Uplink

July/August 2000

SMALL NEWS ORGANIZATIONS

# Driving the big CAR

**By Sean P. Johnson**
*Oshkosh Northwestern*

Sit down. Relax. Take a deep breath.

Now, search the recesses of your mind for that perennial favorite children's story about the little blue engine. Remember what he said: "I think I can. I think I can" all the way up the hill and "I thought I could. I thought I could" all the way down.

For small newspapers everywhere, this must be your mantra for introducing CAR to your newsrooms. Immediately banish from anyone's thinking the notion that CAR is strictly for big papers or big projects. It is a tool that every reporter and editor in every newsroom can use everyday if they approach it just like the little blue engine: You have to think you can do it and look for the opportunity, now matter how small it may seem.

That has been our approach at the *Oshkosh Northwestern*, a 24,000 circulation daily in Wisconsin, and it is an approach that has worked exceptionally well. Our reporters use CAR skills on an almost daily basis, often without knowing they are doing it. Most of the time, the application is small, but it prepares our reporters for tackling the big project when it comes along. They are comfortable with using CAR and believe that they can.

Here's how we have done it.

## Starting simple

It started out really as an effort to make life a little easier and more organized for everyone, but also provided a good primer on basic CAR skills. We encourage our reporters to do enterprise stories that cross beats. However, tracking down sources could sometimes be a pain if the reporter who normally covered a particular beat was not around to give you the information.

So we came up with a solution to create a database of all sources used by all reporters on their beats. We had the reporters type their information into spreadsheets giving names, numbers and other pertinent information. We took the largest Excel file and converted it to an Access database, then appended the files to that. We then created a front-end using the form function of Access, then Presto! A database reporters can use on a daily basis to search for sources, numbers and other information. The database can also be updated and edited through the front end.

Now, before you say "but that's not CAR,"

**Continue on page nine**

AWARD WINNER

## Case Dismissed

**By Gerry Lanogosa**
*WTHR-TV, Indianapolis*

Local law enforcement agencies spend hundreds of thousands of dollars on public service announcements warning that drinking and driving will not be tolerated. It's pretty tough talk: "We've been telling you, drive under the influence and you will lose your license and you will go to jail," said one spot that ran on local stations in Indianapolis.

But do the police always keep that promise?

Our team (Reporter Jeremy Rogalski, Photojournalist Bill Ditton, Investigative Producer Kathleen Johnston, and I) decided to find out. A tipster had told us a well-connected local lawyer was getting a large percentage of his DUI cases dismissed because police failed to show up in court. Sure enough, although we weren't able to prove any collusion, the lawyer turned out to be one of the top beneficiaries of police missing court dates. And our application

**Continue on page twenty**

## Inside Uplink

Computer-assisted reporting is making its way into small news organizations, where journalists are discovering its power in not only improving the quality of reporting, but also the efficiency.

Check out how Joe Kolman used CAR as a bureau reporter for *The Billings Gazette* (see page five) and get some suggestions from *Duluth News-Tribune* reporter Craig Lincoln about being a CAR specialist (see page three). And Charles Lussier, of the *Houma Courier*, tells about teaming up with several sister papers to tackle an environmental project (see page seven).

### Repeat offenders

Jeff Porter, of the *Arkansas Democrat-Gazette*, writes about his paper's investigation of the state's criminal justice system and repeat offenders. And James Walsh of the Minneapolis *Star-Tribune* follows up on his year-long look at offenders on probation repeatedly committing crimes.

SEE PAGE ELEVEN

### SQL Server

David Heath, of the *Seattle Times*, provides all the reasons CAR reporters should switch from Access or FoxPro to SQL Server.

SEE PAGE FOURTEEN

# Uplink

*July/August 2000*

*Volume 12, Number 6*
*A newsletter of the National Institute for Computer-Assisted Reporting*

**EDITOR**
Brant Houston

**DIRECTOR OF PUBLICATIONS**
Len Bruzzese

**MANAGING EDITOR**
Mary Jo Sylwester

**ART DIRECTOR**
Kerrie Kirtland

**SUBSCRIPTION ADMINISTRATOR**
John Green

*Uplink* is published every month by the National Institute for Computer-Assisted Reporting, 138 Neff Hall Annex Columbia, MO 65211. (573) 882-0684. Subscriptions are $40 for IRE members, $60 for nonmembers.

Postmaster: Please send address changes to NICAR. Send e-mail to jgreen@nicar.org

NICAR is a joint effort of Investigative Reporters and Editors and the University of Missouri School of Journalism.

NICAR services include hands-on newsroom training in computer-assisted reporting, special academic and advanced training in data analysis.


**NATIONAL CAR CONFERENCE**

# Kentucky Fried CAR

More than 50 panels and another 40 hands-on training sessions will be offered at IRE's national computer-assisted reporting conference in Lexington, Kentucky, Sept. 14 through Sept. 17.

The Thursday through Sunday conference, known as Kentucky Fried CAR is co-hosted by the *Lexington Herald Leader*. The conference will feature the leading practitioners and have sessions for the beginner and for the expert; for the reporter and for the editor; and for the print journalist, the broadcast journalist, and the online journalist. Conference sponsors include the Gannett Foundation, the Scripps-Howard Foundation, Knight-Ridder, and WKYT-TV.

The conference will take place at the Hyatt Regency in Lexington where rooms start at only $104 a night. We plan to have breakout sessions for people interested in special topics and lots of opportunities for small workshops in the computer demonstration room. And of course there will be opportunities to appreciate the attractions of Kentucky, from horses to bluegrass to distilleries.

To register for the conference, which costs only $150 plus $50 for non-members, you can go to *www.ire.org/training/lexington* or you can call IRE at 573-882-2042.

The conference will include a special batch of sessions on the upcoming Census data. The sessions will be led by experts Steve Doig of Arizona State University and Paul Overberg of *USA Today*. There will also be a track of using mapping software and a close look at the best and safest use of statistics.

## Other topics

- Racial profiling
- Getting ready for the hurricane season and other natural disasters
- The gun trade
- Doing your first CAR story
- Deciphering property records
- Finding the bad doctors
- Using CAR effectively at small newspapers
- Editing CAR stories
- When the plane crashes: From bad rudders to runway incursions
- Training the newsroom
- The latest, greatest CAR stories
- Unlocking the secrets of the prison industry
- Cool stuff: The latest in practical software
- Teaching CAR at the university
- The recipe for illness: Covering the food industry
- Driving CAR on the highway beat- Cam paign finance: What to do after the election
- 20 shortcuts to finding what you need on the Web
- CAR and broadcast: Getting past the time machine
- Children at risk: Daycare and foster care
- Decrypting death records: Ingenious uses
- Crime and violence: What data to use-Beyond the city limit: International data for local stories
- Tracking venture capital
- Measuring school test scores: More ways to an A
- CAR for the statehouse
- Plundering environmental data
- Public employees: Absenteeism, benies and other snafus
- Elderly care: Understanding the nursing home data
- Local government: Budgets, bids, purchases and change orders
- The Military Records Maze
- Building the Intranet for Your Newsroom
- Unveiling the hidden work of lobbyists
- Sports CAR: Stadiums, teams and profits
- The property scam: Landlords, land-flip ping, and blight
- Tracking an individual through on-line resources
- The issues and ethics of public access and the need for privacy

## Hands-on training

You can take up to three one-and-a-half-hour classes, which will include:

- Database managers: Beginning, Intermediate and Advanced (Taught in Access)
- Spreadsheets, Beginning, Intermediate and Advanced (Taught in Microsoft Excel)
- Statistics (Taught in SPSS with special sessions in SAS)
- Mapping: Beginning, Intermediate, Advanced (Taught in ArcView)
- Basic primer in Structured Query Language
- Using SQL Server
- A Primer in Perl
- Data Cleaning: String functions and more

SMALL NEWS ORGANIZATIONS

# CAR specialists

**By Craig Lincoln**
*Duluth News-Tribune*

Let us not dance around the issue any more: All newspapers, no matter how small, need a computer-assisted reporting specialist.

To say otherwise is to say that newspapers don't need copy editors who know how to operate pagination software, graphics artists who know how to make a bar chart, or even photographers who can shoot straight and can work in Photoshop.

This is my conclusion after a decade of trying to talk editors at two newspapers into making CAR a big part of our reporting. I've tried doing it on the sly with the computer sitting on the desk of my publisher's secretary after she went home, on my home computer, and as a sideshow to busy beats. None of that worked well.

**We're becoming the authoritative source on a lot of topics, such as the economy, that we simply quoted experts on in the past.**

I believe what we are doing at the *Duluth News-Tribune* (circulation in the 50,000 range) is working well enough to say it is the best way to go. My first job is taking care of computer-assisted reporting issues. My second is filling in on general assignment tasks when needed, just like any other reporter at a paper this size.

As a result, we're making CAR a regular part of the paper. We're giving our readers something that nobody else has in this market. We're becoming the authoritative source on a lot of topics, such as the economy, that we simply quoted experts on in the past.

So, I'll preach to the choir here. I hope I give you some ideas how to get this job if you are at a smaller newspaper and need ammunition. I'll tell you how we've succeeded here and how we've failed.

If you despair when thinking about the money required to get mapping and statistical software, take heart: We're succeeding without being anywhere near the cutting edge.

## Daily routine

I'm sure my days are quite similar to CAR specialists at bigger newspapers, although I also sense that the daily demands have a lot more variety. Reporters will come to me for help converting files or press releases into a usable format. They'll ask a lot of questions about basic math, or basic file management. Most weeks, I'll search our warehoused data for someone's phone number, address or criminal history at least a few times.

Mostly, however, I do three main tasks: Generate and write stories, do computer work and analysis for other reporters, and help editors plan ahead well enough so we can include data analysis in our stories.

Writing stories is the best part. For one thing, they can be fun. Each year, Duluth throws a big party that starts with a marathon. We've written about runners who have cut hours off their times, the demographics behind marathons' growth, and why our local marathon is financially successful. (It has more to do with the beer and bands after the marathon than the entry fees).

