Uplink

May • June 2004
Volume 16, Number 3

Published bimonthly by the National Institute for Computer-Assisted Reporting

www.nicar.org

FIRST VENTURE

Finding racial tilt in arrests

By Dave Moore
Columbia (Mo.) Daily Tribune

For more than a decade, state and federal government agencies have issued reports and held meetings about the disproportionate number of minorities in jails and prisons.

But by the time people land in jail, they're often too far gone for help. That's a heartbreaking notion, especially when you're talking about teenage children.

When Congress passed the Juvenile Justice and Delinquency Prevention Act of 2002 it intended to address the arrest imbalance for minority-group juveniles by reducing the number of children who get in trouble with the law. For that reason, I decided to look at the racial composition of the children arrested in Columbia by analyzing the police department's arrest data, using the lessons I learned at a recent NICAR Boot Camp.

I found the proportion of minor-

continued on page 18

PEACE CORPS

Analysis exposes attacks on volunteers

By Russell Carollo, *Dayton Daily News*

The idea wouldn't go away.

For more than 10 years, I have routinely scanned inspectors' general reports to Congress for database and story ideas. Thick reports from these internal government watchdogs about large agencies get nearly all of my attention. Still, I usually scan shorter reports from smaller agencies such as the Federal Maritime Commission and Office of Personnel Management, just to make sure I have looked at them all.

Peace Corps reports were the very last places I thought I would ever find tips about computer databases for a major project. But beginning about four years ago, I kept noticing vignettes about rapes and other attacks on volunteers.

These brief, graphic descriptions told of violent attacks, mostly on young women serving overseas. As a citizen, I was shocked this was going on at all; as a journalist, I was sure others would be shocked, too.

continued on page 16

SPOTLIGHT: ELECTIONS

Using CAR to cover the vote

By Megan Clarke, *IRE and NICAR*

It's that time again. Election 2004 is big news already and many newsrooms are preparing to use computer-assisted reporting for stories that go beyond the basics of who-donated-how-much campaign finance coverage.

To get you thinking how you can use CAR between now and election day, here are several examples of how journalists have used spreadsheets, database analysis and mapping to report on voters and election results.

St. Paul Pioneer Press CAR Editor Janet Roberts reported on how particular precincts and types of voters can

continued on page 22

SPOTLIGHT:

For more on election coverage see:

- Federal data helps unravel campaign finance, p. 3
- Getting and analyzing state finance data, p. 3
- Recent election stories using CAR, p. 22

Bits & Bytes

Updated audits files

The IRE and NICAR Database Library recently updated the Single Audit database from the Federal Audit Clearinghouse of the U.S. Census Bureau. The database, which covers 1997 to January 2004, contains information on expenditures of federal awards greater than $300,000 a year. Journalists will find this useful for looking at federal grants used by nonprofits and public agencies.

This data contains more detail than is provided by the federal Consolidated Federal Funds Report and the Federal Award Assistance Data System. The audit data looks at spending and includes audits of indirect recipients.

For more information about these databases *see www.ire.org/datalibrary/databases.*

Sex offenders

KIRO-Seattle in February aired a series of stories about sex offenders living in or close to day-care centers in Washington. The IRE and NICAR Database Library conducted the analysis for KIRO, using ESRI ArcView to map the registered home addresses of sex offenders and the addresses of licensed day care centers. The analysis identified sex offenders living within one city block of the centers. Read the story and view maps online at *www.kirotv.com/investigations/2825037/detail.html.*

To learn more about how IRE and NICAR can help your news

continued on page 4

INSIDE NICAR

Discovering CAR, again and again

By David Herzog, *NICAR and Missouri School of Journalism*

One of the benefits of putting *Uplink* together six times a year is getting to see all the great stories journalists are doing using computer-assisted reporting on the beats and on the big projects.

There's so much going on and so much to get excited about as journalists discover the power (and even fun) of analyzing data for news stories.

As this issue of *Uplink* goes to the printer, hundreds of journalists from television, newspapers, magazines – you name it – will be gathering in Atlanta the day before the IRE Annual Conference officially begins to pick up CAR tips and story ideas.

Aside from the informative panels, the optional CAR day provides a great place for newcomers to hang out with the longtime CAR jockeys and pick up the tricks of the trade.

It's always edifying to see the informal learning that goes on during the CAR day, and how many of the newcomers leave for their newsrooms back home, excited about the possibilities.

Then, as those journalists use CAR in their reporting and turn out some fantastic stories, all of us get to discover the power of computer-assisted reporting all over again.

Contact David Herzog by e-mail at dherzog@nicar.org.

Upcoming hands-on CAR training

IRE and NICAR have several training opportunities for journalists seeking hands-on instruction in using computer-assisted reporting.

This summer marks the last chance this year for in-depth CAR training. There is one remaining CAR Boot Camp in Columbia, Mo., for journalists who want to learn how to acquire electronic information, use spreadsheets and databases to analyze the information and to translate that information into high-impact stories. IRE and NICAR provide follow-up help after participants return to their news organizations.

The Boot Camp is Aug. 1-6. The Boot Camps will be offered again four times next year, along with other specialized workshops.

Journalists interested in learning how to use geographic information systems (GIS) in their reporting can register for the Mapping Data for News Stories mini-boot camp Aug. 20-22 in Columbia, Mo. The sessions will include address geocoding, thematic mapping, overlaying and spatial analysis.

Those who attend the mapping mini-boot camp will be eligible to purchase GIS software at a discount. In many cases, the cost of the training session and discounted software is less than the full price of the software alone.

For more information on the mapping workshop and other training opportunities offered by IRE and NICAR visit *www.ire.org/training.*

SPOTLIGHT: ELECTIONS

Getting a handle on federal campaign cash

By Aron Pilhofer, *The Center for Public Integrity*

There was a time when tracking campaign cash was a straightforward affair. But as a result of actions by Congress and the courts, and the endless creativity of political leaders over the last decade, the money trail has never been more twisted and difficult to follow.

That's particularly true in this election, the first since passage of the Bipartisan Campaign Reform Act of 2002.

Gone are the days when journalists covering state races could blissfully ignore the federal system and vice versa. There is no such thing as a purely federal or purely state election any more.

Both parties, along with a growing number of outside interest groups, have devised sophisticated national strategies and fund-raising networks to put money into races at all levels, from the state house to the White House.

