

## CHARLOTTE

# Get set for 2003 CAR conference

By Brant Houston,  
*IRE and NICAR*

Hands-on classes extended into the evening so you don't have to miss a favorite panel.

A special three-day track for those just starting computer-assisted reporting.

A series of smaller workshops for those more advanced in CAR.

These are just a few of the highlights of IRE's upcoming Annual Computer-Assisted Reporting Conference in Charlotte, N.C., March 14-16.

Although we had to do a quick hotel switch (now we are at the Hilton & Towers because of slowed construction of the Westin) we have just as many panels and classes planned.

This conference marks the 10th anniversary of IRE's first CAR conference, which also was held in North Carolina. In 1993, *The News & Observer* in Raleigh hosted several hundred

*continued on page 17*

## SPOTLIGHT: MEDICINE & HEALTH

# Databases show implant troubles

By Robert Gebeloff, *The (Newark, N.J.) Star-Ledger*

When *The Star-Ledger* first began examining the world of medical implants, we soon faced a major irony:

We were reporting on a flawed regulatory system and, to do it, we were crunching numbers based on data coming out of this flawed system. But we found that troubled data isn't necessarily useless data. It just requires a little more thought.

We already knew from documents and interviews that the Food and Drug Administration's regulation of the medical device industry was lax. The results of our analysis of FDA recall and malfunctioning device data provided numbers that illustrated the scope of the problem in our project, "Foreign Objects: The Risky World of Medical Implants." It showed us what kinds of devices we needed to examine and illustrated what kinds of problems were cropping up with these devices.

The idea of looking into the medical device industry, and specifically implantable devices, came out of our Washington bureau. Reporters Robert Cohen and J. Scott Orr had covered regulatory changes at the FDA for several years and had taken note of some medical device recalls that had made national headlines.

*continued on page 16*

## SPOTLIGHT:



This issue showcases some of the best uses of CAR to report on medicine and hospitals. Journalists in New Jersey used government data and added their own to show flaws with medical implants; others in Texas used a free statistical program to rate hospital performance. See the January-February *IRE Journal* for more about investigating hospitals.

# Digging into health, medicine

By Megan Christensen,  
*IRE and NICAR*

The toll of medical errors – estimated at 44,000 to 98,000 a year – gained national attention after the National Institute of Medicine released "To Err is Human" a few years ago. Citing death rates greater than those from highway accidents or breast cancer, the report challenged the medical establishment to reduce errors by 50 percent in five years.

*continued on page 18*

## Database Library News

### Small Business data (updated):

The Small Business Administration won't let natural disasters and setbacks stand in the way of America's love affair with greasy pepperoni. More than \$5.5 million has been awarded to businesses that serve cheesy slices and whole pizza pies from 1980 through September 2002.

The SBA's Disaster Loans database can offer details about how natural and other disasters affect businesses dealing with pizza, burritos, bagels as well as products and services that aren't edible. Journalists can discover who received loans and why.

The roughly 751,000 records include more than 9,000 recipients receiving almost \$1 billion in loans to help them recover from the Sept. 11, 2001, terrorist attacks.

As winter storms hit some parts of the country, you also can find snow removal companies to which the SBA has awarded more than \$1 million combined through general 7(a) business loans. Since the SBA began in 1953, it has offered loans through the 7(a) program to help small businesses in the United States.

Within the SBA 7(a) database of small business loans are nearly 830,000 records of companies that have received the guaranteed loans since 1953. Journalists can use this data to determine which lenders give small businesses the most

*continued on page 4*

## NICAR UPDATE

# Open Source: Free, but worth the cost?

By David Herzog, NICAR and Missouri School of Journalism

With newsrooms strapped with tight budgets, free open-source software sounds like a great deal for journalists. But is it?

With this issue, *Uplink* starts exploring that question with a series of road tests designed to see whether these open-source packages meet the demanding needs of computer-assisted reporting.

In recent months, the open-source movement has garnered heightened interest with the release of OpenOffice – targeted at users looking for an alternative to Microsoft Office. Also, Lotus Development Corp. founder Mitch Kapor gave the movement a boost when he announced his nonprofit foundation would develop an open-source personal information manager.

Under open-source rules, users may download and use the software at no cost. Programmers are encouraged to view and modify the software source code. Proponents say that quickens

the pace of software development and, in the end, results in better programs.

Already there are scores of open source programs for all kinds of applications. Looking for a geographic information system (GIS)? There's GRASS. Need a structured query language-based database manager? You can use MySQL or PostgreSQL. How about something as simple as a Web browser? You can download Mozilla, based on the Netscape Communicator code released in 1998.

With all the choices available, it's tough to know which programs are robust and which are still buggy. We'll give you answers.

Aron Pilhofer of the Center for Public Integrity, and former director of IRE's Campaign Finance Information Center, has volunteered to write our open-source feature and kicks off our coverage with a review of OpenOffice on page 14.

Contact David Herzog by e-mail at [dherzog@nicar.org](mailto:dherzog@nicar.org).

## Hands-on CAR training upcoming

IRE and NICAR has numerous training opportunities in the coming months for journalists who want to learn computer-assisted reporting. For a complete list of training events visit [www.ire.org/training/otr.html](http://www.ire.org/training/otr.html).

Here are some of the highlights:

There are three CAR Boot Camps in Columbia, Mo., in the coming months for journalists who want hands-on training in how to use spreadsheets and database managers to analyze data for high-impact

stories. The sessions are March 23-28, May 18-23 and Aug. 3-8.

The IRE Annual Conference, June 5-8 in Washington, D.C., will feature demonstrations of CAR techniques and hands-on classes. Members of the NICAR staff will be there to help you develop your stories and answer questions about computer-assisted reporting. Journalists can count on stellar panels and special presentations about national security, public safety, courts, the military, business, education and more.

## RELIGION

# Counting the faithful with church roll data

By Ron Nixon, IRE and NICAR

How many Baptists live in your community? What about Catholics or Lutherans? How many Jews or Muslims are there?

From 1850 to 1936, you could find answers to questions like these in U.S. Census results. But the U.S. government dropped a question about religious affiliation in 1946, partly because the Christian Science Church had concerns about invasions of privacy.

That left little demographic information about religious organizations. When the Census Bureau tried to revive the question for the 1960 Census, opposition from religious groups caused the agency to scuttle the plan.

Fortunately, journalists can find answers by analyzing the Churches And Church Membership In The United States survey conducted every 10 years by a group of religious organizations.

Led by the Association of Statisticians of American Religious Bodies, the group started to conduct its own census of religious group membership in the early 1950s. The survey provides the best available information on the membership of religious bodies in the United States.

More than three dozen news organizations recently used the 2000 data to examine religious beliefs and attendance in their communities.

## Tracking membership

*The Palm Beach Post* examined church membership in South Florida and found that Palm Beach County was the second-most religious county in the state, based on the percentage of people who claim some type of religious affiliation. The paper included color maps and graphics to enhance its stories.

The *Charleston Daily Mail* did a story showing West Virginia ranked second among the states in the number of congregations per capita.

In contrast, *The* (Tacoma, Wash.) *News Tribune* found that Washington ranks next to last among the 50 states in church attendance and membership, dropping from 47th in 1990. One-third of Washington residents attend or belong to a church, synagogue or mosque, compared with one-half nationally. *The Oregonian* of Portland, Ore., found similar results in its review of the religious census.

The *Herald News* of West Paterson, N.J., used the data to look at religious groups that claimed the survey underestimated their membership. Mike Corn of *The Hays* (Kan.) *Daily News* found that several of the reported figures in the survey conflicted with numbers provided by local religious groups. For example, the census showed that church membership for several congregations in one county had declined. Corn traced the decline not to membership losses, but to counting problems in the study. The churches had been included in the population counts of another county.

## Limits in data

The current census is a survey of about 150 religious organizations. The study is the only census to provide a county-by-county breakdown of religious participation. The study provides the total population for a

*continued on page 4*

## Other sources of religion data

While the Churches And Church Membership In The United States survey is considered the best collection of religious data available, it is by no means the only one. There are dozens of places to find data about religion, including the ones listed here:

[www.thearda.com](http://www.thearda.com) Located at Penn State University, the American Religion Data Archive holds dozens of datasets related to religions, some down to the local level.

[www.adherents.com](http://www.adherents.com) This is a collection of more than 41,000 adherent statistics and religious geography citations. Journalists can find adherent numbers by group and place.

[http://hrr.hartsem.edu/sociology/sociology\\_research\\_datasets.html](http://hrr.hartsem.edu/sociology/sociology_research_datasets.html) The Hartford Institute for Religion Research at Hartford Seminary in Connecticut specializes in studying religion in public life and includes

several datasets on its Web site.

[www.gallup.com](http://www.gallup.com) Opinion research firm The Gallup Organization has several polls with religion-based information.

[www.nazareneresearch.org](http://www.nazareneresearch.org) Church of the Nazarene Research Office. Good site with religion data plus interactive maps that can be queried on the Web site.

## For general reporting on religion, see these sites:

[www.religionwriters.com](http://www.religionwriters.com) The Religion Newswriters Association is a nonprofit membership association founded in 1949 to advance the professional standards of religion reporting in the secular press and create a support network for religion reporters.

Journalist's Online links for Religion  
[www.toad.net/~andrews/jrelol.html](http://www.toad.net/~andrews/jrelol.html)

## Database Library News

continued from page 2

money, as well as which businesses are receiving that money.

