just another thing newspapers have to be able to do," said Wallace.

Uplink

MISSOURI INSTITUTE FOR COMPUTER-ASSISTED REPORTING

120 Neff Hall
University of Missouri
Columbia, Mo. 65211