The bulk of campaign cash continues to flow through traditional fund-raising sources, such as party committees and PACs. But a significant amount of money comes under the radar, most often through political nonprofit organizations that do not disclose their activities in the usual places.

Over the past four years, the Center for Public Integrity has tracked at least $500 million in activity by these groups, including just over $100 million in the past 15 months alone. These groups do not disclose their activities to any state or federal election authorities, if at all. This makes them ideal conduits for contributors who wish to have an impact on an election with as little fanfare as possible.

As an unfortunate result, journalists today must be just as sophisticated as those they are covering. They must play the part of attorney, accountant, politico and sleuth to get a clear picture.

Fortunately, there is a wealth of data available either in a downloadable, electronic format or searchable on the Web that can help journalists paint a true picture of who gives to whom and why.

Investigative Reporters and Editors is developing workshops on how to use campaign finance data to report on these political nonprofit organizations.

Downloadable data

Raw campaign finance data for most federal races can be downloaded from two Federal Election Commission Internet sites: the electronic filing site (*http://herndon2.sdrdc.com/dcdev*) or its FTP server (*www.fec.gov/finance/newftp1.htm*). The exception is Senate candidates and party caucus committees (specifically, the Democratic Senatorial Campaign and National Republican Senatorial committees), which are exempt from filing electronically. The only source for electronic data for these Senate candidate and party committees is from the commission's FTP site.

Both sources have their pluses and minuses. The advantage to the electronic filing site is speed and convenience. Reports appear almost as soon as they are submitted by the re-

continued on page 20

SPOTLIGHT: ELECTIONS

Finding data for state elections

By Derek Willis, *The Center for Public Integrity*

State elections account for hundreds of millions of dollars in activity every two years, but tracking the money that fuels those races can be a cumbersome process. Nearly every state has its own method of reporting campaign finances and most don't offer the same kind of data that's available at the federal level.

State data comes in a variety of formats, but electronic access has been increasing lately. Some states provide data in electronic format or display the records on a Web site in a manner that could be imported into a database, while other states provide filings in Adobe Portable Document Files or other image files, or on paper.

During a recent project on state political parties, the Center for Public Integrity found at least 17 states that offered campaign finance data in a database, HTML or ASCII text format. Among the best are Ohio, which has comma-delimited text files for individual candidates and for all party organizations, and New York, which offers a single file containing contributions and expenditures during a given filing period. Washington, Idaho and Colorado also provide downloadable files.

Several other states are going the route of the searchable Web database, where a user can enter search terms or date ranges and get back text files that can easily be put into Microsoft Excel or Microsoft Access. States such as Florida and Maryland do this. Others offer data on a filing-by-filing basis – not exactly the best when you want to download everything – but it's better than nothing. Penn-

continued on page 21

Bits & Bytes

continued from page 2

organization with data analysis projects, contact the Database Library at 573-884-7711.

See page 6 of this issue of *Uplink* to see how WMAR-Baltimore mapped the locations of sex offenders. The May-June *IRE Journal* also has several articles focusing on how to report on child sex offenders.

Campus crime data

New data from the IRE and NICAR Database Library can help journalists find campus crime stories. The Clery Act Campus Crime data for 2001 and 2002 contains 19 tables with information about crimes reported to campus police or local law enforcement agencies.

Offenses such as homicide and sex crimes are included even if they were not prosecuted. There are separate tables for hate crimes and arrest reports, which include totals for liquor and drug law violations. This data is compiled with information campuses are required to publish and distribute as an annual campus security report.

The *Rocky Mountain News* in Denver recently used this data to draw attention to drug and alcohol violations at the University of Colorado at Boulder. Read the story on the Web at *www.rockymountainnews.com/drmn/local/article/0,1299,DRMN_15_2712250,00.html.*

To order the campus crime data, call the Database Library at 573-884-7711 or 573-884-7332.

TRANSPORTATION

Fatal crash data shows truck dangers

By Kurt Rogahn, *The (Cedar Rapids, Iowa) Gazette*

Fatal accidents on Interstate 80 in Iowa are seen in a different light these days.

Here's how another reporter at the newspaper I work for closed a routine news story about an accident that killed two people in a four-vehicle accident that include a semitrailer truck:

"A *Gazette* investigation published in November showed that nearly half the fatal accidents on I-80 from 1994 through 2001 involved semi-trailer trucks. Of the 133 fatal accidents, 62 involved semis. In all, 84 people died in the crashes involving semis."

It was rewarding to see that paragraph.

"Danger Zone," the package of stories it referred to, resulted from months of research, database analysis, more than a few dead ends and a fair amount of uncertainty. The package included stories from KCRG, our affiliated television station, and an online package with animated graphics.

The stories explained how increasing traffic volumes, basic physics and the way drivers of cars behave behind the wheel all contribute to fatal accidents involving semis. Big trucks can't stop as quickly as a car, nor can semi drivers steer out of trouble.

Some people have told me they've changed their driving habits. That's even more rewarding than seeing the conclusions of "Danger Zone" repeated in a short story about a fatality.

My reporting started with a suggestion from a retired *Gazette* reporter who had just returned from a long trip: Look into the way cars and trucks interact on the road. He thought we might learn something from truck inspection reports. But that wasn't as helpful as examining FARS, the federal government's Fatality Analysis Reporting System.

The *Gazette* bought the data covering 1994 to 2001 on CD-ROM from the IRE and NICAR Database Library. I set out to find a story – whatever it might be.

I began by examining outdated reports based on FARS from the Insurance Institute for Highway Safety, and localizing them for Iowa. I looked at how other news organizations had studied cars, trucks and roads. I tried to get trade groups and federal agencies to put me in touch with their researchers, to get coaching and to test some of the queries I created in Microsoft Access.

One of those queries turned up a hot spot with more fatal accidents than any other part of Interstate 80 in Iowa: rural Cedar County, about 40 miles southeast of Cedar Rapids. County traffic volumes have increased to 33,000 vehicles daily from 22,000 in a little more than a dozen years. Every third vehicle there is a heavy truck.

My Internet searches led me to a question about highway traffic deaths that I thought I could resolve: Who's to blame for fatal crashes involving trucks and cars? The *Los Angeles Times* reported that truckers are at fault in about half of all California large-truck crashes. Yet an American Automobile Association study fixes more blame on automobile drivers. Advocacy groups fall on both sides of the argument.