Another database from the SBA contains a list of companies qualifying for the 8(a) minority and other disadvantaged business programs. NICAR has offered this data in the past, but this latest update contains many new fields. Journalists can discover how many women, veterans, blacks, American Indians, Native Americans, Puerto Ricans, Hispanics and other minorities are active in the program.

### Immigration data (updated):

NICAR just updated its database of immigrants admitted to the U.S. as permanent legal residents. The data, maintained by the Immigration and Naturalization Service, contains characteristics of the aliens who were granted legal permanent residence in the U.S. during fiscal years 2000 and 1999.

In fiscal 1999, some 647,000 aliens were granted permanent residence. In fiscal year 2000, there were nearly 850,000. The database includes each immigrant's country of birth and country of last permanent residence, as well as the intended metropolitan area and state of residence. Journalists can track which countries most immigrants come from and which occupations are listed most.

To order or find out more about these databases, visit [www.ire.org/datalibrary](http://www.ire.org/datalibrary) or call 573-884-7711.

## Counting

continued from page 3

county, number of religious bodies and number of members and adherents.

The study is limited in many ways because the numbers are reported by the religious organizations and not all groups participate in the study. For example, in the 2000 census, 14 religious groups declined to participate, including some predominantly black congregations. Several groups participated for the first time, making comparisons between census years difficult.

Also, many Muslim groups have taken issue with the religious study, saying that it underestimates the total number of their members. An additional problem with the data is the way membership is defined by the religious groups that report. While many include only those who attend services on a regular basis, other groups include anyone with an affiliation to the religious body, including children.

To standardize the data, the ASARB uses adherents rather than members to examine growth of religious organizations. Adherents are adult members and their children. Membership totals include only those who are full regular members of a religious group.

Another problem with the data is that in a few counties, the number of adherents actually exceeds the total population. The ASARD says this could be because many people live in one county, but attend church in another.

Despite these limits, the data is the best available information for journalists reporting on religious group numbers.

The data can be obtained from the Glenmary Research Institute in Nashville, Tenn. The data comes on a single CD in Microsoft Access and costs \$100. Older datasets can be obtained from the American Religion Data Archive at Penn State University at [www.thearda.com](http://www.thearda.com).

Contact Ron Nixon by e-mail at [ron@ire.org](mailto:ron@ire.org).

## Better Watchdog Workshops

### Investigative Reporting on the Beat

Investigative Reporters and Editors Inc. and the Society of Professional Journalists, with funding from the Sigma Delta Chi Foundation, have joined forces to offer a series of workshops focused on doing investigative reporting while covering a beat.

The workshops, specifically for journalists at small- to medium-sized news organizations and those in bureaus of larger organizations, will emphasize the use of freedom-of-information laws and address juggling a beat while producing investigative and enterprise pieces.

**"You'll learn enough in the first 15 minutes to keep you busy for a month."**

*Kevin McGrath, The Wichita Eagle*

#### Workshops are scheduled for:

Feb. 22 – Tempe, Ariz.	April 5 – Atlanta, Ga.
March 8 – Storrs, Conn.	April 12 – Spokane, Wash.
March 22 – St. Petersburg, Fla.	April 12 – Long Island, N.Y.
March 28 – Minneapolis, Minn.	Oct. 4 – Eugene, Ore.
March 29 – Columbia, Mo.	Oct. 25 – State College, Pa.
April 5 – Cleveland, Ohio	

For more information, visit [www.ire.org/training/betterwatchdog](http://www.ire.org/training/betterwatchdog).

To request a workshop for your area, contact Executive Director Brant Houston at [watchdog@ire.org](mailto:watchdog@ire.org).



## BROADCAST CAR

# TV pioneer Cox rolls with changes

By Jaimi Dowdell, IRE and NICAR

One of the first times former WCCO-Minneapolis reporter Alan Cox attempted computer-assisted reporting, he used a spare computer loaded with Lotus 1-2-3 for DOS from the TV station's accounting department.

That was in 1990, even before he was officially the station's CAR specialist. Cox's experience of endless keypunching was one of the first in a long line of projects that created respect for the reporter's work and helped change the way broadcasters use CAR.

After a 25-year career in television, Cox left WCCO at the end of June. Since then he has been freelancing out of his home in Minnesota and has done some work for the *Star-Tribune* and KMSP-Minneapolis.

"He showed a whole crop of TV reporters that the same computer-assisted reporting skills that print journalists were regularly using to break big stories could also be applied to TV news," said Mike Wendland, technology columnist at the *Detroit Free Press*.

Brant Houston, executive director of IRE and NICAR, said "Alan has been a true pioneer in computer-assisted reporting in broadcast."

"He not only has led the way with stories, but also helped organize and teach our boot camps focused on the particular skills needed by broadcast journalists."

### Storytelling with CAR

After several years working as a general assignment reporter and then business reporter at WCCO, Cox received an opportunity to learn and use CAR when news director John Lansing approached Cox with a proposal in 1993. Lansing told Cox that

he was thinking of creating a computer-assisted reporting position and wanted him to do the job.

"Alan and I had the discussion, and he made the point that if we do it, we need the proper training and we need to do it well," said Lansing, now senior vice president of Scripps Howard Broadcasting.

**He crunched data and assisted with investigative projects and the day-to-day reporting in the newsroom.**

Cox was pulled off the regular schedule so he could learn how the tools of his new trade full time. Although he had little previous computer experience and did not even own a computer, he was eager to learn.

"My wife remembers that basically for the first year and a half, I stopped reading for recreation and just read computer manuals," Cox said. "I remember one day, all I accomplished was to learn to change the colors in Windows."

Once Cox put his skills to work, the station created a set for him called "Computer Center 4". From there, he crunched data and assisted with investigative projects and the day-to-day reporting in the newsroom.

Freelancer Julie Kramer was senior producer for investigations at WCCO

and worked with Cox on many stories. "A lot of people claim to do computer-assisted reporting, but Alan was what I call an alpha-geek," Kramer said. "He knew his stuff, but he wasn't just all computer; as a journalist he understood the importance of storytelling."

The stories Cox developed with Kramer and the WCCO investigative team were some of his favorites while at the station. Together, they tracked voter fraud in Minnesota by matching a database of criminal history records with a list of registered voters. After this story, Cox said Kramer had the idea to use the same felons data and cross it with a database of hunting licenses.

"We found more than 500 people who weren't supposed to have a gun that were licensed to use guns," Cox said.

### Freelance route

Cox also contributed behind the scenes. After attending a NICAR advanced Boot Camp in 1998, he set up an intranet so WCCO journalists could access information in driver registration, voter and criminal record databases for their daily reporting.

"It helped us find people that didn't want to be found," Kramer said. "It cut the paper trail time down immensely and a lot of reporters who couldn't do it on their own were able to do it quickly."

When Cox's contract came up for renewal last summer, he decided it was time to try something new. While he said he's not sure what that will end up being, he remains very interested in CAR and putting in the time to do it well. Those who worked with him at WCCO recognize his contributions to the station's success as well as broadcast reporting.

"If you want to have a serious investigative unit at a TV station you need a computer-assisted reporting element and we had that in Alan Cox," Kramer said. "We were able to do a lot of stories that we couldn't otherwise."

Contact Jaimi Dowdell by e-mail at [jaimi@nicar.org](mailto:jaimi@nicar.org).



## CAMPAIGN FINANCE

# Judges database helps uncover political games

By Michael Beebe and Robert J. McCarthy, *The Buffalo (N.Y.) News*

When a respected town justice in the Buffalo, N.Y., suburb of Amherst decided to run for state Supreme Court, New York's highest trial bench, he rented some billboards to announce his entry into the race.

In no time at all, he was called on the carpet by the Democratic party chief in Buffalo's Erie County, G. Steven Pigeon.

"Listen you SOB," Pigeon told the judge. "Don't you ever involve the general public in this process. It's my call. Nobody else's."

And when a veteran Supreme Court justice offered to mediate a dispute involving Pigeon at a judicial nominating convention, Pigeon told him, in the presence of witnesses including Buffalo's mayor, "Sit down and shut up you ..."

Both judges went on the record to recount these incidents, a rarity, as anyone who has covered courts knows. Judges, at least in New York state, don't say much outside the courtroom.

But we were able to convince these judges and others to go public in *The Buffalo News* last summer because of the spadework we did in discovering how political party bosses had taken over the system used to nominate judges.

A database we compiled of every state Supreme Court candidacy since 1999, and every local race since 1994, allowed us to go to these candidates and ask some tough questions. Among them:

- Why did those running for judge here, even those endorsed by both major political parties, raise thousands of dollars and contribute to dozens of political campaigns when state judicial

canons forbid them from doing so?

- Why was it that judicial candidates here — and nowhere else in the state — were each charged \$7,500 for judicial nominating conventions, when the actual costs were a couple hundred dollars to rent a hall and put out some coffee and doughnuts for delegates?

- Why were sitting judges putting the arm on local lawyers for campaign contributions when these lawyers could end up appearing before the very same judges?

The three-day report helped bring down the Democratic party chairman, sparking a change in how local judges are nominated and a pledge from party leaders to stop forcing judicial candidates to raise funds for others.

Before it became a series, though, it had been one of those things that gnaws at a beat reporter, having to get a story in the paper but knowing there's a far better tale to tell if only there were time to really pull it together.

### Contributions database

Robert J. McCarthy, the political reporter at *The News*, kept running across the same thing each election cycle. Why were candidates for seats on New York's Supreme Court — even those guaranteed election because they were endorsed by Democrat and Republican party leaders — still raising and spending thousands of dollars?