For another thing, I'm the reporter who gets to needle politicians about touchy subjects, too. For some reason, some of our officials think our region's population and economy are growing. The Census and our other reporting help us remind people (especially our mayor) that we aren't.

## On the beat

We've had a lot of success adding computer work to our beat coverage. The state of Minnesota releases its test scores on the Internet, on the same day at the same time for everyone. We'd look like country bumpkins if we weren't able to download the data and crunch it into something meaningful.

Our business team helped craft a package of economic indicators to track our region's economy. Now, every six months, we put out the most comprehensive and regular report on our economy that people have ever seen up here. It pays lots of dividends beyond helping

**Continued on page four**

**The *Duluth News Tribune* story, "Death Beds," is available in the IRE Resource Center, story #15068.**

**Tipsheets:**
**"CAR in smaller news organizations," tipsheet #769 from the 1998 NICAR conference. An audio tape of this panel is also available through Sound Images, Inc. (303) 649-1811, for $11. Ask for tape #11.**

**"Small and medium newspaper workshop," tipsheets #670 from a 1997 IRE conference includes a list of quick CAR stories and suggestions for starting CAR on a limited budget. Audio tape is also available from Sound Images. Ask for tapes 13 and 14.**

**"A little CAR goes a long way," by David McPherson, of the *Cape Cod Times*, for the 1999 CAR conference, tipsheet #947.**

**IRE Resource Center, (573) 882-3364.**

*Continued from page three:* # CAR specialist

our readers understand what's happening to their pocketbooks. We even have gained respect from our local university economists.

Finally, each week, I swallow hard and spend an hour sequestered in a room with editors so I can find out what's coming up over the next several weeks. That way, I can try to make sure we plan well enough to do all the computer work we should.

We have used computers on a few big projects. One, in particular, won a Headliner Award and was a finalist in the IRE contest (see "Death Beds," in the IRE Resource Center, story #15068). But I don't believe that's how we have made this beat an entrenched part of our culture.

Instead, my beat evolved into something I never expected and wouldn't have wished for, except that it has been successful so far. I have developed a few rules to keep it successful.

**Much of the success comes from a web of relationships that would be hard to dismantle.**

I thought my job would be judged mostly by the number of cool CAR stories I turned out. Nope. It's judged, it seems, by how much I spread around the help among the editors' teams.

I work with the school reporters on test scores, the city reporters on property values, anybody who covers an election on campaign finances, the business reporters on a bunch of topics, the crime reporters looking for scofflaws, and – most important here in the north woods – the outdoor reporters wondering where to find big fish. Much of the success comes from a web of relationships that would be hard to dismantle.

## My rules

So, my first rule: Don't be a Lone Ranger.

I also spend a lot of time on small, simple tasks. We did a quick analysis of property values around bed and breakfasts, for example, and it's common for me to help other reporters find the right demographic information for an info box.

Those good feelings go a long way, hence my second lesson: Think small.

And I also spend a lot of time turning my hard-earned data analysis over to reporters, who make some phone calls and write the story and get their name first on the byline. When I'm cranky, it seems they have fun and get top billing, too. But again, they should remember that help in the future and hopefully editors will want more of it.

So, my third lesson: Swallow your pride.

It also would be nice to have an editor that has the time and expertise to dig for money and technology, but I don't. Instead, I spend far too much time politicking and trying to find out what happened to budget requests. And "far too much time" isn't much, as busy as I am.

Hence, my final lesson: Learn the art of the well-timed memo.

Which brings me to my final piece of advice, whether it's useful or not. As I mentioned early in this article, I tried every way I could think of to get computer-assisted reporting in my newspaper but none of them worked. I even tried to talk editors into making it my specialty for years.

I finally succeeded with a well-timed memo. The message of the memo was simple.

I made two points: We said we wanted sophisticated reporting in our newspaper. CAR, of course, is the best way to go, but we were failing because data analysis always was second priority to beat responsibilities.

We also faced new competitive pressures from a string of community newspapers surrounding us. We had added a lot of community coverage to the *News-Tribune*, but CAR was the way to give our readers something they couldn't get anywhere else. That would make it more likely we would prevail in the newspaper war.

In other words, I said we were failing to get enough sophisticated work into our newspaper. And that making me a CAR specialist would help us win the competitive battle, i.e., make more money. (OK. That's a marketing

**Continued on page six**

**Newsroom Training:**

**IRE and NICAR can parachute into your news organization, bringing our laptop computers, stories from our morgue and time-tested training techniques. These private workshops range from basic computer-assisted reporting sessions for 20 reporters to tutoring experienced CAR reporters in statistical analysis, mapping and advanced data techniques.**

**Training costs can vary depending on the number of people trained and the difficulty in developing training materials. Give us a call and we'll be happy to talk to you about the possibilities.**

**Contact Trainer Tom McGinty at tmcginty@nicar.org or (573) 882-3320 or Executive Director Brant Houston at brant@ire.org or (573) 882-2042.**

# One-man show

**By Joe Kolman**
*The Billings Gazette*

The words of the Montana Department of Corrections spokesman made it all worthwhile: the intense week of NICAR Boot Camp, the hours of frustration inflicted on me by Access and Excel and the constant reassuring I had to give my editors that, yes, the story on felons with guns would be coming soon.

I was checking details of a story I was about to publish that said as many as 100 of Montana's most dangerous ex-cons – including murderers, rapists and kidnappers – bought hunting licenses last year, meaning they were in possession of rifles, shotguns or at the very least, archery equipment.

State officials acknowledged that no one had ever done such a check before. But the Corrections spokesman said they would start doing so.

"We thought your idea was pretty darn good," the spokesman said.

**CAR puts me in a position to do stories that before were the domain of much larger papers with more time, reporters and resources.**

Of course, it wasn't my idea. The week I attended Boot Camp, the *St. Paul Pioneer Press* did a similar story. Full of inspiration upon my return, I knew that such a story would work in Montana, a vast state where there is hunting galore and little computerized cooperation between agencies.

There is also very little computer-assisted reporting, which has been both good and bad.

## Open field

The obvious benefit is that the field is wide open. *The Gazette* is the largest paper in the state with a circulation of just 55,000. None of the papers have CAR reporters. There has been little training offered.

But even in my limited experience, I realized that much meaningful work can be done here with CAR. I obtained Y2K reports from local, state and federal agencies that listed scads of databases no one knew existed.

As I had hoped would happen, CAR has opened up a whole new world of reporting for me. I've always loved rooting through documents and reports. But I frequently found myself wishing I had access to more information.

The opportunities presented by CAR make me want to stay at the office long into the evening. For me – a one-man bureau almost 200 miles from the home office – CAR puts me in a position to do stories that before were the domain of much larger papers with more time, reporters and resources.

I found that I could work on data cleanup in between phone calls and daily stories. I've also found tons of help on the NICAR listserv. It is truly amazing to be sitting here cussing Bill Gates one moment and the next receiving instant help from CAR experts all over the country.

## On the flip side

Downsides include dealing with government officials who have never fielded electronic information requests. The responses ranged from flat refusals to asking me to sign a pledge that I would not use the information for mailing lists.

One state official was very concerned that I would make a mailing list out of his database of municipal sewer systems, as if I moonlighted as a parts salesman for treatment facilities.

To be fair, other officials simply e-mailed information, no questions asked.

## Facing the editors

On the home front, it was difficult to explain to editors just what I was doing and why it was taking so long. It is hard to explain over the phone the process of my data negotiations and the hours it sometimes takes to clean up information once I receive it. Then there are the hours of reporting and writing – not exactly what daily editors like to hear.

**Continued on page six**

**Kolman's stories, listed below, will soon be available in the IRE Resource Center. Search the database at *www.ire.org/resourcecenter* to find the story numbers.**

- **An examination of the low fees the Forest Service charges for such uses in the region as cabins on public land.**
- **Revealing that the number of concealed weapon permits have doubled over the last five years and are held by scores of state employees, including the chief justice of the Supreme Court and three lawmakers.**
- **Finding that Montana leads the country in the number of ice jams on rivers over the last century and also in the number of deaths related to ice jams.**

Here are a couple stories *The Columbian*, in Vancouver, Wash., tackled recently using CAR. They will soon be available in the IRE Resource Center.

Claims made against the city – how they were handled, how much was spent on settling the claims, what types of claims were made, etc. From the data, they put together lists of the top ten types of claims, the top ten costly types of claims, the top ten costly individual claims, which department racked up the most claims and for how much.

Using data from the county's GIS department, in the assessor's office, reporters produced lists of who owns land in the county – who owns the most acreage, the most forest land, the most in terms of value, etc. This took some time because reporters had to clean dirty data and create maps using Arcview, which were then exported for graphic artists to overhaul in Freehand.

*Continued from page five:*

# One-man show

And perhaps I erred by enticing editors with flashy stories like felon hunters and then, while I was working on that story, giving them small CAR stories that were not earth shattering. There seemed to be little interest in the less sexy stories such as which town had the most hunting and fishing licenses per capita or my finding of previously missing child support debtors by matching them up with hunting licenses.

But I was more than inspired enough and received enough encouragement to plow ahead. Heeding advice from the Boot Camp, I requested loads of information my first two weeks back. I continued to make at least one request a week so that I always had new records to examine.

## Story examples

Other stories I've done in the last five months include:

• A study of the financial impacts of non-profit groups in Montana.

• An examination of the low fees the Forest Service charges for such uses in the region as cabins on public land.

• Revealing that the number of concealed weapon permits have doubled over the last five years and are held by scores of state employees, including the chief justice of the Supreme Court and three lawmakers.

• Finding that Montana leads the country in the number of ice jams on rivers over the last century and also in the number of deaths related to ice jams.