My contact with other journalists across the country led me to Daniel Blower of the Transportation Research Institute at the University of Michigan. He warned that the increase of accidents in Cedar County – 15 of 19 fatal accidents on I-80 in eight years had in-

continued on page 12

IRE AWARDS

Broadcast, other winners rely on data work

By Megan Clarke, *IRE and NICAR*

IRE received 595 entries this year in its annual investigative reporting contest. Many of the entries used a wide variety of computer-assisted reporting techniques, including building original databases and using software programs to track shipments by sea. Here is an overview of television, other media and special category winners and finalists that used CAR. See the May-June *IRE Journal* for a full list of winners.

You can order copies of these stories from the IRE Resource Center. Contact the center at 573-882-3364 or rescntr@ire.org.

Television

Network-Certificate

• PBS Frontline/New York Times Television/Canadian Broadcasting Corporation: "A Dangerous Business" by Neil Docherty, David Rummel, Lowell Bergman, Linden MacIntyre, David Barstow, Robin Stein, James, Sandler, Nelli Kheyfets, Remy Weber, Jason Maloney, Lynda Baril, Louis Wiley Jr., David Fanning, Ann Derry and Lawrie Mifflin.

This hour-long documentary, aired in conjunction with a series of articles in *The New York Times*, revealed that cast iron pipe-maker McWane Inc. had accumulated more safety violations than all of its major industry competitors combined. The investigation included an analysis of Occupational Safety and Health Administration (OSHA) and Environmental Protection Agency (EPA) data that uncovered dangerous working conditions.

The documentary also has been awarded a Columbia duPont Award, a Goldsmith Prize for investigative reporting, the George Polk Award and a Peabody Award.

Network-Finalist

• Primetime Thursday, ABC News: "Nuclear Smuggling," by Brian Ross, Rhonda Schwartz, David Scott, Yoruba Richen and Gerilyn Curtin.

On the two-year anniversary of the Sept. 11 terrorist attacks, Primetime Thursday aired a story that revealed gaps in U.S. port security. As a shipping client, ABC was able to track a uranium shipment using the tracking system on the shipping company's Web site. The tracking system was crucial in informing ABC when the shipment passed certain ports.

This report also won second place in the 2004 National Headliner Awards for investigative reporting and the 2004 Excellence in Media Silver Angel Award.

Top 20 Market-Certificate

• KMGH-Denver: "Honor and Betrayal: Scandal at the Academy," by Jeff Harris, John Ferrugia, Kurt Silver, Jason Foster and Byron Grandy.

KMGH's investigation exposed the U.S. Air Force Academy system that deterred and even punished female cadets for reporting sexual assaults. This 10-month project yielded more than 100 stories, including a one-hour documentary, and prompted several investigations and the removal of Academy leadership.

The station created its own database in Microsoft Excel of current and former female cadets in order to research and locate victims.

The investigation also won a Columbia duPont Award, a Peabody Award, the National Headliner Award - Best of Show, a Society of Professional Journalists award, and a CINE award.

Top 20 Market-Finalist

• WFAA-Dallas: "State of Denial," by Brett Shipp, Mark Smith and Kraig Kirchem.

WFAA's seven-month investigation exposed Texas' crippled workers' compensation system. In many cases, state regulators failed to enforce labor laws and overlooked insurance companies that unjustly denied benefits to injured workers.

WFAA built an Excel spreadsheet with hundreds of patients' names, insurance carriers, physicians and status of their benefits and medical treatment.

The series also won the Texas AP Broadcasters Award for investigative journalism.

Top 20 Market-Finalist

• WESH-Orlando: "Building Homes: Building Problems," by Stephen Stock, Michelle Meredith, Shannon Hori, Dave McDaniel, Kathy Marsh, Travis J. Sherwin, Jim Payne, Peter A. Delis, Mark Rice and Jason Morrow.

WESH joined forces with the *Orlando Sentinel* to inspect 406 central Florida homes built in 2001. The investigation found that 80 percent of the homes had at least one major building flaw.

The station obtained a database of inspection records from each municipality and used Microsoft Access to analyze its own database of inspectors' daily reports.

The 14-story series also has won a Peabody Award.

Below Top 20 Market-Finalist

• KAKE-Wichita: "Lake Kahola: A City Manager's Ethical Abyss," by Paul Aker and Craig Lee.

KAKE uncovered a series of no-bid consulting contracts authorized by the Wichita, Kan., city manager and issued to his close personal friend and seller of lakefront property. The four bids totaled $40,000 and KAKE found the manager's friend never completed the majority of the work required.

continued on page 10

MAPPING IT OUT

The latest uses of mapping in news reporting.

Spotting sex offenders

By Tisha Thompson,
WMAR-Baltimore

Nobody wants sex offenders in his or her backyard, and most people lack the technology to see if any live nearby. It's easy enough to go to a sex offender registry on the Web, type in an address and see whether any live on your street. But, since most people are not cartographers, they may not be aware of sex offenders living on streets that wind behind their homes or a block away from a school.

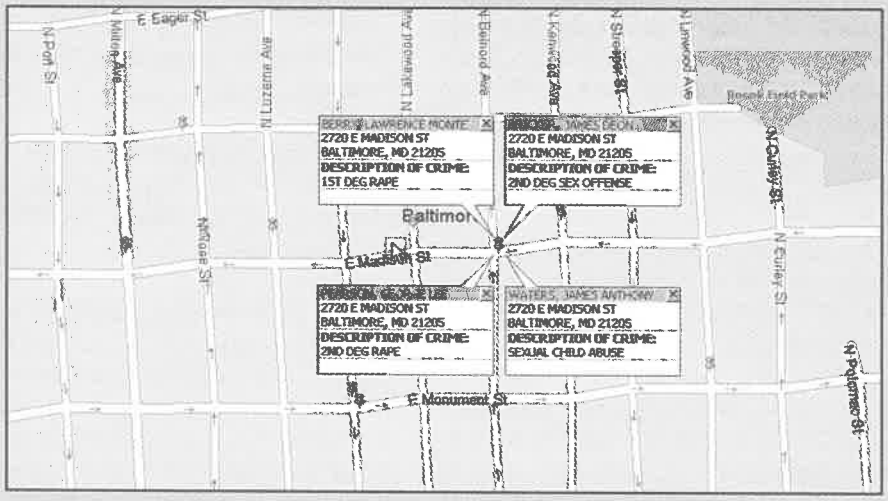
We initially wanted to see if there were any sex offender clusters (large groups of registered offenders living together in the same neighborhood or at the same address) in Baltimore. We filed a public information request with Maryland's Department of Public Safety and Correctional Services for a CD-ROM containing the state Sex Offender Registry. (See the May/June *IRE Journal* for more about reporting on sex offenders).