He wrote different versions of the story several times, but they always seemed to be buried among other election pieces, lost in the crush of more important election stories.

McCarthy tried several times to set aside the time to dig into this, but the

pressure of beat stories kept him from getting to it. He held onto a growing pile of campaign financial reports — the beginning of what came to be a valuable database.

Finally last year, Stan Evans, an assistant managing editor at *The News*, allowed McCarthy to carve out some time from his beat and teamed him with projects reporter Michael Beebe to work on the judges story.

The reporters hooked up with Andrew Bailey, a researcher in *The News*' library and our newspaper's database and spreadsheet expert, and got to work.

The first thing we did was download all campaign finance data for every state Supreme Court race in New York from 1999 to 2002 from the state Board of Elections. The board only started computerizing these records in 1999. So for local judges we used McCarthy's paper records going back to 1994. File clerks typed in the data.

We then created a database of judicial campaign finance reports, with every contribution and expense listed for anyone who ran for state Supreme Court. Although the series would focus on candidates in the western part of New York state, known as the eighth judicial district, we wanted to know how judge candidates here ranked with their counterparts.

We also downloaded the attorney directory from the New York Unified Court System Web site. We merged this with the contributions from our database and got a list of all lawyers who contributed to those running for judge. And because judicial districts don't conform to other, more recognizable jurisdictions, we used ArcView, a GIS program that allowed us more flexibility for graphics when the series ran.

### Interviewing data

Bailey set up everything in a user-friendly format that allowed us to run any number of queries on anything that struck our fancy. After awhile, we got comfortable enough so it was like in-

interviewing someone. In this case, we asked the questions of the data.

We found, as we suspected, that only political parties in Erie County charged a \$7,500 fee for a judicial nominating convention, and that the Republicans had abandoned it in recent years.

We found that we had the most expensive state Supreme Court races in New York. We found a judge without an opponent who still managed to spend \$75,000 and hire the party chairman's roommate as a campaign director.

And we found that party chairmen got around the ban on judicial candidates donating to other campaigns through a loophole, one big enough to fit a buffet table through. The exemption allowed candidates to buy tickets to political dinners, some of them costing \$5,000 each. So while the party chairman kept these candidates waiting months and months for endorsements, these potential judges kept getting billed for more and more tickets.

Armed with this data, we were able to interview with considerable authority and confidence the judges and judicial candidates. Once judges and candidates knew what we had and what we intended to say, the judicial veil of secrecy was dropped. It turned out they didn't like the system either, that it was demeaning and that judges were being nominated more for who they knew and how much they could raise, rather than what their qualifications were.

After the series ran, Pigeon, who had faced opposition before, withdrew from the race when it became clear he didn't have the votes to stay as party chairman. The new party chairman pledged there would be no more \$7,500 judicial nominating fees, and one prominent Democrat ran for judge on a pledge he would accept no more than \$100 from any lawyer.

Contact Michael Beebe by e-mail at [mbeebe@buffnews.com](mailto:mbeebe@buffnews.com).

Contact Robert J. McCarthy by e-mail at [rmccarthy@buffnews.com](mailto:rmccarthy@buffnews.com).



WASHINGTON, D.C.

June 5-8

### Don't miss out!

The best in the business will gather for dozens of panels, workshops and special presentations on public safety, national security, courts, the military, business, education, local government and many more.

### PLUS!

Hands-on computer classes, exhibitors and job recruiters.

### CAR HELP

Demonstrations of computer-assisted reporting tools and techniques, plus useful Internet sites. NICAR staff on hand to help you with your stories.

### Conference locations:

JW Marriott Hotel  
1331 Pennsylvania Ave.

Medill School of Journalism  
1325 G St. NW, Suite 730

National Press Club  
13th Floor, 529 14th Street NW

**For registration and more information,  
visit [www.ire.org/training/dc03/](http://www.ire.org/training/dc03/)**

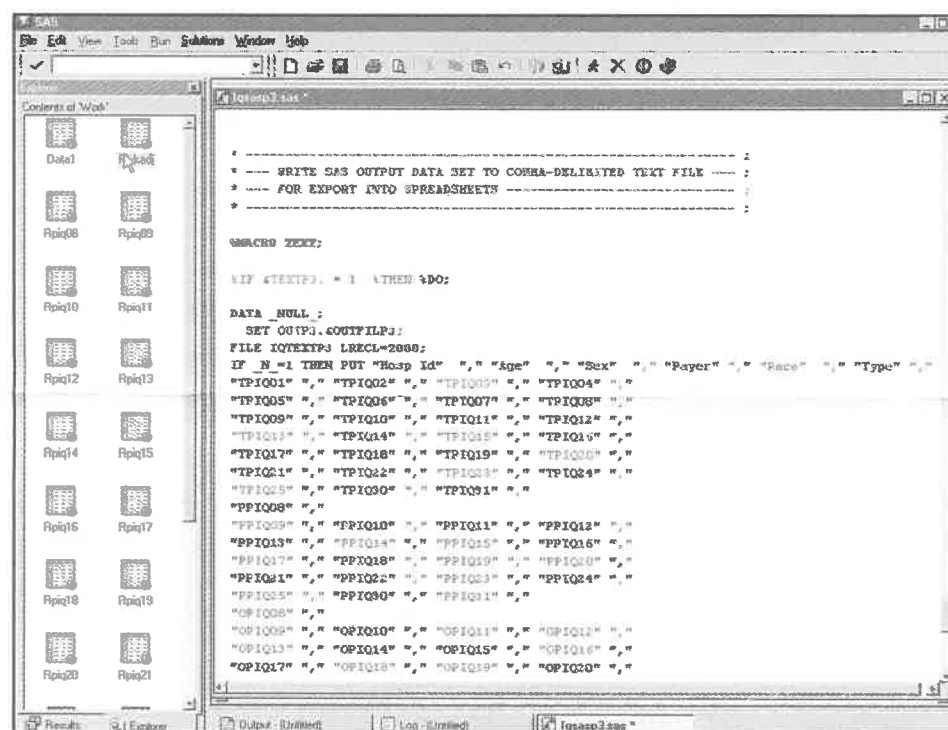
# SPOTLIGHT: MEDICINE & HEALTH

## Free stats programs compare hospitals

By Andrea Adams, *Missouri School of Journalism*

Analyzing millions of hospital discharge records might seem daunting, but customized statistical programs created for the federal government can churn through huge datasets in just a couple of hours – and help uncover interesting patterns in patient care.

cial report focused on specific medical topics, such as cardiac procedures and care for the elderly. The *Star-Telegram* posted the results of its analysis on the Web so readers could compare hospitals based on specific procedures or diseases.



The *Fort Worth Star-Telegram*, working with an outside researcher, used one of the programs to help analyze 5 million patient discharge records from the Texas Health Care Information Council for "Vital Signs," a July 2002 special report about hospital quality in Tarrant and Dallas counties.

The newspaper found discrepancies in death rates for a variety of diseases and in the performance of Caesarian sections. Also, the analysis showed that hospitals using new procedures and medicines were the most effective in helping patients. Many of the dozen stories in the spe-

### First evaluation

The Texas Health Care Information Council was created in 1995 to publish report cards about hospitals' performance rates for particular medical services. Nothing had ever been published by the agency. As a result, the *Star-Telegram* decided to conduct its own analysis to show readers how North Texas hospitals compared.

Obtaining the data was easy but costly. The newspaper paid \$8,000 for two years of data, which was provided on a CD-ROM for each quarter. Jeff Claassen, the newspaper's computer-assisted reporting specialist,

had decided to examine information from 1999 and 2000, the only two full years of data, to decrease the chance that a hospital's ranking might be distorted by an abnormally bad year.

The *Star-Telegram* decided to use the same computer program that the Texas Health Care Information Council used to analyze the patient medical procedure data: In-Patient Quality Indicators.

In-Patient Quality Indicators was created at Stanford University for the Agency for Healthcare Research and Quality, part of the U.S. Department of Health and Human Services. The program comes in two formats: SAS program files and SPSS syntax files. The agency provides the files and documentation for free download on its Web site ([www.ahrp.gov](http://www.ahrp.gov)) but users need to have a SAS or SPSS license to run the programs.

Rather than obtain a license, the newspaper hired Wayne Lehman, a research scientist at Texas Christian University, to conduct the analysis using SAS.

The program works by assigning codes to each patient discharge record based on that patient's type and severity of illness. It then compares the patients undergoing a surgery or treatment to a national sample of patients undergoing the same, and adjusts the hospital's mortality rate using the difference between the two. The risk-adjusted rates take into account how old or sick the patients are to make hospital comparisons fairer. The program also calculates the number of times a certain procedure was performed and the rate at which a hospital performed a certain procedure.