One of the most fun CAR stories I've done came not from a spreadsheet or database but from negotiation. The state refused to even give me a sample of driver's history records. Instead, they insisted I buy them for $4 each from my local bureau.

So I bought my record, the record of a well-known traffic violator and the record of Gov. Marc Racicot. Guess which one of us had an expired license and had been driving illegally for six months?

Some of the governor's aides were sure that I'd been tipped. I simply told them that if they had turned over the information when I first requested it, it's likely I would have never found out. Unfortunately, they still refuse to provide the entire database.

It's hard for me to believe that it has taken me this long to catch the CAR bug. Being able to do this work has been rewarding both personally and professionally – I found that even with my limited experience my value in the job market increased almost immediately after I started using CAR for my stories.

My enthusiasm for CAR work is not shared by all the reporters I've talked to about it. The very idea of spreadsheets and databases makes their eyes glaze over. But for those who see the benefits, the rewards are great.

Joe Kolman, now a reporter with the *Omaha World-Herald*, can be reached by e-mail at joekolman@hotmail.com

*Continued from page four:*

# CAR specialist

argument. But remember my third rule: Swallow your pride).

It worked.

I also was lucky. I had some good editors at the time, good enough to understand where newspapers were headed. They didn't just want enough quotes to fill the white space. They actually wanted a newspaper.

We're far from being as far along as we should be. I haven't persuaded editors to make CAR training a required part of a reporter's experience here. We have been slow keeping up with technology over the past few years, and my pleas for mapping software haven't been successful yet.

But then the Census is coming up and I can't justify trying to report it without mapping software. I'm getting busier and busier and adding mapping to my duties means other people are going to have to start doing the simpler CAR work.

I'm just waiting for the right time for my next memo.

Craig Lincoln can be reached by e-mail at clincoln@duluthnews.com

# Sampling for answers

**By Charles Lussier**
*The Courier*

It seemed easy. Go to the state capitol in Baton Rouge for two or three months and see how well the Louisiana Department of Environmental Quality is protecting the earth, air and water.

The agency has long been a lightning rod for controversy, some of its own doing and some because of the sensitivity of its duties. Of particular interest was a 1998 study by the Tulane Environmental Law Clinic suggesting the DEQ was rarely penalizing polluters for environmental violations. The study, however, raised as many questions as it answered.

Louisiana's three *New York Times* Regional newspapers – *The Courier*, the *Daily Comet* in Thibodaux and the *Daily World* in Opelousas – decided to look closer. It was not an easy decision; the three newspapers have fewer than 25,000 readers each. But the editors decided the story was important, so we pooled our resources and tried to tell the story as best we could. I was picked to lead the investigation.

The difficulty in finding clear answers to simple questions, however, quickly turned three months into seven.

In the end, the resulting five-day series, published Feb. 21-25, showed that the state's top environmental agencies failed on many occasions to resolve long-standing pollution cases and hold the wrongdoers accountable. It also showed that the agency lacked the kind of critical information that could allow ordinary citizens to make their own assessment of the agency and how well it protects the state's environment.

## Data surprises

From the beginning, the plan was to obtain DEQ's electronic data on enforcement and let that tell the story.

That was easier said than done. As it turns out, DEQ has lots of data. But little of it covers more than any one branch of the agency, and most of it is full of holes and riddled with errors. In fact, the agency had more than 250 databases from which to choose. We eventually submitted public records requests for about half a dozen databases.

To help us extract the truth from this mess, upon the advice of a couple of IRE members, we sought outside help. I had had some limited experience dabbling with spreadsheets, but I felt I needed someone with more expertise to come along for the ride. Unfortunately, Louisiana is not blessed with many journalists familiar with databases and other computer-assisted reporting techniques. So, we placed want ads in a couple of newspapers. We ended up reaching outside the profession and hired Ashley McManus, a computer programmer living in Baton Rouge.

For close to two months, Ashley and I wrestled with this information, but were largely defeated by its many inherent flaws. The data had no common denominator. Typos were common. Many seemingly important data fields were started, but often unfilled or contained infrequent entries. In fact, only a few fields were even close to complete.

It was a classic "dirty data" situation. We tried to compare the numbers with official totals, but we found to our surprise that the agency collected no reliable official totals of its own which we could use to check our results. We were able to use this information only sparingly.

To get a more complete picture, I sought advice from Steve Doig, a professor at Arizona State University, an IRE member, and an unofficial adviser to this project. He recalled a crime study he had performed at *The Miami Herald* years ago in the days before desktop computers. Instead of obtaining a database, he pulled an old-fashioned statistical sample of a couple of hundred paper files. It was not of the caliber of a Gallup Poll, but the sample was helpful for discovering overwhelming trends.

## Sampling

We decided to pull a random sample of our own. It consisted of 200 DEQ cases selected at random from more than 11,000 cases contained in the agency's main enforcement database. On two laptop computers, we created Excel spreadsheets and filled the headings with an ambitious list of items we wanted to check. And then for another month, we visited DEQ public file rooms

**Continued on page eight**

**"Polluter's Paradise" can be seen online at *www.dailycomet.com/pollute.html***
**It is also available from the IRE Resource Center (story #16455), by calling (573) 882-3364.**

**Other recent environmental stories:**
**"Big polluters pay small price," by James Eli Shiffer and Susan Ebbs of *The News & Observer* (Raleigh, N.C.), found the state has the ability to fine companies that are polluting the environment, but state regulators have avoided using the full force of the law. (story #16431)**

**"Environmental Injustice," by David Armstrong of *The Boston Globe*, found that certain U.S. federal agencies preside over some of the deadliest waste sites in the country, and there is little that can be done to stop them. (story #15872)**

Tipsheets on environmental data:

"Investigations in health, environment and safety," by Penny Loeb of U.S. News & World Report and Ron Nixon of *The Roanoke Times*. (#1133)

"CAR on the Environmental Beat," by Ken Ward of the *Charleston Gazette*, (#859) includes a list of Web sites for environmental reporters and tips for starting environmental reporting projects using CAR.

Check for other tipsheets at *www.ire.org/resourcecenter*
To order, call the IRE Resource Center at (573) 882-3364.

*Continued from page seven:*

# Sampling

and pored over documents.

But once again, poor record-keeping caused problems. We were unable to locate a third of the cases. So one of the stories became the poor quality of the agency's data. We were also able to show that a significant proportion of the cases were either still open or had a status that could not be determined. This confirmed the impressions of many citizens that the agency was slow to act decisively and would tolerate pollution for years without bringing the cases to a conclusion.

The search also yielded specific stories. Three of the cases uncovered had local significance to the readers of our three newspapers. They were cases of pollution which DEQ proved ineffective at stopping. Reporters at each of the papers then pursued these stories in greater depth.

## One warning

Be careful about borderline trends. Our error ratio was 6.6 percent, which would be too high for a professional statistician, though it was fine for our purposes. If we had declared that a slim majority, or 54 percent, of the case files were missing, we would have been wrong. The real answer would have been that between 47.4 and 60.6 percent of the cases were missing. That's probably a majority, but not necessarily. We would have handed our critics ammunition.

Obviously the sample wasn't the end of our investigation. We still had a lot of work to do, interviewing sources and amassing various government documents. We also managed to use some of the more recent DEQ electronic data for some limited comparisons.

The sample, though, was crucial. Besides the authority it lent to our work, it also provided a worthwhile framework with which to examine paper files. Those days reading case files proved the most illuminating of the investigation.

It's ironic that an old-fashioned sample was more fruitful than modern-day database crunching. Obviously, I would have preferred to have the data gift-wrapped for me to analyze, but with poorly run government agencies that often doesn't happen. So, if you have the time, or even if you're just fitting in this kind of work while doing other things, a study of the sort we undertook can pack a lot of punch and shine a light on worthy information that you would not have otherwise obtained.

Charles Lussier can be reached by e-mail at charles.lussier@houmatoday.com

ENVIRONMENT

# State-level data

*Editor's note: This is an excerpt from a 2000 IRE National Conference tipsheet compiled by MaryJo Sylwester, of IRE and NICAR. The entire tipsheet is available from the IRE Resource Center. Check the database, www.ire.org/resourcecenter, for the order number.*

1) There are literally hundreds of databases on environmental issues, both at the state and federal level, and many have had little scrutiny from the media. Federal data are typically 2 years old by the time it is released, while you can usually get more current information from the state level.

2) Budget extra time to get the information. It could take months from your first request until you actually get the data. Environmental agencies at the state level seem to be even less experienced than other state agencies in handling FOIA requests.

3) Beware that environmental laws can vary from state-to-state, which will change not only the record keeping but also the interpretation of the data. This will also make it difficult to draw comparisons between states.

4) Study the topic thoroughly before diving into the data – particularly the regulations for the topic and state/region you're looking at. Covering the environment has enough stumbling blocks because of the technical jargon, but computerized environmental data can be even more cryptic.

5) The usual warnings about dirty data are even truer when it comes to environmental data. Be prepared to look through as many paper records as possible.

*Contiued from page one:* **Big CAR**

let me explain. The middle letter in the acronym is A, for assisted. That's just what this little database does: it assists reporters in finding sources that can make their stories better. That's what we are after. It also gives them practice at day-to-day database skills.

Of course, we do a lot more as well.

## Elections

One of the great strengths of CAR is the ability to quickly do math, organize information and show trends. One of the most helpful little projects taken on at the *Northwestern* is our election template, which is nothing more than a simple spreadsheet that does instant election math and analysis for us.

I came up with the idea for the template shortly after my first election night as city editor. Oshkosh is divided into some 40 election wards, which are grouped together into 12 voting districts. Now, the problem reporters were dealing with was the way the wards and districts were organized. As the city had grown, wards had been added. While they may have been geographically contiguous, they were not numerically sequential. It all added up to chaos on election night, as reporters tried to add the right wards together to get district totals for maps, graphics and analysis. Essentially, they were not able to do it on election night for the next morning's paper, and even the day after mistakes were often made.