Even though that information is readily available on the Internet in a Portable Document Format file, the department initially didn't want to give us the information. It claimed that because the registry changes every day, any information burned onto a CD would immediately be outdated. But when I responded that I wanted to map the information, the department suddenly became super helpful. "Our computer guys want to know what you can do with the data," they said. And I promised to verify the address of any sex offender we specifically referred to in our story.

I imported the comma-delimited text data into Microsoft Access and Excel. I used Access to find the clusters by sorting and grouping and counting specific addresses and ZIP codes. I used Excel to import the data into MapPoint, a Microsoft mapping program.

We bought MapPoint instead of ESRI ArcView because it's less expensive ($299 list price) and easier to use. It took me about 10 minutes to somewhat master the software. MapPoint imported the more-than 3,500 sex offender rows from Excel within two minutes, including name, address and offense of each offender. The software already contains the addresses of places such as police departments, libraries, museums and other public buildings. I asked MapPoint to map all school buildings and quickly saw there were clusters located near school buildings.

The MapPoint address matching process did have some shortcomings. I had to manually approve some addresses because a ZIP code didn't match, a street name was slightly off and so forth. I also noticed a few school names were out of date, even though the addresses were correct. And I realized the software would only mark each location once, so there is no way to tell if several people live at one address without first grouping and summarizing the data.



Our biggest hurdle came when we tried to incorporate the maps into the story. Like many television stations, our graphics department uses Macintosh computers and MapPoint runs on Windows. Our computer guru tried everything he knew, but we were working on deadline. We never were able to figure out how to import the maps I created into the graphics computers.

We ended up simply shooting the PC screen as I zoomed in and out of the maps in MapPoint. Surpris-

ingly, the video ended up being much more effective than any graphic we could have made. (It wouldn't have worked if our photographer hadn't managed to get rid of that annoying computer flicker). We started with a map of Maryland, then zoomed in to show how densely packed the sex offenders are in Baltimore, zoomed in again to a ZIP code that had more than 100 offenders, and then zoomed in once again to a street where almost a dozen offenders lived in the same residence.

The maps showed sex offenders living all around Baltimore.

We did a lot of old-fashioned reporting – verifying addresses, getting court records and finding people. But the red dots on the maps told the story best. Even experienced reporters at our station were surprised at what they saw on the computer screen. Maryland looks like a big, red morass. After zooming in, Baltimore became a glaring red dot. I'm convinced that no single piece of video could have illustrated the statement that "we found more than 100 living in the same neighborhood" as well as the map that showed the dots lining a road.

The maps showed sex offenders living all around Baltimore. Practically everyone in the newsroom – as well as some viewers – asked me to type in their addresses so they could see how many dots were in their neighborhood.

And the department's computer staff? They are trying to figure out if they can incorporate the mapping technology onto the Sex Offender Registry's Web site. All of which convinced me that if you have the right kind of story, maps will not only help research, but dramatically enhance storytelling as well.

Contact Tisha Thompson by e-mail at Thompson@wmar.com.

Tisha Thompson is a graduate of the University of Missouri School of Journalism.

Would you be willing to share a mapping example with fellow journalists? Send an electronic copy of the map along with details to David Herzog at dherzog@nicar.org

EDUCATION

Scholarship program takes from the poor

By Maurice Tamman, *The Atlanta Journal-Constitution*

Georgia's lottery and HOPE college scholarship program recently celebrated its 10th birthday. Until recently, pols have been able to sing its praises by noting that 700,000 Georgians have had their college education funded by the programs and that HOPE has distributed about $2.1 billion.

In a recent *Journal-Constitution*/Zogby America poll, nine out of 10 Georgians applauded, saying HOPE is a good use of lottery dollars.

Despite all the adulation, elected state leaders are faced with the reality that without changes to the HOPE program, the cash paid out to Georgia's college students will likely outstrip lottery sales as soon as 2007.

After the spring legislative session, Gov. Sonny Perdue signed a budget that froze the money students can get from HOPE and tightened how the 3.0 grade point average is calculated.

The budget also limited the number of credit hours for which technical school students can receive HOPE money. But, for the first time, part-time private college students would be eligible for the scholarship.

Wholesale changes to the program, however, could be treacherous. The HOPE program most benefits those in politically powerful and wealthy communities, usually with the best high schools. In many ways the program had become an entitlement for middle- and upper-class, largely white, Georgians.

The challenge for the newspaper in examining the HOPE program was to find data that would measure the programs' successes and shortcomings. These, however, are not heady days for public data-gatherers. State education bureaucrats and lottery officials refused to share any individual student financial and academic information or store lottery sales and winner data.

As a result, we were forced to employ the "dead flesh" method made famous by Victor Frankenstein when building his monster.

When we were done, however, we had documented issues many would have preferred the villagers not gaze upon:

• The HOPE programs are largely subsidized by lottery players living in predominately black and poor ZIP codes.

• HOPE scholarships were awarded in the highest numbers in wealthier, white ZIP codes.

• Some proposed changes to the HOPE program likely will result in fewer poor and minority students qualifying for the scholarship.

• Some high schools had lowered standards so more students initially qualified for HOPE but were entirely unprepared for college academics.

• Less than 40 percent of students could keep their HOPE scholarships past their freshman year. (Generally, a B average qualifies a student for a scholarship.)

• Many of the students who received HOPE would have gone to college regardless of the scholarship. And a majority of HOPE scholars at several of the state's largest universities came from families with income exponentially higher than the cutoff for federal needs-based assistance.

The paper's director of computer-assisted reporting, David A. Milliron, and I split the universe in half. He focused on academic issues, such as grade inflation and how proposed changes to HOPE would alter who qualified as a HOPE scholar. I focused on the economics and social cost of HOPE and the lottery.