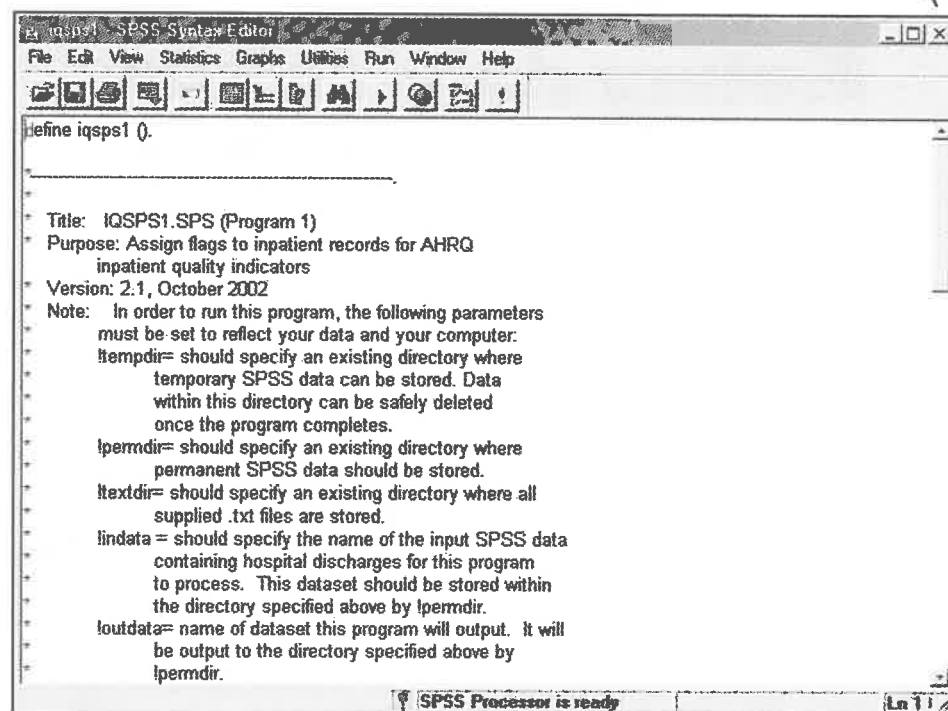
Before analyzing the SAS program results, which were compiled in an ASCII text file, Claassen verified the numbers. He made sure the calculations were done using the entire dataset and checked them against other sources. He focused his analysis on death rates for 10 medical pro-



cedures, primarily those involving heart problems and birth. Following advice from a researcher at the agency, Claassen included only those hospitals that performed the procedure at least 60 times during the two years to reduce the risk of unreliable statistics.

Claassen used FoxPro and Microsoft SQL Server to run queries, ranking the hospitals that performed best and worst at each procedure and comparing local hospitals to others in Texas. Charts of these results were compiled in Excel for medical reporter Charlotte Huff, business reporter Trebor Banstetter and Claassen to use. The three reporters called area hospitals with the best and worst rates to find out what kinds of factors can contribute to high or low mortality rates.

Claassen recommends using programs like the In-Patient Quality Indicators because it provides the first widely recognized risk-adjustment methodology, decreasing the chances that a news organization's



analysis would be criticized. Claassen said journalists who would like to do a story like this elsewhere should keep in mind that many hospitals are non-profit organizations and therefore not

subject to open-records laws. To get around this limitation, he suggests reading medical journals so you can ask better questions of hospitals with high or low mortality rates.

## Heart bypass mortality

Medical research suggests that hospitals with lower mortality rates are more likely to do a procedure well. For more information, click here.

Hospital (click for more)	City	Deaths	Patients	Risk-adjusted rate
Medical Center of Arlington	Arlington	7	202	2.08%
North Hills Hospital	Richland Hills	4	109	2.13%
Doctors Hospital	Dallas	3	260	2.72%
The Medical Center of Mesquite	Mesquite	4	239	3.16%
Baylor University Medical Center	Dallas	67	1,943	3.34%
Baylor Medical Center at Irving	Irving	26	427	3.84%
Harris Methodist H.E.B.	Bedford	11	251	3.85%
Medical City Dallas Hospital	Dallas	36	928	4.08%
St. Paul University Hospital	Dallas	11	284	4.17%
Methodist Medical Center	Dallas	9	297	4.31%
Plaza Medical Center of Fort Worth	Fort Worth	23	797	4.37%
Presbyterian Hospital of Dallas	Dallas	40	888	4.48%
Baylor Medical Center at Garland	Garland	14	370	4.55%
Arlington Memorial Hospital	Arlington	16	466	4.86%
Zale Lipshy University Hospital	Dallas	9	244	4.94%
Harris Methodist Fort Worth	Fort Worth	31	730	5.03%
RHD Memorial Medical Center	Dallas	6	118	5.75%
Baylor All Saints Medical Center at Fort Worth	Fort Worth	32	646	6.51%
Osteopathic Medical Center of Texas	Fort Worth	12	178	6.62%

Mortality rates are calculated for all patients except those transferred to other acute-care hospitals.

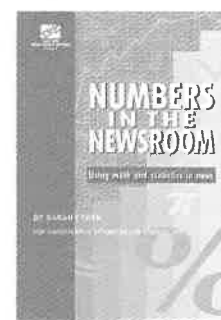
Star-Telegram

## REQUIRED READING For Your Newsroom

### Numbers In The Newsroom

Using Math and Statistics in News, by Sarah Cohen

Pulitzer Prize-winning journalist Sarah Cohen guides reporters through fractions, rates, percents and per capitas. Making inflation adjustments. Understanding averages. Doing the budget story. Questioning surveys and polls.



### ORDER TODAY!

Call 573-882-3364 with your VISA or MasterCard -OR-

Visit our Web site at [www.ire.org/store](http://www.ire.org/store) for online ordering or order form downloads

IRE MEMBERS: \$15 each • NONMEMBERS: \$25 each

## PUBLIC HEALTH

# Eatery data shows poor food handling

By Fred Vallance-Jones, *The Hamilton (Ontario) Spectator*

The cook in the small family restaurant wanted to make sure the eggs weren't greasy.

So before the heaping plate of breakfast went to the customer, she used a reusable cloth lying on the counter to dab the eggs. She then put the grease-laden rag down, ready for the next meal. It would be used for six hours before being washed.

Such was the scene witnessed by a public health inspector in Hamilton, Ontario, an industrial city with nearly half a million residents, about an hour's drive southwest of Toronto.

The inspector wrote about the incident in his notes, one of more than 2,000 pages of inspection reports released to *The Hamilton Spectator* as part of our investigation into restaurant safety.

The resulting five-part series, called "Reservations," revealed shocking food-handling practices in the back kitchens of well-known eateries.

But the series was more than just a recitation of horror stories about mouse infestations, rotting food and cooks who never washed their hands. Using computer-assisted reporting, it also revealed that the city health department was frequently failing to do the inspections required to protect the public, and even when serious problems were found, little or nothing was being done.

Senior managers at the newspaper had wanted to do a story on restaurant safety for some time. There would be high reader interest, and similar CAR reports by other newspapers had led to reforms in several North American cities, including Toronto.

The responsibility eventually fell to me

because of my special interests in CAR and public documents research. Little did I know that getting out of the starting gate was going to take the better part of a year.

## Obstruction 101

My goal was to obtain a copy of the city's database of food premise inspections so I could see which establishments had the worst records and whether the city was doing the minimum inspections required to protect the public.

But the municipal health bureaucrats were in no hurry to hand over a disk.

Ontario has an open-records law that specifically applies to municipalities. Under it, the city was supposed to make a decision on my request within 30 days of my filing it in March 2000. In reality, it took five months and a meeting with the city clerk before I finally got my first response.

The answer was no. I could not have the database. But if I was willing to hand over a check for more than 1,000 Canadian dollars (about \$630 U.S.), they'd be happy to make paper printouts.

The health bureaucrats said they couldn't provide an electronic copy.

The provincial health ministry had designed the database system and provided it to municipalities to make it easier for them to report inspection data to the ministry. In addition to being able to spit out what the province needed, the system allowed for some standard printed reports, including a list of inspections for any one premise. The city's staff had no idea how to make a copy of the underlying data they had themselves entered.

When questioned they did provide a

key piece of information, however. The application was written in Microsoft FoxPro.

As I had used FoxPro extensively in the past, I now knew it should be straightforward to write a query to produce the data I wanted. I called the ministry people who had designed the application, and after some initial reluctance, they agreed to write the query and provide it to the city to run.

By November 2000 I finally had a disk in hand, the cost about \$30.

## Table troubles

Once my disk arrived, I was not surprised to find the data was dirty.

There were two tables. One contained information about establishments, each identified with a unique ID. It linked to detailed inspection information stored in the second table.

In the establishment table, a handful of the ID numbers were duplicated, with more than one establishment having the same number. Usually it happened because a premise had changed hands (and names) and someone had neglected to create a new ID. In two instances the mistake was inexplicable. I chose to eliminate those two premises from any further analysis.

The inspection table was twice as dirty. Some inspection entries were duplicated, and a later review of paper records revealed that a small number of inspections were not entered at all.

I had to design queries to eliminate the duplicates, which could be identified from one of the fields. As for the missing inspections, they seemed few in number and health bureaucrats maintained the data was at least 98 percent complete. The deficiencies were later revealed to readers.

To decide which paper files I wanted to see, I used a field that recorded the number of critical violations found on an inspection visit. A critical violation is a mistake in food preparation or handling that

has the potential to cause food poisoning, such as storing foods at temperatures where bacteria thrive and grow.

I compiled a list of premises that either had large total numbers of critical violations or a large number of failing inspections. I then submitted an informal request for the detailed paper files.

Without the data, we could never have zeroed in so well on premises that had consistently had problems.

### Dirty dining

I received the paper records in early 2001. From there, the project sat on the backburner until May, when City Editor Jim Poling assigned me to work on it full time.

By that point, I had received an updated run of the inspection data from the city. Using Microsoft Access, I queried the database extensively and determined how many times each premise had been inspected, as well as how many inspections the department was doing.

Using Excel, I calculated how long passed between individual inspections of each establishment. All of this revealed that the city was falling far short of the inspection schedule required by provincial law. In some cases, years passed between visits, a fact that I confirmed with the owners of the establishments themselves.

There was still a key piece missing.