**Our reporters use CAR skills on an almost daily basis, often without knowing they are doing it.**

To fix the problem, I spent a couple of hours one afternoon setting up a spreadsheet that would automatically add the ward totals and report them out for the corresponding districts. I also created portions of the spreadsheet that would calculate percentage of the vote and percentage turnout. And it all began to happen as the reporter typed in the information.

The result: instant election night analysis that allowed our reporters to write better election stories on deadline. We moved past reporting the basics of who won and how many votes to telling you where those votes came from and what that meant.

In the April 1999 elections, the spreadsheets proved their worth when we were the only source covering a hotly contested judges race to report that not only was the race decided by fewer than 100 votes, but that the losing candidate had lost the election in his home ward. That same year, our coverage of Common Council elections was the only reporting that showed how local college kids had taken to the polls and ousted an incumbent and helped a newly formed neighborhood group put two of its candidates on the council.

That's powerful stuff for readers, and it's all because of a simple spreadsheet.

## On to bigger things

OK. So we can do the simple stuff. But what about "major projects?" Well, it's that simple stuff that gave our reporters the ability to do just that.

In 1998, the newspaper began working on a story, the effects of which are still being felt today.

For several months, the *Northwestern* had questioned the city's handling of rental housing in this college town. Its regulation seemed ineffective. The potential for a disaster loomed. There was not an existing database of complaints to work with – only the paper inspection and notice logs kept by the city.

However, because of the skills our reporters had gathered from creating the source database and the election spreadsheet and other smaller projects, they were confident enough to build their own databases of those paper records. They first typed the information into spreadsheets, then converted it into database files. Then, they started looking for trends.

What they found shocked a lot of people. Property owners with political connections or financial resources were simply able to skirt the rules. But they didn't stop there. Using

**Continued on page ten**

**Stories the *Northwestern* has done using CAR:**

- Rental properties
- Street conditions
- Parking tickets
- Lottery numbers
- Age and gender trends
- Increases, decreases and types of court cases
- Dam inspections
- Gas prices and taxes
- Train derailments involving dangerous materials

**Examples of good everyday uses for CAR:**

- Bus accidents
- Farm machinery accidents
- Plane crashes
- Railroad grade crossing accidents

*The Oshkosh Northwestern* story, "No Easy Fix," is available in the IRE Resource Center, story #14992.

Tipsheets:

"Investigations at small papers: tips for editors," by David Stoeffler of the *Lincoln Journal Star* (1999 IRE Conference), tipsheet #997.

"Small newspaper investigations," 1998 IRE Conference. Tipsheet #849 includes a list of Web sites to make your reporting job easier and a list of small newspaper investigations.

"Big Ideas at small papers," by Tamar Stieber of the *Albuquerque Journal*. Tipsheet #44.

IRE Resource Center, (573) 882-3364.

*Contiued from page nine:*

# Big CAR

census and budget information, they were able to tie the addresses of chronic problem properties to poorer areas of the city and show how the declining property values there were driving up the taxes of property owners everywhere.

Now, that's a pretty good story right there. But the data work was only the beginning. Our reporters spent two months interviewing all the parties involved, using that data to confront those who denied there was a problem and to ask probing questions of all.

The end result was a three-day series that no one challenged as containing any false information. The city, realizing it was not capable of policing the problem on its own, forged an agreement with the local prosecutor and the state to improve enforcement and criminally prosecute problem cases. Some landlords got out of the business, and those who remained began meeting with city inspectors regularly to review housing code issues. (See a copy of "No easy fix" in the IRE Resource Center, #14992)

It was a big project accomplished by a small newspaper because reporters believed they could do it and had developed the skills to make it happen.

## Quick hits

Certainly, that project was large scale and the *Northwestern* has not pursued one as large since. However, we have done some good size projects within the last year, notably an analysis of city development policies that showed the poorest neighborhoods were often overlooked for street and other improvements, despite the fact they needed them most.

But, the *Northwestern* does a lot of "quick hit" CAR. All of our reporters have Internet access, as well as spreadsheet and database programs installed at their workstations. We encourage them to use them, and they do. It is common for reporters to download small data sets and spreadsheets, analyze them and include the results in deadline stories. They do it without thinking it's CAR.

The result is better stories with greater context and depth that mean more to the reader. And isn't that what all newspapers, including the small ones, are after?

CAR is the answer to getting us there, if only we allow ourselves to think we can.

Sean P. Johnson can be reached by e-mail at johnson@oshkoshonline.com

SMALL NEWS ORGANIZATIONS

# Cheap & easy data

**By Craig Lincoln**
*Duluth News-Tribune*

Here are a few ideas for CAR stories and for getting data without having much money or time:

• Demographics data are cheap. The key is, of course: *www.census.gov*. We have used the IRS county-to-county migration databases (about $10 a state) and a variety of small-area population estimates from the Census and Minnesota's State Demographer.

• Putting big charts of school tests scores in your newspaper can make you look awfully smart to your editor and your readers. If your state is like Minnesota, there will be a lot of demographic information in the databases about each school.

• Employee salaries are handy to have, even if it's just for the inevitable standoff just before the strike by a public employees' union. It's better just to get the data each year, rather than make last-minute requests.

• You can make your newspaper the expert on the local economy. We have used employment data, house sales data (work with real estate agents), car sales (work with your advertising department), spending data and other readily available economic indicators to write a report on the regional economy.

• Finally, this is how we cheat. We get national data from a central Knight Ridder purchasing pool, and a lot of Minnesota data from the *St. Paul Pioneer Press*.

If I weren't in that network, I would find a non-competing newspaper with similar interests and try to set up a similar bartering arrangement – especially for databases that are used more on a regular, look-up basis.

# Repeat offenders

**By Jeff Porter**
*Arkansas Democrat-Gazette*

Convict me once. Convict me twice. Convict me once again.

And I still might not get locked up.

That's the key finding in an *Arkansas Democrat-Gazette* story looking at the state's criminal justice system and repeat offenders. We reported:

Some repeat offenders get second, third and fourth chances before going to prison. They're out there, writing bad checks or stealing cars or breaking into houses. Some are doing worse.

Such is the case in many states, not just Arkansas. The reasons are complex and overlapping – crowded prisons, overworked courts, law enforcement's emphasis on drug crimes. While the system tries to sort out its problems, a few people take advantage.

## Starting point

The primary starting point was a database from the Arkansas Administrative Office of the Courts. The file contains more than 400,000 circuit court records. In Arkansas, circuit court judges preside over felony cases. Data fields include offender names, dates of birth, charges and outcomes of the cases.

**Although we set out to write a people story, numbers had a place. We used the courts database to create most of the statistics used in the story.**

One field missing: any unique identifier for the defendants. We semi-overcame that when we began our first step, finding out which offenders had the most convictions. We simply grouped by first and last names, plus dates of birth, to identify each unique offender. Using database software Visual FoxPro, we counted each conviction of each unique offender.

We also weeded out records we didn't want. At this point, we included only felony convictions. While circuit courts deal with felony cases, they also deal with misdemeanor appeals, and some felony cases are bargained down to misdemeanors. And we decided to not count, at this point, those charged with violating their probation. We did eventually fold back into the story the misdemeanor plea bargains, cases that didn't result in convictions and probation violations.

## Paper files

Next step: We chose a handful of repeat offenders and obtained both court and prosecutor files – a time-consuming process that involved trips from one side of the state to the other. But the case files generated details that the electronic information didn't include, plus a long list of source names: criminals, victims, witnesses, police, prosecutors, lawyers, judges. That turned the story into a story.

We went on to profile four repeat offenders, choosing them to represent the most common types of crimes and circumstances. One of our four was a violent criminal, but, typically, most repeat offenders commit property or drug-use crimes.

One such offender committed at least 10 felonies before a judge sentenced her to prison. Wrote a prosecutor to a judge, before her prison sentence: "In my seven years as a prosecutor, I have never witnessed a nonviolent offender who has attempted to manipulate the criminal justice system more." The story pointed out that since her release, she'd faced a series of hot-check warrants. That's not a big crime for most, but her history was different. In 1993, during a stay in Milwaukee, she bounced an $8,500 check at a car dealership and drove away with a Cadillac Fleetwood.

A few days after our story ran, her parole officer picked her up. She was later released again.

## The place for numbers

Although we set out to write a people story, numbers had a place. We used the

**Continued on page thirteen**

The *Democrat-Gazette's* story is available in the IRE Resource Center, story #16665

The *Star-Tribune* series, "Catch and Release Justice," is available from the IRE Resource Center, story #15606, by calling (573) 882-3364.

**It is also online at *www.startribune.com/news/metro/catch/***



# Eliciting change

**By James Walsh**
*Star Tribune*

*Editor's note: Walsh wrote about the first installment of this project in the September 1999 issue of Uplink.*

After all the delays, after the legal fights and technological obstacles, we were finally ready to proceed with the stories we'd planned to do all along.

Months earlier, the *Star Tribune* had requested jail booking records from Minnesota's 87 counties and probation records from the seven largest counties in the Twin Cities metropolitan area. We wanted to quantify the anecdotes we'd heard for years: Offenders on probation were repeatedly committing new crimes without receiving the threatened sentence from their previous convictions.

**The stories went beyond data analysis to find real offenders and real victims. It also helped elicit real change.**

The original idea was to build a database of the Twin Cities offenders on probation over the past five years, link it to another database of all offenders booked into jail over the same period to find the people who committed new offenses while on probation for old ones.

From the beginning, however, there were problems.

## No statewide database

Minnesota has no statewide criminal justice system database. Each county not only keeps its booking and probation records in different formats and in different systems, there are no unique identifiers to ensure that one John Smith isn't another John Smith. Those differences not only made it difficult for counties to share records with other counties, it made it nearly impossible for one county's probation agency to share data with the same county's sheriff's department.

OK. We'd create our own way to link the databases. Converting first name, middle name, last name and date of birth into a character string, we could create keyfields that would allow us to pull a thread through nearly 100 different databases of offenders.