The lottery folks gave us data for ticket sales and winners of amounts over $600 aggregated by ZIP code and year since the lottery began. We would have preferred the lottery data by store and winner, but the legislature specifically exempted the lottery from the state open records law.

Having the winners data, however, proved to be critical. We had looked at the sales data in the past and it was impossible to know if sales were driven by those living in a ZIP code or those driving into or through a ZIP code. (ZIP codes on interstates and major roadways, especially those near the state line tended to have disproportionately higher sales.)

Given the millions of tickets sold, the number of winners in each ZIP code seemed to be a fair predictor of how popular the lottery was there.

We married that data with the HOPE awards aggregated by ZIP code from the Georgia Student Finance Commission and Census demographic data.

What we found, unsurprisingly perhaps, is that ZIP codes with the highest number of lottery winners had relatively few HOPE scholars while those with highest number of HOPE scholars had relatively few lottery winners. The lottery-winning ZIP codes were poor, largely black; the HOPE scholarship ZIP codes were significantly wealthier and predominately white.

Consider this:

The ZIP code with the most lottery winners was 7 percent white, 87 percent black, and 6 percent other. Median household income was $36,000 in 2000. The state paid out $1.3 million in lottery winnings. There were 369 HOPE scholarships totaling about $900,000.

The ZIP code with the most scholarship recipients was 79 percent white, 8 percent black, and 13 percent other. Median household income was $71,000 in 2000. The state paid out $391,000 in lottery winnings. There were 1,331 HOPE scholarships totaling $3,850,094.

This income redistribution was best illustrated by a University of Georgia student working at an Athens, Ga., convenience store.

Reporter Matt Kempner wrote: The 25-year-old economics major says the scholarship money he got for four years was nice, but it had little to do with his decision to go to college.

"All the time, moms come in and cash their welfare and child support checks and buy tickets," he says. "Bums hand me a quarter, three dimes, a handful of pennies, whatever they've got on them so they can play. They're just buying a dream."

[The student] says he feels bad for the players and tries to make sure they know they're appreciated.

"I look them in the eye and tell them thanks for paying my tuition."

Some teachers and educational professionals questioned the basic qualification for a HOPE scholarship – a B average – especially for incoming freshman. Once again, examining this with data wasn't easy since the state doesn't release student level grade data. Instead, Milliron looked at aggregate grade results and post-secondary data compiled by the school systems, the College Board, the Board of Regents and Georgia colleges and universities.

At Cobb County's Walton High School and DeKalb County's Towers High School, the average GPA for students headed to state public colleges in 2000 was a B, reporter Patti Ghezzi reported. Walton, however, averaged 1117 on the SATs in 2000 compared to 833 at Towers.

Perhaps predictably, we found Walton graduates had an average GPA of 2.9 at the end of their freshman year. Towers graduates, however, had an average GPA of 2.2.

Clearly, some school systems were softer than others. Looking at data collected at the college level, we found a disproportionate number of DeKalb graduates who needed remedial work in college. For example, at Lithonia High School in Gwinnett County, 41 percent of graduates who landed at in-state public colleges in fall 2001 needed remedial courses. Yet two-thirds of the graduating class had high enough grades to qualify for HOPE.

One of the proposals now being considered to restrict HOPE is connecting HOPE scholarships to SAT scores. By looking at data from a number of metro Atlanta school districts, however, Milliron was able to show that doing so would disproportionately affect poor, often black students.

For instance, 53 percent of DeKalb County graduates reportedly earned a B average in 2002 and thus were eligible for HOPE. But Milliron's analysis showed that if the state were to require a minimum 1000 on the SAT, nearly 48.2 percent of the minority graduates in DeKalb would be ineligible for HOPE in a school district that is 90 percent minority.

Unfortunately, it is impossible to know the race of HOPE recipients because no one collects that data.

In fact, most colleges and universities stopped collecting students' family income information in recent years. (A federal financial aid form is not needed for HOPE applicants.)

For those that did have data, the State Board of Regents, which oversees the state public university system, directed the schools not to give out individual student records, even if they scrubbed the data to completely obscure students' identities.

We battled with the board over this issue early in project. Like petrified villagers, however, the board armed themselves and the targeted schools with the federal Family Educational Rights and Privacy Act (FERPA) and refused to budge.

So we focused on three fairly representative institutions. We asked them to look at the data from the last year they collected income data and aggregate the number of students based on ranges of family income that we defined and the average income in each range. (Almost all schools had stopped collecting this data earlier this decade.)

We found what the ZIP code analysis had suggested: HOPE scholars tended to come from higher income families. During the 1999-2000 school year at Georgia Southern University, for example, HOPE scholars' family income averaged $67,000, about $25,000 more than the state's median income that year and $10,000 more than students who didn't qualify for HOPE; 46 Georgia Southern Hope scholars' families earned more than $250,000 a year, averaging $453,000.

Further, using data from the Georgia Financial Aid Commission we were able to determine that three-quarters of HOPE scholars come from families who either have incomes too high to qualify for federal needs-based aid program or who don't bother to apply for it.

The writers capably and carefully constructed project from these varied data elements and presented reasoned findings.

Even HOPE founder and former Gov. Zell Miller could only muster this criticism in a letter to the editor:

"... You fought the lottery tooth and nail from the first day I mentioned it in 1988. And now, you're going after HOPE like a vulture picking apart roadkill."

Given Miller's choice of words, I think it's time to abandon the Frankenstein analogy before it's too late.

Contact Maurice Tamman by e-mail at mtamman@ajc.com.

Winners

continued from page 5

KAKE used a proprietary database that included a property's appraisal and its address that it used for preliminary information and verification.

Below Top 20 Market-Finalist
• WITI-Milwaukee: Still Drunk, Still Driving," by Bob Segall and Michele Murray.

WITI's investigation found thousands of drunken drivers with multiple driving while intoxicated convictions who were still driving vehicles that should have been seized by law enforcement.

WITI created a database of court information for the repeat drunken drivers. The station also analyzed the number of cases in which judges ordered vehicle seizures.

The report also won a regional Emmy and a Milwaukee Press Club Award.

Other Media

Magazine-Certificate
• *Time*: "U.S. Energy Policy," by James B. Steele and Donald L. Barlett.

A nine-month investigation revealed that the 30-year failure of the United States to adopt and stick with a national energy policy will force the nation to struggle with recurring oil and gas crises in the future.