Roger Gillespie, the night managing editor at *The Spectator*, observed: It was important that the reader understand the potential consequences of a system that failed. That meant documenting cases of food poisoning.

I filed a new freedom of information request to another branch of the public health department, asking for records detailing outbreaks from the previous year and a half.

While I waited for those documents, I interviewed restaurant owners, food

safety experts, the inspectors and top health bureaucrats. By early September I had received records documenting several dozen outbreaks or suspected outbreaks of food poisoning and was tracking down one of the victims, who had suffered the humiliation of seeing the guests at her wedding meal fall ill.

When "Reservations" launched on Nov. 24, 2001, we led with the wedding story and documented the failure of the city to live up to its obligations to protect the public.

On Monday (*The Spectator*, like many Canadian papers, does not publish on Sunday), we wrote about the restaurants with the worst inspection records. Three more days of stories followed, including a look at systems used by other cities to disclose the results of inspections to the public. We also ran sidebars giving necessary background on food safety issues and posted summary information derived from the city database on our Web site.

The response from the public and politicians was overwhelming in the true sense of that overused word.

My voicemail was overloaded, letters and e-mails poured in and our Web site was swamped as outraged readers found out for the first time about filthy conditions in restaurant kitchens and the city's failure to make them clean up.

By the second day of our series, the politicians were already promising a cleanup. By the time the last of the stories ran, councillors had voted to hire additional inspectors and to develop an action plan to fix the problems identified by the newspaper.

In the meantime, shortly after our series ran, the owner of that family restaurant ordered his staff to stop using the "egg cloths." He still insisted the cloths posed no risk to the public, but his diners apparently thought otherwise.

Contact Fred Vallance-Jones by e-mail at [fvjones@hamiltonspectator.com](mailto:fvjones@hamiltonspectator.com)

## Annual Computer-Assisted Reporting Conference

— MARCH 14-16, 2003 —



### ★ Panels that touch on every beat: ★

Census, crime, education, local and state government, freedom of information and transportation

### Hands-on computer classes:

Spreadsheets, database managers, mapping, statistical software and more

### For hotel reservations call:

800-445-8667 and ask for the Investigative Reporters and Editors room block before Feb. 14

### NEW LOCATION

Hilton Charlotte & Towers  
222 East Third St.  
Charlotte N.C. 28202

### Sponsors:

*The Charlotte Observer*  
NBC News

### REGISTER ONLINE AT

[www.ire.org/charlotte](http://www.ire.org/charlotte)  
or call 573-882-2042

# Tech tip...

## Using SAS code in other programs

By Jeff Porter, IRE and NICAR

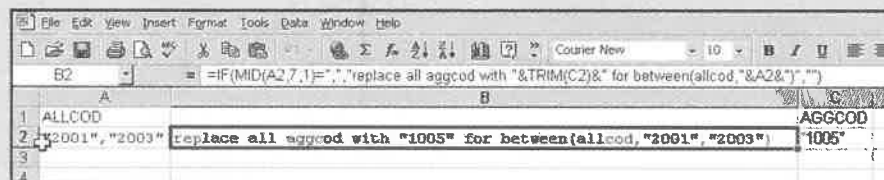
SAS out of your budget? Think again. You can use some SAS files with the more widely used programs such as Excel, Access or FoxPro to ease your work considerably.

SAS programs abound in state and federal agencies, and are often downloadable from the Web. They're valuable even for non-SAS users.

Using language in SAS programs – ending with the .sas suffix and easily read with Notepad or any text editor – you can create lookup tables or UPDATE queries in Access or REPLACE commands in FoxPro.

The IRE and NICAR Database Library recently did some analysis for a project on the causes of death for the *Columbia Missourian* newspaper using a mortality file containing data of every death in Missouri during a 10-year period. A SAS program provided by health department authorities in Kansas helped us write a FoxPro program to categorize causes of death.

The official codes of causes of death, called the International Classification of Diseases, have changed over time. So the data has a different set of codes for deaths 1999 and after, and those before then. Those classification codes are extremely detailed. For example, there are several codes that represent a death caused by a faulty heart valve. Depending on what you need to know, that might be helpful. For the purposes of the *Missourian* project, though, we needed fewer, broader categories. For example, we wanted a "heart disease" category that would cover those heart valves and several



other causes. And we preferred to have causes of death reported in plain English.

Fortunately, Kansas has a SAS program to do that.

Here's a snippet:

```
proc format library = death
print;
    VALUE $TIR31GP
'2004','2005'='1001' /* TUBERCU-
LOSIS */
'2009'='1002' /*
SEPTICEMIA */
'2010'='1003' /*
SYPHILIS */
'2015'='1004' /* AIDS/
HIV */
'2001'-'2003','2006'-
'2008','2011'-'2014',
'2016','2017'='1005' /* OTHER
INFECTIONS-PARASITES */
'2018'-'2040'='1006' /*CANCER */
'2043'='1007' /*
DIABETES */
'2048'='1008' /*
ALZHEIMERS DISEASE */
'2049'-'2059'='1009' /* HEART
DISEASE */
```

```
'2060'='1010' /*
HYPERTENSION */
```

```
'2061'='1011' /*
STROKE */
```

```
'2062'='1012' /*
ATHEROSCLEROSIS */
```

This part of the program takes deaths from several categories and aggregates them. For example, "2049" through "2059" are grouped together into another code: "1009." The text inside of the slashes are notes that are not part of the SAS program.

A couple of preparations, then we're off to write some simple lines of SQL. We need two additional fields in the Missouri death data. One will include the new codes such as "1009." Let's call it "AGGCOD." (Stands for aggregated cause of death.) The other will include English translations of those codes, also courtesy of the SAS program borrowed from Kansas. We'll call that field "TRANCOD." The field that already includes the range of codes we're going to aggregate is called "ALLCOD."

Now, with a lowly word processor or Excel, you can write commands to FoxPro.

Example: Copy and paste a section into a Word document, then on the tool bar, activate the button that looks like a paragraph mark. It al-

lows you to view hidden characters. In this SAS file, we noted that the line before "OTHER INFECTIONS-PARASITES" has an extra paragraph return. We deleted it. Also, some, but not all, of the category notes were preceded by two tabs before the first slash mark. We inserted those to make all of them the same.

Then, we copied that into an Excel spreadsheet, then used Data > Text to Columns... to make a new column, using the equal sign as the delimiter. Now, we have the new codes in a separate column. We made the appropriate number of columns between the old codes and the new ones – by simply counting the commas – and used Text to Columns... again to delimit the columns with commas. Now, we have single codes or a range of codes in separate columns.

Rather than go through old ground we'll touch on just a couple of items regarding a function called BETWEEN. (MaryJo Sylwester, of *USA Today*, wrote an article for the September/October 2002 *Uplink* showing how to use Excel to write FoxPro program language.) Notice in several categories there are ranges of codes. For example, the "OTHER INFECTIONS" codes include "2001"- "2003." We'll do a simple find-and-replace in Excel to change those hyphens to commas, then we're ready to write a program line to throw in a BETWEEN function.

In our example, the codes "2001", "2003" are in cell A2. The codes we eventually want to populate the just-created data field called AGGCOD are in cell C2. So in the cell B2, we'll type in this language:

```
=IF(MID(A2,7,1)=",","REPLACE all
AGGCOD with "&TRIM(C2)&" for
BETWEEN(ALLCOD,"&A2&"),"")
```

In other words, if there's a comma separating the two items – that is, if

language will eventually wind up in the field called TRANCOD. Again using Excel, with the equal sign as the delimiter, the spreadsheet software becomes a tool of writing SQL commands. No BETWEEN function



there's a range in the particular cell and not a single value – then plug in the language needed. If not, then leave the cell blank.

The result:

```
REPLACE all AGGCOD with "1005"
for BETWEEN(ALLCOD,"2001","2003")
```

That line tells FoxPro to go to the new field AGGCOD and plug in the code "1005" every time the field ALLCOD values fall between "2001" and "2003." We just repeated these steps a few more times, resulting in a long list of SQL commands.

Then we were ready to move on to the next step, again from a short example of the Kansas SAS file:

```
proc format library = death print;
VALUE $TIR3IDS
'1001'='TUBERCULOSIS'
'1002'='SEPTICEMIA'
'1003'='SYPHILIS'
'1004'='AIDS/HIV'
'1005'='OTHER INFECTIONS-PARASITES'
'1006'='CANCER'
'1007'='DIABETES'
'1008'='ALZHEIMERS DISEASE'
'1009'='HEART DISEASE'
'1010'='HYPERTENSION'
'1011'='STROKE'
'1012'='ATHEROSCLEROSIS'
```

The first codes, of course, are in the AGGCOD field. The rest of the

is needed this time; it's a straight translation for one code to a text label. Assuming that the codes are in A column and the terms are in C column, in B column, the Excel language would be:

```
= "REPLACE all TRANCOD with
"&TRIM(C2)&" for AGGCOD = "&A2
```

The resulting SQL line:

```
REPLACE all TRANCOD with 'TUBER-
CULOSIS' for AGGCOD = '1001'
```

If one wanted to do the same work in Access SQL language, it would be slightly different. The Excel formula looks like:

```
= "UPDATE deaths SET TRANCOD
= "&TRIM(C6)&" WHERE AGGCOD = "&A6
```

The results:

```
UPDATE deaths SET TRANCOD = 'TU-
BERCULOSIS' WHERE AGGCOD = '1001'
```

To download the FoxPro program created from the SAS file, go to [www.nicar.org/techtips.html](http://www.nicar.org/techtips.html). There are even more techniques for using SAS materials without the SAS price tag. We'll share them in a future edition of *Uplink*.