One problem: The state's largest sheriff's department and several of the largest probation agencies insisted that the dates of birth of offenders were not public. Without those dates, how could we be certain that we were indeed finding the same offenders with multiple records in multiple counties? We filed several lawsuits.

We had dates of birth from all the counties but Hennepin, the state's largest. Using data sets from the state Sentencing Guidelines Commission, state Supreme Court information system and other counties, we found birth dates for more than 28,000 of the 89,000 people booked in Hennepin County from 1994 through 1998. More importantly, we tracked down the dates of birth for the county's 1,200 most frequent offenders.

The result of our work was the July 1999 publication of "Catch & Release Justice," a series of stories that showed how thousands of chronic offenders continue to cycle through jails over and over again with little serious consequence. The series also chronicled the pain and damage they cause by repeatedly creating new victims. But we weren't finished yet.

## Dates of birth

The lawsuits still were out there. And soon after publication, *Star Tribune* editor Tim McGuire took local officials to task in a column for withholding offenders' dates of birth. A week later, the newspaper reached settlement agreements with probation and sheriffs' officials and we finally began receiving the dates of birth.

McGuire urged us to write the stories that we had originally proposed - find the repeat offenders on probation at the time they committed new crimes. We moved forward.

Again, we faced problems.

Ramsey County was unable to provide

**Continued on page thirteen**

*Continued from page twelve:* # Eliciting change

offenders' conditions of probation in a computerized format, making impossible a cross-county analysis of how well criminals complied with those conditions. And Hennepin County, which supervises about 30 percent of the state's probationers, provided incomplete data. The probation department gave us data on 64,461 cases over five years but most were missing dates on which the probation was to end.

They also seemed not to know how many people were actually given probation.

When we turned to the Hennepin County District Court, which provided us with data on all sentenced offenders over the same five years, we found that there were really 102,000 probation cases. When we asked probation officials about this, they simply instructed us to "go with the higher number."

Using nearly 220,000 records, CAR Editor Dan Browning tracked overlapping probations – offenders sentenced and given new probation while still on the books for an older offense. We also examined whether officials knew when the offenders they supervised were booked into jail on suspicion of committing new crimes.

## Results

The stories went beyond data analysis to find real offenders and real victims. It also helped elicit real change. State lawmakers earlier this year approved development of a new statewide criminal history information system – signed into law by Gov. Jesse Ventura – that will eventually allow officials across the state to know when offenders are booked, charged and sentenced anywhere in Minnesota.

In addition, the series sparked action by the state's business community.

By lending their expertise and financial muscle, business leaders are helping speed the development of a statewide database of all offenders placed on probation. Two counties – Hennepin and Anoka – now download the names, offense information and conditions of all offenders with probation into a state computer system each night. Other counties will be added over the next couple years.

Charlie Weaver, Minnesota's commissioner of public safety, has credited the newspaper for providing the spark to change what had been a critical – but largely invisible – problem.

"Everyone is responsible for how we got to this place. And the offenders know how to play the system," Weaver said. "If there aren't consequences for the repeat offenders, they're going to continue to reoffend."

**James Walsh can be reached by e-mail at jwalsh@startribune.com**

*Continued from page eleven:* # Offenders

courts database to create most of the statistics used in the story.

For example:

Almost one of every four people convicted of a felony in Arkansas will be found guilty of a second crime.

They may not go to prison for a first offense, particularly if it is a nonviolent crime, such as burglary or credit-card fraud. Fewer than half of first-time offenders do in Arkansas.

They may not go to prison the second time, either. Or even the third or fourth. About three in 10 repeat offenders will simply be put on probation.

And if they violate the rules of their probation, as about one of three felons is at least accused of doing, they still may not do time. Arkansas' circuit judges send just six of every 10 violators to prison or jail. The rest get only a longer probation or a bigger fine or, sometimes, only a scolding.

We were fairly confident of our numbers, but we did bounce the statistics off several experts, including a set of professors at a local university. None were quoted in the main story, but they confirmed key statistics and made a series of helpful suggestions.

**Jeff Porter can be reached by e-mail at jporter@ardemgaz.com**

**Some Web sites recommended by Porter:**

**Bureau of Justice Statistics**
**Offers statistics, studies and surveys on criminal justice, including profiles related to probation and parole violation.**
*www.ojp.usdoj.gov/bjs*

**About.com's page devoted to crime and punishment**
**Lists sources and subjects to aid research.**
*http://crime.about.com*

**Findlaw Links for legal information, searches for court decisions and directories of lawyers.**
*www.findlaw.com*

Despite its title, one of the best books for learning SQL is "SQL for Dummies" by Allen G. Taylor.

A good book specific to SQL Server is "SQL Server 7 Developer's Guide" by Michael Otey and Paul Conte.

Believe it or not, the official documentation that comes from Microsoft – the Books Online on the SQL Server 7 install disk – is excellent and comprehensive.

TECH TIP

# SQL Server pays off

**By David Heath**
*The Seattle Times*

Here's a data nightmare. You're on deadline and you need to do an analysis of on-the-job deaths.

You've got OSHA's workplace safety data – so that's good – but the data's kept on four CD's. It takes you an hour just to copy the data to your hard drive. Then you discover that if you combine all the data into a single two-gigabyte table, Access chokes. You have no choice but to break it up.

What you thought would be a simple question – how many deaths has OSHA investigated in the past 25 years – takes you all day to answer. And that's just the start.

There's a simple way to avoid situations like this: Use SQL Server.

**SQL Server can handle a terabyte of data, so I've never come close to seeing it choke.**

SQL Server is high-powered database software from Microsoft that, as the name suggests, works on a server over a network. You can use the software without a server, but you lose many of the benefits that way. While I don't have personal experience with this, I assume that you could get the same benefits from other server-client software such as Oracle or Informix.

SQL Server has gained a reputation in newsrooms as a great tool for putting databases on an Intranet. In fact, it's also great for the kind of slicing and dicing of data that CAR folks do all the time. I have hardly touched Access or FoxPro in the last three years. There are many reasons for this, the chief one being that I haven't needed to.

Here's why:

• SQL Server can handle a terabyte of data, so I've never come close to seeing it choke. Access cannot handle anything exceeding a gigabyte, which is not a lot these days and can be quite frustrating.

• Your data is always warm. By keeping it on a server, you no longer have to waste time copying from a CD or restoring from a backup tape. As I was writing this, a reporter asked me to search for a name in several different databases. I was able to do it in less than five minutes because the data was ready to go.

• Because the server is doing most of the work, my computer isn't tied up when I'm doing queries. In fact, I can run several queries at once and still be free to do other things on my computer while I'm waiting for the results.

• SQL Server doesn't have as many point-and-click features for doing queries, which means you have to master SQL. But one advantage is that you can easily connect to a SQL Server database with Access or FoxPro and use their mouse-friendly querying tools. You can also easily import data directly into SPSS through what's called an ODBC connection. They are easy to create.

• That means that you can set up any reporter on the network to do queries directly from SQL Server in Access. And you can do analysis from any PC on the network. You don't have to worry about copying data from one machine to another.

• If your server is on the Internet, you can do queries from any computer on the Internet. You could go home, log onto America Online, open up Access and, with an ODBC connection, do queries on data kept in the office. The data is password protected, so it's secure.

## Intranets

On top of all that, SQL Server is in fact a great tool for putting databases on the Intranet. Here's my routine:

When I get any data, I automatically "upsize" it to SQL Server, which is no more difficult than importing the same data into Access. As always, I index the fields I'm going to be querying. Then I'm ready to go.

If the data have people's names in it or may be of interest to others, I create a view and put it on our Intranet. The view simply creates a façade that converts field names and codes into plain English. To create search pages on the Web, I've used Uniquery and

**Continued on page sixteen**

# Hidden costs

**By Robert Gebeloff**
*The Record*

A veteran police chief retires. The town council goes into closed session. And before you know it, the chief is relaxing on the beach with a cool drink in one hand and a six-figure lump sum payment in his bank account – far away from the angry mob at the next council meeting who just read about the deal in your paper, two weeks after the fact.

Sound familiar?

In New Jersey, lucrative buyouts are fairly common for long-time public officials, and the primary justification is unused sick and vacation time. While most private sector employers do not allow workers to carry over or accumulate these benefits, public employees enjoy a different standard.

Recently, we decided to take a stab at measuring the overall impact of such benefits in a way that went beyond the "outrage of the week" type story.

While the lucrative payments being handed to outgoing police chiefs and school superintendents made for good copy, we wondered what was going on at lower levels, when long-time police sergeants and teachers retired.

What reporters Jan Barry, Peter Sampson and Don Stancavish discovered was staggering – that local public employees had banked more than $200 million worth of unused sick and retirement time.

Moreover, they found, this cost was pretty much hidden from public view. In addition to settling the big cases in closed session, most government agencies failed to itemize the cost of paying retirees in budget documents, instead rolling the figures inside line items for salaries or miscellaneous costs.

## The back door

Fortunately, the reporters found a back door. State regulators had, since the mid-1990s, urged local officials to upgrade accounting practices to meet national standards. And as part of this effort, auditors for most local agencies had begun calculating the estimated payoff cost for unused sick and vacation time as part of the annual audit process.

The audits were a godsend. Instead of recapping the five most notorious buyouts, our reporters could tell readers in Ho-Ho-Kus how much they'd have to eventually pay retiring town employees and readers in Wanaque how much they'd have to pay retiring public school workers – with official figures generated by auditors working for the government agencies themselves.

But as the reporters began plowing through hundreds of audit documents – we cover 93 towns, 103 school districts, and a couple dozen county agencies – they quickly encountered a journalistic obstacle.

Context.

It would have been very easy to just stick all the big numbers we were finding in these audits into the newspaper, but the truth is that the numbers alone would mean very little to readers.