Using Lotus 1-2-3 and Excel, Barlett and Steele constructed spreadsheets that charted energy trends containing information from oil wells drilled to gallons of oil refined. They also extensively used the Department of Energy's Energy Information Administration, which has an online storehouse of energy data.

Magazine-Finalist
• *The Atlantic Monthly*: "Texas Clemency Memos," by freelancers Alan Berlow and Toby Lester.

After two years of research, Berlow found that George W. Bush, while governor of Texas, had approved the executions of 57 criminals without "carefully" and "thoroughly" reviewing their cases as he publicly promised.

Berlow reviewed hundreds of pages of reports and communications between Bush and his chief legal counsel, Alberto Gonzales, now White House counsel.

Berlow used Microsoft Access to sort through data to find documents relating to executions and to filter out correspondence to Gonzales.

Radio-Certificate
• WNYC-New York Public Radio: "Handshake Hotels," by Andrea Bernstein and Amy Eddings.

WNYC's seven-month investigation uncovered how New York City spends more than $180 million a year to house the homeless in hotels without contracts.

WNYC created a database using Microsoft Excel and Access to track the biggest operators of hotels for the homeless and calculate how much the city was spending.

The series also won the Society of Professional Journalists Sigma Delta Chi Award, Heywood Broun Award, Regional RTNDA Edward R. Murrow Award, Newswomen's Club of New York Front Page Award and the New York State Associated Press Broadcasters Association Award.

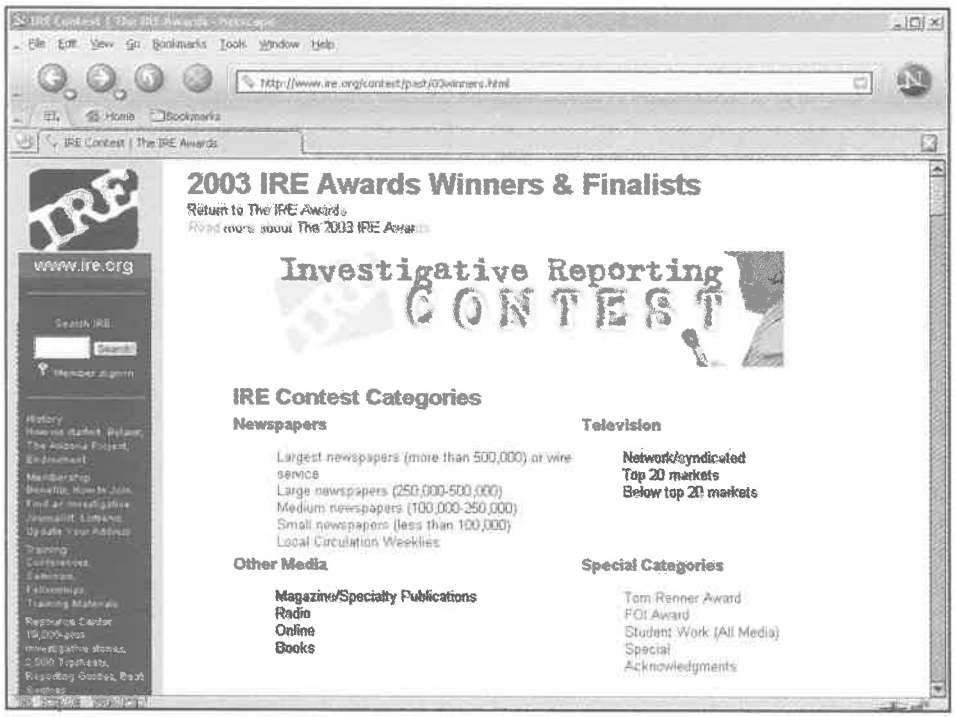
Online-Certificate
• Center for Public Integrity: "The Water Barons," by William Marsden, Maud Beelman, Daniel Politi, Erika Hobbs, Bill Allison, Andre Verloy, Laura Peterson, Samiya Edwards and Aron Pilhofer.

A yearlong investigation by the International Consortium of Investigative Journalists, a project of the center, found that public drinking water systems worldwide are being privatized.

The investigation included the analysis of five databases, including one from the World Bank containing data about contracts for drinking water projects. ICIJ used several programs including Microsoft SQL Server, Access, Excel and ESRI ArcView.

Online-Finalist
• Center for Public Integrity: "Well Connected," by John Dunbar, Bob Williams, Scott Singleton and Morgan Jindrich.

Ten months of research found that Fed-



IRE's Web site *www.ire.org* has detailed information about current and past reporting contest winners.

eral Communications Commission officials were "showered with nearly $2.8 million in travel and entertainment over the past eight years, most of it from the telecommunications and broadcast industries the agency regulates."

The series involved the analysis of data from the commission's Cable Operations and Licensing System.

"Well Connected" also won the Online News Association's award for enterprise journalism.

Book-Medal

• "Perfectly Legal: The Covert Campaign to Rig Our Tax System to Benefit the Super Rich – and Cheat Everybody Else" by David Cay Johnston and published by Penguin Group USA.

Johnston's book unravels the complexities of the U.S. tax system, and shows how the middle class subsidizes the rich and the income gap is greater than previously reported.

Johnston used Excel to create hundreds of spreadsheets using IRS statistical reports and similar data sources. He also used Access.

Special Citation for International Entry

• *The Sydney Morning Herald*: "Public Land, Private Profit," by Gerard Ryle and Brian Robins.

A seven-month investigation revealed how public property across Australia was being leased to corporations, wealthy individuals and political donors at rates far below market value.

Journalists for the newspaper built several databases using Cyber. The paper also compiled in Excel spreadsheets a list of 18,000 leases from the Department of Lands.

Student Work-Certificate

• *Daily Tar Heel*: "Raising the Cap," by John Frank, Jamie Dougher, Matt Hanson, Joe Rauch, Suzanne Presto and Lynne Shallcross.

The independent newspaper used poll and admissions data to counter the University of North Carolina at Chapel Hill's introduction of a policy to increase out-of-state enrollment.

After publication, the university decided to put the debate on hold indefinitely.

Excel was used to analyze admissions data and retention surveys. The paper also used SPSS to look at poll data.

Student Work-Finalist

• Capital News Service: "Effluent Escalation," by Dan Wilcock, University of Maryland.