Contact Jeff Porter by e-mail at [jeff@nicar.org](mailto:jeff@nicar.org)



# OPEN SOURCE Office package needs more work

By Aron Pilhofer, *Center for Public Integrity*

Last spring, Sun Microsystems released a new version of its StarOffice suite for PC, promising it would deliver the same functions as Microsoft Office at about one-tenth the price.

About the same time, an open-source version of the StarOffice suite — OpenOffice.org — emerged from beta testing, with nearly the same features, but at no cost.

spreadsheet and database work), stick with MS Office for now.

## A bit of history

A few years ago, Sun Microsystems bought StarOffice, then a commercial product, and later decided to release it under open-source rules. Under these rules, anyone may download, view or alter the programming source code behind the application, while the copyright owner maintains its rights.

OpenOffice, StarOffice includes printed documentation and technical support.

## Installation

I have installed OpenOffice and StarOffice on Windows ME, Windows XP Home and Professional editions, and Mandrake, Gentoo and Red Hat Linux. All went smooth as silk.

But there's one huge warning for Windows/MS Office users: During installation you must change a default setting that makes Open/StarOffice the default program for MS Office files. Otherwise, Windows will specify Open/StarOffice as the startup program for these files.

Windows users may also have to download and install the free Java Runtime Environment from Sun Microsystems ([www.sun.com](http://www.sun.com)).

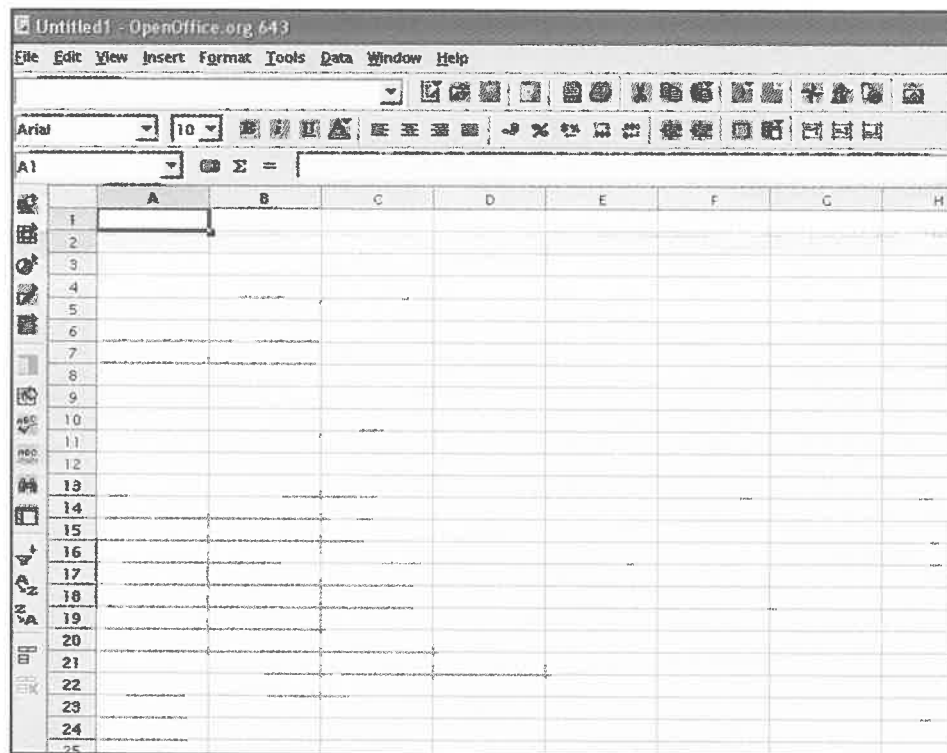
## What's inside?

Both products include a word processor, spreadsheet, presentation program, a database front-end and a drawing program. Of these, I have used the first three extensively, the database front-end somewhat and the drawing program not at all. StarOffice includes a database manager, a stripped-down version of Adabas. But it is limited to databases of 300 MB and allows no more than three concurrent users. These make the program largely useless for CAR. Otherwise the two office suites are nearly identical.

### • Word processor

The people at Sun probably realized that a good word processing application is a must for an office suite. So they obviously spent most of the development on this part of the Star/Open office suite, and it shows.

This is the one application that rivals its MS Office equivalent. I edited large and small files in OpenOffice, including a 100-page book filled with odd formatting, embedding comments and change tracking codes — and OpenOffice handled that flawlessly.



After putting both to the test, I can say that they are very impressive and deliver many of the same features as MS Office. But there are enough small shortcomings — and one critical limit to the spreadsheet — that I'm not quite ready to recommend giving up on MS Office just yet.

For serious computer-assisted reporting (minimal word processing, heavy

The StarOffice code was placed in the hands of an organization of largely volunteer programmers and was renamed OpenOffice.org. Meanwhile, Sun continued to develop StarOffice internally and released it in May.

The OpenOffice project and the Sun developers worked very closely, so it's no surprise that StarOffice and OpenOffice are almost identical. Unlike

Everything imported from MS Word perfectly, with some tiny exceptions. Images were slightly larger than before, and sometimes just a hair off their correct position. MS Word users will find it easy to switch to the OpenOffice word processor because almost every feature in Word can be found in OpenOffice. There are some nice extras in the OpenOffice package that makes working in that environment easier.

### • Spreadsheet

This application, called Calc, is slightly less polished, and has one huge shortcoming: It only accepts 32,000 rows of data, compared to 65,500 or so in Excel. This is bad for journalists doing CAR, but not necessarily fatal. If you import a spreadsheet that is too long, it will simply lop off all rows past number 32,000 and tell you so.

Beyond that, there is a lot good about Calc. Excel users will feel right at home. You'll find similar features used in CAR: string functions, pivot tables (called "datapilots") and charts.

It also has some nice extras. For example, you can do a three-level filter in Calc, compared to two levels in Excel. You also can insert graphics, form elements and other automation tools right into the sheet.

Connecting the spreadsheet to an existing data source (such as an external Excel spreadsheet or .dbf table) is also very easy. Like the word processor, Calc imports data seamlessly from any version of MS Office, right through version XP.

One annoyance: Unlike Excel, Calc does not allow you to select non-adjacent cells using the control key.

### • Database manager

Good news and bad news here. The good news: Connecting OpenOffice to an external data source is an easy, wizard-driven process. Once you connect you can manipulate the data through a number of means. You can use the database to create a merge file with your word processor. You can manipulate the

numbers in Calc. Or you can send SQL commands directly to the data source.

OpenOffice also offers a nice front-end that looks and works very much like the Access query by example grid, though this feature is buried. You can link tables and send queries, just like you do in Access.

Now the bad news: You can't connect to an Access database directly. You can connect via ODBC, but that is slower and somewhat limits what you can do.

In addition, it seems you are restricted to select queries via the OpenOffice grid, which limits how much crunching you can do without writing SQL. And if you're going to write SQL, you might as well use a database application with some power, such as MySQL or Postgres.

This is the one part of the OpenOffice suite that has changed very little since Sun's purchase of StarOffice. Unfortunately, it is the one program in the suite that journalists doing CAR would care about most.

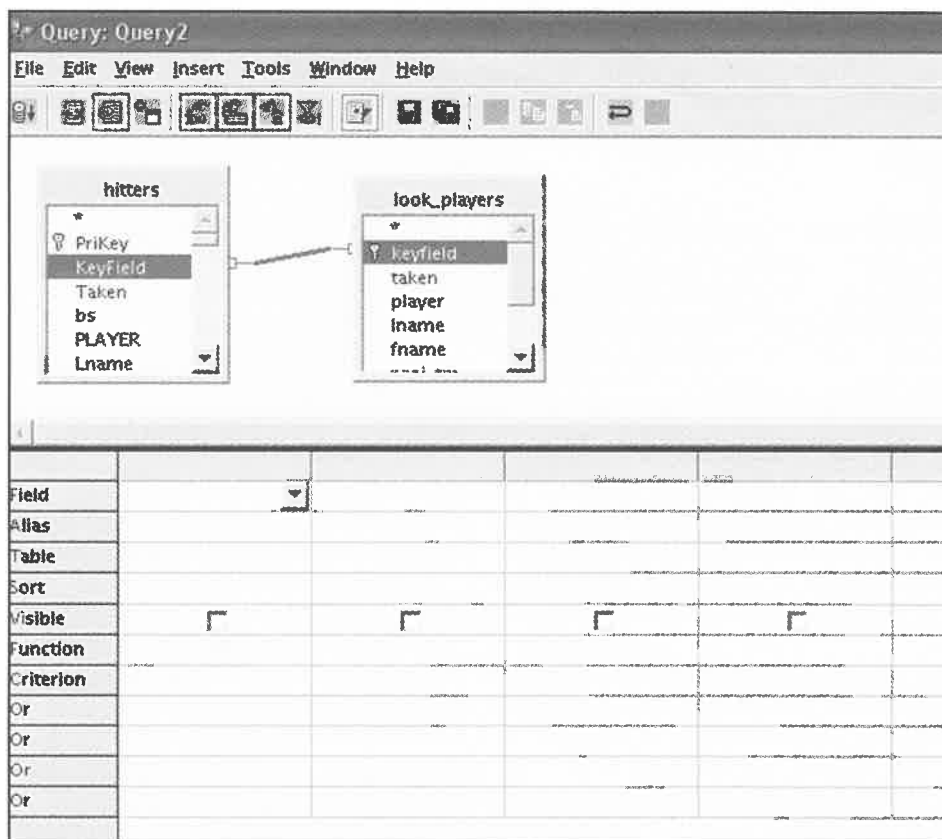
### Recommendation:

Those who aren't able to afford any office suite software would do well to download OpenOffice at [www.openoffice.org](http://www.openoffice.org). It does just about everything as well as MS Office – with a few exceptions.