A $100,000 obligation in a small school district is not the same as a $100,000 obligation in a large county agency. And what does $100,000, $500,000 and $1 million mean to readers anyway – those are just abstract large numbers. Clearly, we needed to do some analysis before we could present this information in the paper.

## Putting it in context

We decided to crunch numbers on two levels. First, in an effort to see the story through the eyes of the reader, we decided to express our findings on a town-by-town basis. This meant combining the municipal, school and county obligations into one total figure for each town.

It wasn't as simple as adding up three numbers. Many of our school districts serve more than one community. And all of our county agencies cover multiple towns. So we developed a method for distributing the obligations among communities.

For county agencies, we made a spreadsheet of assessed property value and calculated the percentage of the county total in each town. Then we assigned that percentage to the county obligation. In other words, if Hackensack contains 2.2 percent of the county's assessed value, then Hackensack taxpayers are ultimately responsible for paying 2.2 percent of the accumulated unpaid sick and vacation days of county employees.

We created similar worksheets for regional school districts. Only, we based the break-

**Continued on page sixteen**

You can see *The Record's* entire series at: *www.bergen.com/news/payday09200004091.htm* It is also available in the IRE Resource Center, story #16526

**Recent CAR Stories:**

**"Bankrolling Suburban Politics"**

***The Detroit News***

**June 4, 2000**

***www.detnews.com/specialreports/2000/campaign/index.htm***

**A four-month investigation of campaign contributions to local politicians found local politicians raised $2.1 million for last year's campaigns in more than 40 Metro communities, tapping the real estate and construction industry, city contractors and employees and unions.**

**"Vulnerable Lives: The business of foster care"**

***The Denver Post***

**May 21-25, 2000**

***www.denverpost.com/news/foster0521a.htm***

**This investigation of Colorado's foster care businesses revealed "that the state's private foster-care system too often fails to protect abused and neglected kids."**

*Continued from page fourteen*

# SQL Server

Cold Fusion. Both work great because they require no programming. Who has time to become a Web developer?

(This could be the subject of a separate article, but the big advantage of Uniquery is that it allows you to search all databases at once or individually. The big advantage of Cold Fusion is that you can easily customize the search page.)

If you are especially ambitious, a la Tom Torok at *The New York Times*, you can use SQL Server 7.0's advanced features, such as English query. This allows users to ask questions in plain English, such as "How many firefighters were paid overtime last year?" SQL Server converts that to an SQL statement, such as "select count(occupation) from payroll where overtime > 0 and year = '1999' and occupation = 'Firefighter'".

SQL Server 7.0 also allows for full-text indexing, which means that you can search for a word or words in a field of text. That's much faster than using a simple LIKE query.

## What you need

What do you need to make SQL Server work? You can install it on your personal computer. But to get all the benefits I've mentioned, you really need a nice Windows NT server – one with at least 500 megabytes of memory and 96 gigabytes of disk space. SQL Server is insatiable when it comes to memory and space.

You also need to learn how to create an ODBC connection. With your mouse, click Start | Settings | Control Panel | ODBC Data Sources | Add | SQL Server. Then type in a name, type in the name of the server, click on Next, chose either NT authentication or a password, click on Next, change the default database, click on Next and then on Finish.

The biggest downside to using SQL Server is that to get full benefit from it, you really need to master SQL. The good news is that there's not a steep learning curve. Once you've gotten SELECT, UPDATE, CREATE TABLE, and INSERT down, you will be able to do almost any query with it. Also, you can use the SQL you learn in Access or FoxPro.

The challenge of learning SQL Server pays off, because you can use it for any database need you have. You can slice and dice. You can query data that's ready to go. You can connect data savvy reporters with Access. And you can make the data available to everyone via the Intranet. The software is specifically designed to work efficiently with multiple users. So why bother with anything else?

**David Heath can be reached by e-mail at dheath@seattletimes.com**

*Continued from page sixteen:*

# Hidden costs

downs on enrollment. In general, towns support regional districts based on the proportion of enrollment generated by each participating town.

So we split regional district figures along the same lines.

Once we finished this round of crunching, we were still left with a fairly abstract large number for each community. To bring this figure home, we decided to perform a property tax analysis.

We set up a spreadsheet laying out the property tax characteristics for each town: the tax rate, the total levy, the value of a tax point, etc. We then wrote formulas that calculated how much property taxes would go up if the obligation for unused sick and vacation days were added to the tax levy.

While no town would actually have to bear such a burden all at once, we explained to readers that this was a real cost and that it was a cost they would have to pay eventually in the coming years.

Meanwhile, I'm happy to say that our reporters did an excellent job of NOT bogging down the narrative of their stories with these details. The data gave the story real backbone, but the actual writing was filled with anecdotes and discussion of the issues.

For the paper, we printed a town-by-town chart that contained only the summary data, plus an explanation of our methods. We then put several more charts online that contained the nuts-and-bolts data.

**Robert Gebeloff can be reached by e-mail at gebeloff@bergen.com**

# Keep an eye on Ferrett

**By Paul Overberg**
*USA Today*

A new tool for digging into big federal data sets shows promise, and it's useful now, even in its buggy beta phase.

It's the new version of Ferret (Federal Electronic Research and Extraction Tool), renamed Ferrett (*http://ferret.bls.census.gov/java/netscape.htm*). Built in Java and available free to anyone with a Web browser, it lets users tap microdata – individual records – in federal economic, demographic and health surveys.

Let's say you want to know whether registration and voting rates for 60- to 65-year-olds are different from those of 65- to 69-year-olds.

Or how the age breakdown of union members compares with that of workers covered by union contracts.

**"Our goal is to develop an expert- system-type approach about the data," opening it to new audiences, said Cavan Capps, Census Bureau Ferrett project leader.**

Or the workforce distribution of your state's foreign-born population.

You can get all this from the Current Population Survey, a monthly 45,000-household sample that yields a wealth of annual demographic data snapshots in addition to its main job, producing the monthly unemployment rate. But only a few people know how to wade through the survey's hierarchical file structure, various weightings for household, family and person records, top-coded variables and other subtleties.

## In its infancy

Before the Web, using CPS required buying computer tapes and mastering arcane concepts with codebook and phone help. Ferret 1.0, released three years ago, simplified the task tremendously. It offered a Web interface to build queries, and also allowed users to slice out just the variables they wanted to download. It carried pretty good online documentation, but it still required subject familiarity.

The newest version is better. It's been renamed Ferrett to avoid trademark entanglements. It walks a user through the process of picking a survey and variables. You can also check off which values you want from each variable. Finally, Ferrett allows on-screen tabulation or download.

Once you've chosen the survey and variables, arranging them is much like using Excel's pivot table wizard. An animated graphic of a ferret wearing a mortarboard demonstrates.

It's not without problems. It's slow. A memory leak causes it to occasionally forget where you were in building a query. Some of the point-and-click steps require a second or third try before they're executed. Windows sometimes take three tries to close. The navigation is pretty good, but not great.

But even with all that, Ferrett's power makes it worthwhile if you need to slice up the CPS for answers that go beyond published tables.

## Future potential

Ferrett's real potential lies in its promise to open up other government datasets over an effort called the Data Web. The National Center for Health Statistics has helped with Data Ferrett's development, and some of its surveys should be available soon. Among them:

- National Health and Nutrition Examination Survey (HANES)
- National Health Interview Survey (NHIS)
- National Ambulatory Medical Care Survey (NAMCS)

Other Census-data sets that may soon be added include annual population estimates, annual County Business Patterns and the 1997 Economic Census.

Even state and local data sets may become available, said Cavan Capps, the Census Bu-

**Continued on page eighteen**

**Accessing Ferrett:**
***http://ferret.bls.census.gov/java/netscape.htm***

**Another viewpoint on Ferrett:**
**"Burrowing for Data in Ferret," a tipsheet by Jennifer LaFleur, formerly of the *San Jose Mercury News*, for the 1999 National CAR Conference. It is part of a group of tipsheets in file #946.**

**Recent CAR stories:**

**"Detroit police: Lethal force, lasting questions"**
***The Detroit Free Press***
**May 15-18, 2000**
***www.freep.com/news/locway/pomenu16_20000516.htm***
**An analysis of FBI statistics found that Detroit leads the nation's largest cities in the rate of fatal shootings by police. The paper also found "that the system intended to monitor police shootings has become a scale out of balance."**

**"Kozmo's digital dividing lines" MSNBC.com**
**April 11-13, 2000**
***www.msnbc.com/news/373212.asp?cp1=1***
**An investigation of Kozmo.com – a digital company specialized in delivery services – based on data from the U.S. Census Bureau reveals "that in the cities it serves, the company does not offer delivery to many neighborhoods with high concentrations of black residents."**

*Continued from page seventeen:*

# Ferrett

reau project leader. Already, Ferrett taps 350 gigabytes of data on computers scattered around the country.

"A Web browser is designed to be a document browser, but then you happen on data," Capps said. "Ferrett is designed to be a data browser."

## Experts to audience

Capps said that only a handful of labor economists and demographers understand CPS concepts, variables and weights and how to link them in time series to measure change. Building the rules that experts know into Ferrett, Capps said his team hopes to make the data self-aware, and useful to new audiences.

One example is state-level CPS data. Each March, the CPS panel is asked a host of questions on household members and their income sources the previous year.

The panel typically doesn't include enough people in each state to produce sound results. But two years of March data does, and that's how Census experts use it to publish reports. Capps hopes Ferrett will be able to understand such rules and make them transparent.

"Our goal is to develop an expert- system-type approach about the data," opening it to new audiences, he said.

In addition to the March demographic and income questions, the CPS each October carries school enrollment questions, and every other November surveys voting and registration in the just-past national election. Periodic supplements address computer usage, fertility, marital history, displaced and contingent workers and job tenure.

## An example

A simple example might help illustrate. I entered Ferrett and logged in with my e-mail address. (If you can't trust the Census Bureau's promise to never disclose it, who can you trust?)