The investigation revealed that overflows of raw sewage from municipal systems into Chesapeake Bay increased more than 1,200 percent in the past three years. The reports exposed errors in the Maryland Department of Environment's data and reporting system.

Wilcock used Excel to analyze three years of sewage overflows.

FOI Award-Medal

• *Argus Leader*, Sioux Falls, S.D.: "Governor Bill Janklow's pardons and commutations," by David Kranz, Patrick Lalley, Stu Whitney, Terry Woster and Jon Walker.

After dealing with inadequate open record laws in South Dakota, the *Argus Leader* published several articles on the hidden activities of former Gov. Bill Janklow.

The paper's six-month investigation uncovered that Janklow commuted more prison sentences than any other governor in the nation. He sealed about 60 pardons and didn't inform the Board of Pardon and Paroles.

Using Excel, the *Argus Leader* created its own spreadsheet and calculated the average age of inmates, average length of commutation and prevalence of serious crimes.

The series also won the Scripps Howard Foundation's public service award and Associated Press Managing Editors' Freedom of Information Award.

FOI-Finalist

• *The Sacramento Bee*: "Liberty in the Balance," by Sam Stanton and Emily Bazar.

The investigation uncovered how the war against terrorism is eroding civil liberties of Americans and immigrants. The four-part series focused on immigrants jailed indefinitely on minor charges because of unwarranted suspicions of terrorist activity.

The paper used Zoomerang to catalog the results of surveys completed by California libraries being asked for patron information by federal officials. The paper also compared immigration statistics from each year.

"Liberty in the Balance" won an investigative reporting award from the Associated Press News Executives Council, second place National Headliner Award for public service and third place in the Best of the West contest.

Contact Megan Clarke by e-mail at megan@ire.org.

ORDER NOW!

Get the scoop
on top investigations

The latest compilation book from IRE not only salutes the winners of the prestigious IRE Awards, it offers information on how the work was completed, which techniques proved most valuable and what resources were used to win. A great guide for launching projects in your own back yard.

www.ire.org/store

Fatal

continued from page 4

volved semis – could be the result of chance alone. He also pointed out that the FARS data does not disclose who is to blame.

But from studying more detailed accident data, Blower has concluded that actions by passenger vehicle drivers contribute more often to fatal crashes than do the actions of the truck drivers.

So I returned to the computer and ran more queries, this time concentrating on particular roads. By comparing a database containing only fatal accidents in which semis were involved (developed with help from our corporate trainer, Steve Cope) against the entire FARS database for Iowa, a major finding finally jumped out of the numbers. Nearly half the fatalities on I-80 in the eight-year study period involved semitrailer trucks. No other rural interstate in Iowa had such a high proportion of accidents or such a high percentage of use by trucks.

As I told KCRG reporter/anchor Renee Chu, "The accidents involving trucks occur where the trucks are." Chu interviewed me on camera about my findings for one of her stories.

During my reporting I found a former truck driver who had left the business after having been involved in a fatal accident four years ago on I-80 in Cedar County. I also found the family of a woman who in January 2000 was killed on another segment of I-80 in Cedar County. I wanted to tell about an accident from the viewpoint of the trucker, and from the viewpoint of the family of a victim.

Both the *Gazette* and KCRG gained from teaming up for the "Danger Zone" project.

"The joint project allowed the station to have access to concrete data that we would typically not devote the time and resources to uncover," Chu said. "It also allowed us to convey relevant safety information that affects every person who drives."

The televised report received positive responses, primarily from people whose loved ones worked as truckers and those who are tired of drivers who disregard trucks on the highway.

The *Gazette*, meanwhile, gained from the exposure on television.

But Iowans may have gained the most from the stories that showed how traffic on part of the interstate system had neared its capacity.

Contact Kurt Rogahn by e-mail at Kurt.Rogahn@gazettecommunications.com.

readme.txt

"Danger Zone" is available in the IRE Resource Center. To order the story, contact the center at 573-882-3364 or rescntr@ire.org and ask for Story No. 21020.

Looking for news stories that have used the U.S. Department of Transportation's Fatality Analysis Reporting System data? You can find some in the IRE Resource Center or on the Web.

Newsday analyzed three years of FARS data as part of a 12-part series, "Death on the Roads." The paper found that Suffolk County's traffic fatality rate ranks among the worst for large urban areas nationwide and is the worst in the New York metropolitan area. (Story No. 20601 in the IRE Resource Center)

The *San Francisco Chronicle* examined eight years of FARS data and found that California leads the nation in fatal hit-and-run accidents. Hit-and-runs accounted for 7.8 percent of the state's fatal crashes in 2001, more than twice the national average. Traffic safety experts cited the high number of drivers with a suspended or revoked license as one explanation for the state's problems with hit-and-run accidents. Read the story at *www.sfgate.com/cgi-bin/article.cgi?file=/c/a/2003/07/27/MN292165.DTL.*

Reader's Digest looked at fatal accident data from 1998 to 2001, focusing on incidents at intersections and removing crashes blamed on driver error or impairment. The magazine found that more than 24,000 people were killed. One-third of them were at intersections, where confusing lanes, blind spots and inadequate signs can be deadly. Read the story on the Web at *www.rd.com/common/navindex.jhtml?channelId=5&subChannelId=21&articleId=9526014&wtGroup=SPLASH_ITI&wtID=SPLASTHISI_003.*

The Indianapolis Star used three years of FARS data to track fatal pickup truck accidents on Indiana roads in which victims were not wearing seat belts. The paper's analysis found that more than 90 percent of motorists who are thrown from the truck cab during an accident die or suffer a disabling injury. Indiana doesn't require adults in pickup trucks to wear seat belts. The death toll from these accidents is rising, but efforts to change the seat-belt law have failed. Read the story online at *www.indystar.com/articles/6/135115-5226-092.html.*

To order any of the stories in the IRE Resource Center, call the center at 573-882-3364 and provide the story number.

For more information about the FARS database see *www.nicar.org/data/fars.* Journalists and journalism organizations can order the data by calling the IRE and NICAR Database Library at 573-884-7711 or 573-884-7332.

Upcoming CAR Workshops

Computer-Assisted Reporting Boot Camps

These unique seminars, taught by IRE and NICAR's experts, train journalists to acquire electronic information, use spreadsheets and databases to analyze the information and to translate that information into high-impact stories. In addition, IRE and NICAR provide follow-up help when participants return to their news organizations.