For users stuck with an older version of MS Office (Office 95 or earlier), the OpenOffice programs would provide significant improvements, except for the database manager.

Even older versions of Access are superior to the database manager in StarOffice. But, for CAR power users, there are plenty of other excellent open source database applications that run on Windows, Mac or Linux. MySQL rivals Oracle in terms of power and speed. The typical journalists doing CAR will find that the simplicity of Access cannot be beat, and users of Office 97, 2000 and XP should probably stick with what they have, at least for now.

Contact Aron Pilhofer by e-mail at [apilhofer@publicintegrity.org](mailto:apilhofer@publicintegrity.org).



# Implants

continued from page 1

So when editors asked them to propose a project with a national scope, they decided that it would be worth digging beyond these scattered stories to take an in-depth look at problems with medical devices and the government's regulatory role.

Assembling information and looking for sources, they began talking with investigative editor Tom Curran about possible data analysis and brought me onto the project.

## Data entry time

Our first goal was to acquire a database of FDA product recalls, and we were quick to learn that the agency had no such database.

The FDA rarely orders recalls. Instead, manufacturers voluntarily recall their products and the FDA publishes a list of voluntary recalls on its Web site.

If you want to find out if a product has been recalled, you can search. But if you want to find out how many heart valves have been recalled, which manufacturers have the most recalls, what reasons are most often given for recalls, you're out of luck.

That is, unless you create your own database, which we did. Orr spent weeks going through the recall notices, pulling out recalls of implants, and capturing vital information about each product and problem.

So this was a database built painstakingly by hand. But it was worth it. We found that the FDA had published recall notices for 573 implant devices since 1992 – far more than the handful that received national attention.

We also found the number of recalls to be on the rise. From 1992 through 1998, an average of 40 implanted medical devices were withdrawn per year. That number jumped to 96 in 1999, 79 in 2000, and 117 in 2001.

## MAUDE matching

Next we wanted to match our recall data to the FDA's Manufacturer and User Facility Device Experience Database. MAUDE, for short, is a key component of the FDA's "aftermarket surveillance" program.

The IRE & NICAR Database Library makes the MAUDE database available to news organizations. News organizations may purchase the data for the entire United States on CD. For more information see [www.ire.org/datalibrary/databases/maude/](http://www.ire.org/datalibrary/databases/maude/). Also see "Digging into health, medicine" on page 1 of this issue of *Uplink* for more information about how MAUDE has been used in news reporting.

Doctors, distributors and manufacturers who become aware of cases in which a device may have malfunctioned or otherwise injured a patient are supposed to fill out an incident report and, ultimately, manufacturers are required to forward any incident reports they receive to the FDA.

We learned from the outset that MAUDE is incomplete. Not every doctor fills out an incident report each time there is a problem with a medical device.

Also, as we discovered in our reporting, manufacturers don't always report what they know. The FDA doesn't have the staff to conduct as many inspections as they'd like, but when they do get around to auditing the companies they regulate, they sometimes find stacks of incident reports that had never made it into MAUDE.

We knew that going into it. What we didn't realize is just how much of a bear MAUDE can be to work with. We bought the data on CD from NICAR, and if there were two words to describe MAUDE from a technical perspective, they would be big and messy.

How big is big? If you buy the disks, be prepared to process about 1.5 GB of raw data, which will swell up to more

than 2 GB once you create field indexes. The database is composed of interlocking tables, including one that contains a narrative of the incident. It took Microsoft SQL Server all night to index that one.

## Messy data

Once you get past that, then you have to deal with the mess. No, the data is not badly formatted. It's just poorly reported. You'll find manufacturers identified 50 different ways in the database. You'll find the same product identified 20 different ways.

We wanted to get an idea of how many deaths involving implantable devices were reported to the government, and what kinds of devices were most often cited in these death reports. Fortunately, each incident record contains a three-character product type code that can be joined to a lookup table containing more than 5,000 types of medical devices.

Unfortunately, the coding system seems to have been devised at random. The letters in the code have no bearing on the type of product. Products related to artificial hips don't all begin with H, and implantable products don't all begin with I.

We got a little outside help here. A source at ECRI, a nonprofit medical device research organization, sent us a list of what they consider implantable medical devices. It was tied to their own coding system, but by eyeballing their list and eyeballing the FDA list, we were able to create a table of FDA product codes for implants. (Our ECRI source was gracious enough to review this list for us).

After running the list of implants against the MAUDE tables, we were able to report that since 1991, manufacturers, doctors and hospitals filed more than 155,000 adverse event reports involving 422 types of implants. Included in these were 2,333 cases in which a patient died.

These are big numbers, and they're far

from perfect. For one, MAUDE reports only include incidents where devices may have caused a problem – there's no requirement for each incident in the database to be investigated.

Moreover, the entire question of cause of death is tricky. Many patients who receive certain kind of implants are dying to begin with and recognize the risk of their procedure – yet in the database, a death is a death is a death.

But the numbers did accurately reflect what was reported to the government, and in the context of all of our reporting, served to make the story more factual and more interesting.

Read "Foreign Objects" online at [www.nj.com/special/projects/implants/](http://www.nj.com/special/projects/implants/).

Contact Robert Gebeloff by e-mail at [rgebeloff@starledger.com](mailto:rgebeloff@starledger.com).

## Charlotte

*continued from page 1*

journalists who gathered to share their knowledge – although with a bit more conflict. For example, some panel topics pitted one type of software against another (database managers vs. spreadsheets!) instead of looking at how all these tools work together. (Ironically enough, we also had to move at the last minute in Raleigh from the hotel to the nearby convention center because of a hotel snafu.)

The conference, composed of more than 50 panels and workshops, begins on Thursday afternoon with small roundtable sessions for those who have been practicing CAR for years. These workshops will look at more complex topics: intranet services for the newsroom, the latest in programming and data cleaning, the possible uses of social networking software, and new advances in GIS (mapping) software.

Co-hosted by *The Charlotte Observer* and NBC News and with major spon-

sorships by the Gannett Foundation and Knight-Ridder, the conference will offer continual hands-on training classes. The classes include sessions for both the beginner and advanced practitioner in effectively using the Internet, spreadsheets, database managers, statistics and mapping.

The panels will, as always, cover specific beats and showcase the best stories of the past year. Unlike previous years, the planners have sketched out a recommended path for those new to CAR, sort of a mini-boot camp.

Friday morning will open with an overview of CAR stories and techniques when just starting out. Most examples will involve smaller datasets that are accessible on the Web, have been built by other reporters, or are slices of data that can be quickly obtained from the IRE and NICAR database library.

The Friday panels for beginners will

### NICAR SERVICES

## Data Library tackles specialty data jobs

By Jeff Porter, IRE and NICAR

The IRE and NICAR Database Library has looked at every block group in the United States and every speeding ticket in Cleveland. Those are just two of the processing or data analysis jobs the Database Library has completed during the past few months.

Under contract with The Associated Press, NICAR downloaded Census 2000 Summary File 3 data files, transformed the raw text files into dBASE IV files and added some geographic elements to every table. NICAR's work for the AP began in March 2001, when the redistricting Census data became available and continued through the Summary File 1 and the SF3 releases.

NICAR analyzed a year's worth of

speeding ticket data from the state of Ohio and the city of Cleveland. For WEWS-Cleveland, we discovered the most likely places to get a ticket, found the officers who wrote the most, and discovered how much faster than the speed limit a driver can go before getting a ticket.

Transcripts of the stories can be found at [www.newsnet5.com/theinvestigator1781686/detail.html](http://www.newsnet5.com/theinvestigator1781686/detail.html) and [www.newsnet5.com/theinvestigator/1784946/detail.html](http://www.newsnet5.com/theinvestigator/1784946/detail.html).

For a nominal cost, the Database Library can conduct data analysis and transfer data from tape to CD for news organizations. For more information, call 573-884-7711.

Contact Jeff Porter by e-mail at [jeff@nicar.org](mailto:jeff@nicar.org).

offer an overview of CAR, including stories and techniques from the most frequently covered beats, and how to get good data from the Web. In the afternoon beginners can go to hands-on classes to learn how to do simple analysis with spreadsheets. Some of the data will be from the Census.

On Saturday morning, the beginners' track will discuss how to weave CAR into a beat, how to use CAR techniques to investigate municipal government and using Census data in everyday reporting. The hands-on sessions that afternoon will focus on learning these skills.

Sunday morning panelists will talk about getting a story off the ground, and offer story ideas and organizing and writing tips. The track will finish with a hands-on session to walk through importing data to analysis to writing a lead.

Contact IRE Executive Director Brant Houston by e-mail at [brant@ire.org](mailto:brant@ire.org).

# Medicine

continued from page 1

Journalists were already on the story, using databases to examine faulty medical products and shortcomings in patient care. A variety of data sources can help with big-picture analysis and breaking public health news, whether it's suspicious hospital deaths or a product recall.

When Michael Berens investigated the impact of errors by nurses for the *Chicago Tribune* (Story No. 17136 in the IRE Resource Center), he found that at least 1,720 patients had died and more than 9,500 had been hurt due to registered nurses "besieged by cuts in staff and other belt-tightening in U.S. hospitals" from 1995-2000.