On the introductory screen, only one option's available. Clicking on that button opens Ferrett's tabbed palette. A Windows-style left pane allows browsing and selection of the survey you need.

I chose the March 1999 CPS and searched for "age" and "union." Ten variables came back. I selected two and clicked on the "Browse variable" button at top right.

That gave me a screen where the variable's details were laid out, and check boxes allow the option of filtering by certain values. Selecting a variable there pushes it into the Data Basket.

Moving on to the "Get Data" tab, the option is to "Extract" (download) or "Tabulate."

Clicking on the latter opens a screen where the selected variables can be dragged and dropped as row and column variables, just like Excel's pivot table wizard.

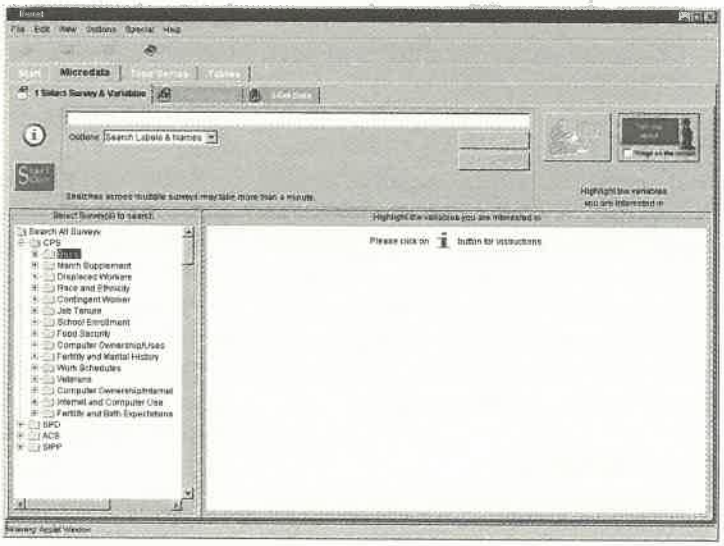
Clicking the green "Go" button launches the process. Going back and choosing "Extract" produces a download option.

Ferrett automatically weighted the tabulation so it got national totals, not just survey respondent totals. It also added a MARSUPWT variable to the download.

This is the weighting by which I'd multiply each record by to produce the national total. I didn't have to specify this, Ferrett just knew.

This tool isn't ready for quick answers or deadline work. A query can take several tries, Capps admits. Ferrett should come out of beta this fall.

**Paul Overberg can be reached by e-mail at poverberg@usatoday.com**



TECH TIP

# More on Perl

**By Matthew Ericson**
*The Philadelphia Inquirer*

*Editor's note: In last month's issue, Ericson explained the various ways a reporter could use Perl as another computer-assisted reporting tool. Here's a specific example of how he used it for a project at The Philadelphia Inquirer.*

One of the best examples of how regular expressions can be used to clean up data came when *The Inquirer* obtained a 700,000-record database of every reported crime in Philadelphia since the early 1990s.

Unfortunately, police had been sloppy in entering addresses in a standard format. Here's just a few of the ways that a crime at 400 N. Broad St could have been recorded:

```
BROAD 400 N
N BROAD 400
400 N BROAD
BROAD 400N
N400 BROAD
```

After looking at the data, we found that about 90 percent of the addresses fell into the first category – "BROAD 400 N."

However, with a data set of 700,000 records, that still left about 70,000 records with bad addresses. After a little more playing with the data, we found about a dozen different styles for addresses.

**The data gave the story real backbone, but the actual writing was filled with anecdotes and discussion of the issues.**

To standardize the addresses, I wrote a Perl program that went through the database record-by-record – as part of a SQL Server DTS script – and used regular expressions to pick out the various parts of the street address.

For each record, the program checked to see if the address field matched this pattern: a word, a space, a number, another space, and a direction. The syntax Perl uses to express this is:

```
(\w+) (\d+) ([NSEW])
```

How this worked with an address of BROAD 400 N:

- The \w indicates an alphanumeric character. The + says look for one or more of them. So \w+ means look for one or more alphanumeric characters in a row. So BROAD meets this condition. It's five alphanumeric characters in a row.
- The \d+ tells Perl to look for one or more digits in a row; 400 meets this condition.
- The brackets tell Perl to look for a single character. The letters in the brackets are the acceptable values. In this case, a street direction of N, S, E or W.

But where Perl outshines most other search-and-replace utilities is in taking pieces of the search term and rearranging them into the replacement. By placing parentheses around groups in the search term, the script remembered what it found.

The value matched by the first set of parentheses was saved in the variable $1. The value matched by the second set was saved in the variable $2, and so on.

So to turn BROAD 400 N into 400 N BROAD, the script looked for (\w+) (\d+) ([NSEW]) and replaced the address with $2 $3 $1.

Since the block number needed to be first, then the street direction, and then the street name, the program took the block numbers from the second set of parentheses – the (\d+), stored in $2 – and put it back as the first item in the replacement.

The street direction – stored in $3 – was the next item, and the street name – the first item searched for – became the last item in the replacement.

What if a particular record didn't match the pattern (\w+) (\d+) ([NSEW])? The program moved on to the next possible pattern - out of the dozen or so we were trying - and checked to see if it matched. If it did, it took the pieces, rearranged them in the proper order, and moved to the next record.

If it didn't, it tried yet another pattern. And if it exhausted all possible patterns, it flagged that record as an error and moved on to the next record.

Matthew Ericson can be reached by e-mail at mericson@phillynews.com

**More information about Perl, including the latest version to download for free, is available at www.perl.com**

**Perl, "Practical Extraction and Report Language," is a programming language that started as a utility to read text files and produce reports on Unix-based computer systems. It recently has been widely used for creating dynamic Web pages.**

"Case Dismissed," won a certificate (below top 20 markets) in the 1999 IRE Awards contest.

To see the full list of award winners and finalists, go to the IRE Web site, *www.ire.org/contest/1999awards.html*

*Continued from page one:*

# Case dismissed

of some very basic computer-assisted reporting principles allowed us to broaden our story by demonstrating a system-wide pattern of cases thrown out because of police no-shows.

## Finding the cases

Using a database obtained from the Marion County Justice Agency (our city's criminal justice information service), we found nearly one in 10 DUI cases dismissed in that manner. From the time we got our data, it took about three months to get our story on the air.

Our dataset covered all alcohol-related cases (we separated DUI cases out later) from 1993 through 1998, including case ID numbers; charge codes and descriptions; case disposition codes; judges, attorneys and officers involved with each case, and personal data on the defendants. It took the county three months to pull it together.

The law in our state allows agencies to charge for "enhanced access" requests like this. The local rate is a hard-to-swallow $50/hour for programming or processing. They didn't charge us for CPU time, but we ended up paying $650 for programming.

**Relentlessly question your data, and then be willing to accept the logical conclusions. Sometimes, that means throwing out cherished premises**

The records came to us via e-mail in three dozen Excel files containing more than a million records (about 151 MB). We imported them into an Access database for our analysis. After the initial clean-up, we had about 30,000 closed DUI cases. A few simple queries later, we determined that the number of dismissals because of police no-shows was about 25 percent.

It was a tantalizing figure – and it was wrong.

## Integrity checks

We've done enough of these projects - and heard enough horror stories at IRE conferences – to know that you never go with the first set of numbers. After some sober reflection and some painful integrity checks – a process that took several weeks and numerous consultations with the county's data guru – we came up with a final number of about 10 percent. That's still a significant finding.

Those integrity checks also involved going early and often to the paper files – that old-fashioned practice a dazzling database often tempts reporters to forget. By the time it was all over, we had pulled around three hundred hard-copy case files to make sure the data meant what it said. And we found that often it did not, which brings us to an overarching point that can't be overemphasized: Relentlessly question your data, and then be willing to accept the logical conclusions. Sometimes, that means throwing out cherished premises.

In our case, that turned out to be both a positive and a negative.

## The good and the bad

On the plus side, we started out thinking we might get a story about a lawyer working the system, perhaps by using his connections with the police. We couldn't prove that, but found a larger story about a system letting victims down by the thousands.

On the other hand, there were things we wanted to show but realized we couldn't. Take officer attendance rates. First, the records often failed to identify the arresting officers at all, so it was impossible to reliably determine individual rates.

Once we gave up on that one, we tried to determine the overall dismissal rate as a proportion of cases that went to trial. In other words, we knew the percentage of DUI cases thrown out because officers failed to appear, but what was the figure for the smaller number of cases that actually went to trial – cases in which officers were required to appear (as opposed, for instance, to cases settled through

**Continued on page twenty-one**

*Continued from page twenty:*

# Case dismissed

plea agreements)?

Obviously the rate would be higher. But, as we discovered in the paper files, what the database termed a "completed" trial event wasn't always accurate. Sometimes the trial was postponed and the case later settled; sometimes – inexplicably – there wasn't a trial at all, even though the data said there was. So another hypothesis went into the trash can.

Frustrating? You bet. But if we don't root out flaws and aberrations in our stories, the officials we cover might – and embarrass us in the process. However, when we do our job and ask those questions, it usually makes for a better story.

Case in point: Refiled cases. Halfway through our analysis, we wondered: What if prosecutors are refiling all these dismissed cases? It took us another week's work to find out, but turned out to be well worth it, because we found it was an extremely small number of cases – another example of the breakdown of the system.

Finally, this story bore out the truism that computer-assisted stories aren't about numbers alone. Our foraging through the court files yielded the individual cases that brought mere data to life. In the end, it wasn't the numbers (however startling) that made the excuses of the bureaucrats seem hollow. It was the heart-rending stories of victims like Charita Harris, left widowed by a drunken driver who might have been in jail if an officer had shown up in an earlier DUI case.

And you know what? Charita Harris's name doesn't appear once in our DUI database. Her name was buried in a paper file in the basement of the City-County Building, proving once again that many of the best computer-assisted stories still hinge on "old-fashioned" shoe-leather reporting.

**Gerry Langosa can be reached by e-mail at glanosga@wthr.com**

---

**EXCEL TIP**

# Subtotal to median

**By Mark Houser**
*Pittsburgh Tribune-Review*

The Subtotals function in Excel 97 is handy for a number of things, but if you want to use it to calculate medians, you're out of luck. That is, unless you outsmart Bill Gates' programming minions...

Let's say you have an Excel spreadsheet of 10,000 state employees, and you want to know the median employee income in each department. You sort by department to get all your employees bunched together according to where they work, then highlight the whole table and choose Data, Subtotals.

The dialogue box prompts you to tell it where to calculate the subtotals – in this case, at each change in department. It also asks you what function you want the subtotal to be – a sum for each department, an average, a count ... But "median" isn't one of the choices.

Damn the torpedoes! Choose "sum". Check mark the column you want Excel to do the subtotaling on – that'd be income – then click OK, and let Excel chug away and figure the sums where you want medians.

Now here's the trick. Highlight the column that contains the subtotals.

Select Edit, then Replace. In the Find What box, type: "SUBTOTAL(9," (without the quotes). In the Replace With box type: "MEDIAN(".

The original formula that figured the subtotals was something like "=SUBTOTAL (9,B2:B7)" – assuming your income figures were in the B column, with this particular department ranging from the second to the seventh row.

The replace all function changes the formulas, for all subtotals at the same time, to their appropriate Median formulas. So, the above example would be changed to "=MEDIAN(B2:B7)". Keep in mind this will completely replace the subtotals.

Voila! You have cleverly foiled Bill Gates' attempt to control you.

**Mark Houser can be reached by e-mail at mhouser@tribweb.com**

**Uplink story ideas:**
**Have you or one of your colleagues recently published a story using CAR that has not been done before or involved particularly difficult data work?**

**Do you know of a technical problem (or its solution) that others may like to hear about?**
**Or is there some issue or beat that we haven't covered?**

**If you have a story idea, we'd like to hear from you. Please contact managing editor MaryJo Sylwester either by e-mail at maryjo@nicar.org or by phone at (573) 884-7711.**

New data in the NICAR Data Library:

The Aircraft Registration Database, which is maintained by the Federal Aviation Administration Civil Aviation Registry, includes information on more than 320,000 registered aircraft. This dataset also includes two separate databases: A listing of all individuals/businesses registered as aircraft dealers; and another listing of all individuals/entities that requested to "reserve" a particular N-number.

NICAR will be updating this data every three months. You can arrange to receive each of the four annual updates on a subscription basis. For more information, including cost, go to *www.ire.org/datalibrary/databases*

TIPSHEET OF THE MONTH

# Newsroom training

**By Jennifer LaFleur**
*St. Louis Post-Dispatch*

*Editor's note: This is an excerpt from a tipsheet (#1158) that LaFleur, former NICAR training director, created for the 2000 IRE National Conference. The complete tipsheet is available from the IRE Resource Center, (573) 882-3364.*

From veteran nerds to those just learning the basics of computer-assisted reporting, we are all in the process of developing better ways to spread these skills throughout our newsrooms. Many models are developing, no single model is the answer. That's why sharing experiences at gatherings such as this is so important.

Here's an outline of some training models and the benefits and drawbacks of each.

## Throw it at everyone and see where it sticks

Almost five years ago, when the *San Jose Mercury News* first put computers on desktops in the newsroom, we did intensive training in all the basics (Internet, spreadsheets, databases) for everyone.

Benefits:
- Everyone gets exposed
- Most basic skills are widespread
- Get a lot of buy-in from the newsroom

Drawbacks:
- A lot of skills spread very thinly
- Very time-consuming
- Less pay-off, training a lot of people to create a few users

## The fellowship/the specialist

The *Post-Dispatch* plan requires reporters to apply for a six-week specialist program. They must have a story idea, data and be pulled off their beat to work with the CAR editor. Additionally at the P-D: they are then required to go back and be the CAR coordinator for their team and are part of a larger virtual CAR team.

Benefits:
- Greater pay-off
- Other people serve as mentors for their team
- Skills level is more in-depth

Drawbacks:
- Difficult to get team editors to commit
- Fewer skills throughout the newsroom
- Others may feel left out

## Outside training

Benefits:
- "Prophet in his own land" – outsiders are smarter
- Better commitment from participants
- Someone else is doing the work

Drawbacks:
- Cost
- Follow-up
- Specialization to your needs – for example, corporate spreadsheet training won't focus on journalism.

## Dos and Don'ts for training:

- Tailor the training to the reporters who will be using it. Use data that are relevant to them.
- Use real data that can be used to do real stories.
- Don't train more than 8 people at one time, unless you have a coach.
- Don't train reporters unless they will have access to the same software very soon afterward on which to practice their new skills.
- Don't hold a "gee whiz" CAR session if the reporters will not have hands-on training shortly afterward.
- Conduct a skills evaluation BEFORE designing the training program.
- Don't train folks with very different skills in the same room.
- Don't train reporters how to use resources that you don't plan to have available.
- Start CAR with some small projects with realistic goals.
- Don't expect too much immediately — everyone needs practice.
- If you use a commercial vendor, have them model your training around journalism.
- Provide lots of examples of what others have done.
- Encourage lots of questions.
- Provide follow-up with someone to turn to as they set out on their first projects.
- Don't just explain how, explain why and show them the process. Save shortcuts for later.
- Use workbooks they can practice with later.
- Evangelize and find experts.
- Don't make it seem too easy. Don't make it seem too hard.

Jennifer LaFleur can be reached by e-mail at jlafleur@postnet.com

# KENTUCKY-FRIED CAR: finger-clickin' good!

Investigative Reporters and Editors, Inc. and The National Institute for Computer-Assisted Reporting present...

## The National Computer-Assisted Reporting Conference 2000

Learn to find facts on the Internet, do database reporting on deadline, map crime, analyze budgets, perform statistical analysis and more. For more information, call us at 573-882-0684 or see the IRE Web site at www.ire.org

**Hotel Information:**

Hyatt Regency Lexington
401 West High Street
Lexington, KY 40507
Phone: 606-253-1234
Room Rate is: $104.00 single/double/triple/quad

To register, please complete this form. Visit our Web site at www.ire.org for the latest details.

REGISTRATION FORM

Please write carefully! This information will be used to make your name-tag.

Name: _____

Affiliation: _____

Address: _____

_____

City, State: _____

Zip Code: _____-_____ E-mail: _____

Phone: _____ Fax: _____

To attend this conference, you must be a current IRE member through 10/1/00.

_____($150) I'm an IRE professional member and would like to attend the conference Sept 14-17.

_____($100) I'm an IRE student member and would like to attend the conference Sept 14-17.

_____($200) I would like to attend the conference Sept. 14-17 and need to join or renew my U.S. membership.

_____($205) I would like to attend the main conference Sept. 14-17 and need to join or renew my international membership.

_____($125) I would like to attend the conference Sept. 14-17 and need to join or renew my student membership.

_____Total – Add ($25) late fee for registrations postmarked after August 25, 2000.

To register, mail this form and a check to NICAR, 138 Neff Annex, Missouri School of Journalism, Columbia, MO, 65211. To register by credit card, you must have a Visa or Mastercard. We cannot accept American Express. You may fax your registration to (573)882-5431 or register on-line at www.ire.org.

**Investigative Reporters and Editors, Inc.**
**138 Neff Annex**
**University of Missouri**
**School of Journalism**
**Columbia, MO 65211**

NON-PROFIT ORG.
U.S. POSTAGE
**PAID**
PERMIT NO.286
COLUMBIA, MO. 65211

# Bits, Bytes and Barks

## NICAR data updates

**FBI Uniform Crime Reports:** The 1998 version of UCR is now available from the NICAR data library. It is comprised of six databases and includes crime information reported to the FBI by law enforcement agencies around the country. Most of the data consist of the "index" crimes: murder, non-negligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny-theft, motor-vehicle theft and arson. Data from earlier years are also available.

**ATF Firearms Licensees:** This dataset was recently updated to reflect those individuals, organizations or businesses licensed to sell guns, as of March 2000. The database includes stores where guns are sold, rifle clubs, museums and pistol ranges who want to own guns that are off limits to them without the permit.

**IRS Tax-Exempt Organizations:** This database is a listing of the organizations filing an IRS 990 form – the form required for nonprofit groups in lieu of an income tax return. The data includes the organization's name, address, contact person, total annual income, total assets, and codes describing their activities generically. The records in this dataset include filings for tax periods ranging from 1974 through the first two months of 2000.

More information is available at *www.ire.org/datalibrary/databases.*

## File extensions

You just received data in a file with some obscure extension that you've never seen before. How do you figure out what it is?

A discussion on the NICAR-L listserv recently yielded several ways to hunt up the answer.

Properties: In the Explorer window, right click on the file and select "properties." Although this is not comprehensive, it will often show the type of file.

Glossary of File Extensions: This Web site lists thousands of file extensions for PC's. It does not have descriptions, however. *www.ace.net.nz/tech/TechFileFormat.html*

Every file format in the world: This Web site has thousands listed alphabetically. Most only list the name, but a few have more detailed descriptions. www.whatis.com/ff.htm

Another Web site: Similar to the ones listed above, this also lists file extensions used by various programs. *www.delanet.com/~pparish/filetype.htm*

## IRE Resource Center

Dozens of new tipsheets have been added to the IRE Resource Center's tipsheet database from a variety of panels that were held at the 2000 IRE National Conference in New York.

Tipsheet topics include Census 2000, day care, housing issues, environment, education, juvenile crime and justice, military, business and others.

The database can be searched at *www.ire.org/resourcecenter.* Many of the tipsheets can be downloaded or viewed in Adobe Acrobat. You will need to sign in as "guest" with "guest" as the password.

More information is available from the Resource Center at (573) 882-3364.