A major resource for the investigation was the Manufacturers User and Device Experience Database (MAUDE). This database from the U.S. Food and Drug Administration contains reports about injuries, deaths and accidents related to medical devices. It also contains information about the malfunctioning devices.

*The Star-Ledger* recently used MAUDE in its investigation of faulty medical devices. (See "Databases show implant troubles" on page 1).

One of the most useful features for journalists is a memo field detailing the incidents. It also contains references that can be used to request detailed records for each report.

It takes patience to cull through MAUDE, because incident reporting is voluntary and thus prone to gaps. Data entries also are inconsistent. Journalists using MAUDE can examine specific products or manufacturers, as the *Tallahassee Democrat* did for its report about automated drug pumps linked to accidental overdoses (Story No. 17311).

MAUDE is available from the IRE and NICAR Database Library. For more detailed information about the data see

[www.ire.org/datalibrary/databases/maude/](http://www.ire.org/datalibrary/databases/maude/).

Berens also used a database of investigative records from HCFA (now known as the Center for Medicaid and Medicare Services, or CMS) and a database of disciplinary records collected by the National Council of State Boards of Nursing ([www.ncsbn.org](http://www.ncsbn.org)), obtained from the state licensing agency and scanned into electronic form.

Access to health-system data varies by state. For example, *Wilmington (N.C.) Morning Star* reporter Cory Weiss knew that local hospitals would not share their internal investigation into rising staphylococcal infection rates. Instead, the state's death certificate database provided a window on staph-related deaths (Story No. 16583).

Do some research to find out what your state regulates, what kinds of reports the statutes specify, and whether any records exist as databases. Or, if the state yields little, find local records collected nationally by the CMS.

When planning a project, keep in mind that the medical profession has its own rigorous scientific standards for evaluating performance, and the media's methods need to be stringent. Many news organizations consult with independent experts. One major concern is making an accurate comparison among patient outcomes because of all the variables that affect risk in a diverse population with many different procedures and diagnoses. The *Fort Worth Star-Telegram*, for example, worked with a research scientist to analyze patient records (see article on page 8).

## Here are some available data sources:

- Birth and death records.
- Inspection reports, mandated by state law or under the Medicare program.

- Health maintenance organization financial data plus information on participating facilities and patient complaints.

- Uniform billing records. These databases describe each patient's illness, any surgeries, age, insurance type, town or ZIP code and are available from state agencies. Everything will be anonymous, but you can ask for a unique patient identifier as a substitute for names.

- Medicare billing data can be obtained from CMS. Check with your local branch office for details on these requests; private companies often maintain and provide the records.

- For more information on data sources, see tipsheet Nos. 1710, 407 and 643 in the IRE Resource Center.

- The FDA Adverse Event Reporting System contains incidents related to pharmaceuticals and biological products like blood (available from NICAR; for details, go to [www.ire.org/datalibrary/databases/AERS/](http://www.ire.org/datalibrary/databases/AERS/)).

## Other resources can help:

- National Association of Health Data Organizations ([www.nahdo.org](http://www.nahdo.org))

- ECRI ([www.ecri.org](http://www.ecri.org)) is a nonprofit specializing in medical device safety issues.

- The Centers for Disease Control tracks public health data ([www.cdc.gov](http://www.cdc.gov)). See its National Center for Health Statistics Data Warehouse ([www.cdc.gov/nchs/datawh.htm](http://www.cdc.gov/nchs/datawh.htm))

- The National Association of Health Data Organizations, 703-532-3282.

See the January-February 2003 *IRE Journal* for even more about investigating problems with hospitals. Berens of the *Chicago Tribune* details how he used patient billing records to help uncover infections in hospitals. The newspaper reported that three-quarters of the infections could have been prevented. Berens also shares tips for investigating your local hospital.

Contact Megan Christensen by e-mail at [megan@nicar.org](mailto:megan@nicar.org).



## IRE and NICAR Services

Investigative Reporters and Editors, Inc. is a grassroots nonprofit organization dedicated to improving the quality of investigative reporting within the field of journalism. IRE was formed in 1975 with the intent of creating a networking tool and a forum in which journalists from across the country could raise questions and exchange ideas. IRE provides educational services to reporters, editors and others interested in investigative reporting and works to maintain high professional standards.

### Programs and Services

**IRE Resource Center:** A rich reserve of print and broadcast stories, tipsheets and guides to help you start and complete the best work of your career. This unique library is the starting point of any piece you're working on. You can search through abstracts of more than 19,000 investigative reporting stories through our Web site.  
**Contact:** Carolyn Edds, carolyn@ire.org, 573-882-3364

**Database Library:** Administered by IRE and the National Institute for Computer-Assisted Reporting. The library has copies of many government databases, and makes them available to news organizations at or below actual cost. Analysis services are available on these databases, as is help in deciphering records you obtain yourself.  
**Contact:** Jeff Porter, jeff@ire.org, 573-882-1982

**Campaign Finance Information Center:** Administered by IRE and the National Institute for Computer-Assisted Reporting. It's dedicated to helping journalists uncover the campaign money trail. State campaign finance data is collected from across the nation, cleaned and made available to journalists. A search engine allows reporters to track political cash flow across several states in federal and state races.  
**Contact:** Brant Houston, brant@ire.org, 573-882-2042

**On-the-Road Training:** As a top promoter of journalism education, IRE offers loads of training opportunities throughout the year. Possibilities range from national conferences and regional

workshops to weeklong boot camps and on-site newsroom training. Costs are on a sliding scale and fellowships are available to many of the events.

**Contact:** Ron Nixon, ron@nicar.org, 573-882-2042

### Publications

**The IRE Journal:** Published six times a year. Contains journalist profiles, how-to stories, reviews, investigative ideas and backgrounding tips. *The Journal* also provides members with the latest news on upcoming events and training opportunities from IRE and NICAR.

**Contact:** Len Bruzzese, len@ire.org, 573-882-2042

**Uplink:** Bimonthly newsletter by IRE and NICAR on computer-assisted reporting. Often, *Uplink* stories are written after reporters have had particular success using data to investigate stories. The columns include valuable information on advanced database techniques as well as success stories written by newly trained CAR reporters.

**Contact:** David Herzog, dherzog@nicar.org, 573-882-2127

**Reporter.org:** A collection of Web-based resources for journalists, journalism educators and others. Discounted Web hosting and services such as mailing list management and site development are provided to other nonprofit journalism organizations.

**Contact:** Ted Peterson, ted@nicar.org, 573-884-7321

### For information on:

**Advertising:** Pia Christensen, pia@ire.org, 573-884-2175

### Membership and subscriptions:

John Green, jgreen@ire.org, 573-882-2772

### Conferences and Boot Camps:

Ev Ruch-Graham, ev@ire.org, 573-882-8969

### Listserves:

Ted Peterson, ted@nicar.org, 573-884-7321

### Mailing Address:

IRE, 138 Neff Annex, Missouri School of Journalism, Columbia, MO 65211

## Computer - Assisted Reporting Boot Camps

These unique seminars give journalists a jumpstart in computer-assisted reporting techniques. Participants are trained in how to acquire electronic information, use spreadsheets and databases to analyze the information and to translate that information into high-impact stories. The National Institute of Computer-Assisted Reporting provides follow-up help when participants return to their news organizations.

- March 23-28 – Columbia, Mo.
- May 18-23 – Columbia, Mo.
- Aug. 3-8 – Columbia, Mo.

### What participants have said about IRE and NICAR Computer-Assisted Reporting Boot Camps:

*"The workshop and the conference have convinced me that the investigative reporting approach and techniques can be easily applied to beat reporting and daily journalism."*

— Afi-Odelia Scruggs, Professor of Journalism at Ohio Wesleyan University

*"Overall this is a wonderful seminar. This is a great start for working with CAR for someone who came with no experience."*

— Anonymous (from seminar evaluation)

*"Well worth the money!"*

— Anonymous (from seminar evaluation)

**More information is available at**  
**[www.ire.org/training](http://www.ire.org/training)**

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

NON-PROFIT ORG.  
U.S. POSTAGE  
**PAID**  
Jefferson City, MO.  
Permit NO. 89



Ira Chinoy  
Philip Merrill College of Journalism  
1117 JOURNALISM BLDG  
UNIVERSITY OF MARYLAND  
COLLEGE PARK MD 20742-0001

## Uplink Info

A newsletter of the National Institute for Computer-Assisted Reporting

### Director of Publications

Len Bruzese  
len@ire.org

### Advertising Coordinator

Pia Christensen  
pia@ire.org

### Subscription Administrator

John Green  
jgreen@ire.org

### Subscriptions

IRE members \$40, nonmembers \$60

### Uplink Address:

IRE-NICAR, 138 Neff Annex  
Missouri School of Journalism  
Columbia, MO 65211

Postmaster: Please send  
address changes to IRE-NICAR.

### NICAR is a joint program

of Investigative Reporters and

Editors Inc. and the Missouri

School of Journalism.

NICAR services include

supplying journalists with

government databases,

training programs, tipsheets and

data analysis.

### Editorial

573-884-7711

Subscriptions  
573-882-2772

### Editor

Brant Houston  
brant@ire.org

### Managing Editor

David Herzog  
dherzog@nicar.org

### Asst. Managing Editor

Jeff Porter  
jeff@nicar.org

### Art Director

Lisa Tiefenbach

### Copy Editor

Pia Christensen

### Contributing Editors

Megan Christensen

Jaimi Dowdell

Holly Hacker

Stephanie Kang