Aug. 1-6 – Columbia, Mo.
Several 2005 dates will be announced soon.
Check the IRE Web site:
www.ire.org/training/bootcamps.html

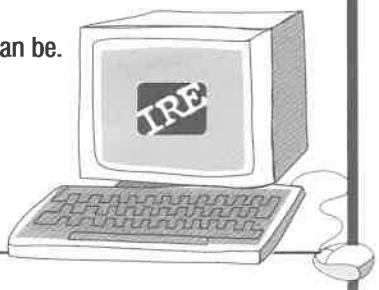
Aug. 20-22 – Columbia, Mo.
Mapping Data for News Stories Boot Camp
Learn how to uncover interesting news stories by mapping data with geographic information system software.

What previous Boot Campers have said about the experience:

"They were able to talk in a real way about issues that reporters face when practicing computer-assisted reporting skills ..."
– Analisa Nazareno, *San Antonio Express-News*

"The intense, small group workshop helped me understand how powerful database research can be. ... I now consider CAR one of the most powerful tools to expose wrongdoing."
– Holly Whisenhunt, WOAI-San Antonio

"I realized I hadn't been fully tapping my beats, and you don't have to be a computer nerd to do computer-assisted reporting."
– Teresa Taylor Williams, *Muskegon Chronicle*



Fellowships

A limited number of minority and small-news organization fellowships are available for IRE and NICAR workshops. Visit **www.ire.org/training/fellowships.html** for more information and an application.

Better Watchdog Workshops

Investigative Reporting on the Beat

IRE and SPJ, with funding from the SDX Foundation, the Chicago Tribune Foundation and the Ethics and Excellence in Journalism Foundation, have joined forces to offer a series of workshops focused on doing investigative reporting while covering a beat.

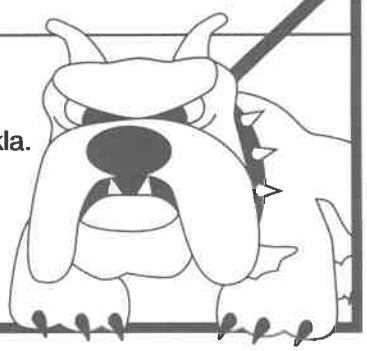
The workshops, specifically for journalists at small- to medium-sized news organizations and those in bureaus of larger organizations, will emphasize the use of FOI laws and address juggling a beat while producing investigative and enterprise pieces.

"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:

- Aug. 4, 2004 – Washington, D.C.
  In conjunction with the UNITY Conference
- Sept. 18-19, 2004 – Richmond, Va.
- Sept. 25, 2004 – New York, N.Y.
- Nov. 6-7, 2004 – Oklahoma City, Okla.
- Feb. 5, 2005 – Orlando, Fla.

For more information, visit **www.ire.org/training/betterwatchdog**
To request a workshop for your area, contact Executive Director Brant Houston at watchdog@ire.org.



Tech tip...

Making a point with geocoding

By Holly Hacker, *IRE and NICAR*

In Washington state, child molesters cannot live within a city block of day-care centers.

At least that's what the law says. In reality, more than 600 sex offenders live within 100 yards of child-care centers, KIRO-Seattle reported in February.

The IRE and NICAR Database Library assisted KIRO with its investigation by analyzing offender and sex offender data with geographic information system (GIS) analysis.

KIRO provided files with the addresses of the state's 15,000 convicted child molesters and 9,000 licensed day-care centers. The database library mapped the files, and then found child molesters and day-care centers within 100 yards of each other.

We used the geocoding function to transform the street addresses contained in the tables into geographic points. In a nutshell, geocoding takes a street address and converts it into latitude and longitude, so the location can be displayed on a map. Common GIS programs – like the program we used, ESRI's ArcView 8.3 – can geocode addresses. The process can save hours when journalists want to map crimes, schools, businesses, or anything else with street addresses and, preferably, ZIP codes.

To geocode street addresses, you need a map file of streets for the area you're interested in. To map day-care centers and sex offenders in Washington state, we downloaded the files in ESRI shapefile format for the entire state from ESRI's Web site at *www.esri.com/data/download/census2000_tigerline*. The files are based on data from the 2000 Census.

Each street is represented in the map file by a number of line segments and records in the map data table. Each street segment has two address ranges reported in the data table: one for each side of the segment.

The beauty of geocoding, at least in ArcView, is that the program can map locations even when the address in the data table fails to exactly match what's in the street file.

For example, there might be a segment for East Broadway that includes the address ranges 100-150 and 101-159. ArcView's geocoding tool attempts to match the day-care and offender street addresses to the street segment address ranges.

So if there is a day-care center at 125 East Broadway (whose address range falls exactly midway between 100 and 150), the geocoding tool will place the point at the physical mid-point of that line segment.

The beauty of geocoding, at least in ArcView, is that the program can map locations even when the address in the data table fails to exactly match what's in the street file. Whether a day-care center is listed at 203 W. Maple St., 203 West Maple Street or even 203 Maple St. West, ArcView will do its best to match it to the corresponding street segment. Each match receives a score from 0 to 100, with 100 being a perfect match.

That also means many matches are less than perfect. And you can't accept or reject them based solely on the score. Take two cases from the sex offender database. ArcView matched an offender at E. 2909 37th to 2901-3099 East 37th Ave. (identical ZIP codes) with a score of 63. But with the same score, it matched an offender at 225 South Warren to 201-299 North Warren Ave., both in ZIP code 99156. So either the sex offender's address is wrong, one of the ZIP codes is wrong, or South Warren Avenue doesn't show up in ArcView's street files. (It turns out there is a street segment for South Warren, but it doesn't include an address range.)

It's one thing if you are geocoding a few businesses in your town and you need to verify them by driving by the businesses or calling to check their locations. But what about when you're mapping 15,000 anonymous sex offenders

across an entire state that you don't live in? (KIRO received the list of all child molesters but without the names.) We didn't want to mistakenly place molesters – or day-care centers, for that matter – in the wrong location. Because we had to select molesters within 100 yards of day-care centers, putting either in a bogus spot would compromise our results.

That's where the geocoding options in ArcView came in handy. You can set the minimum score for a match. The default is 60, so any scores of 60 or more will result in a valid match. We upped that threshold to 80.

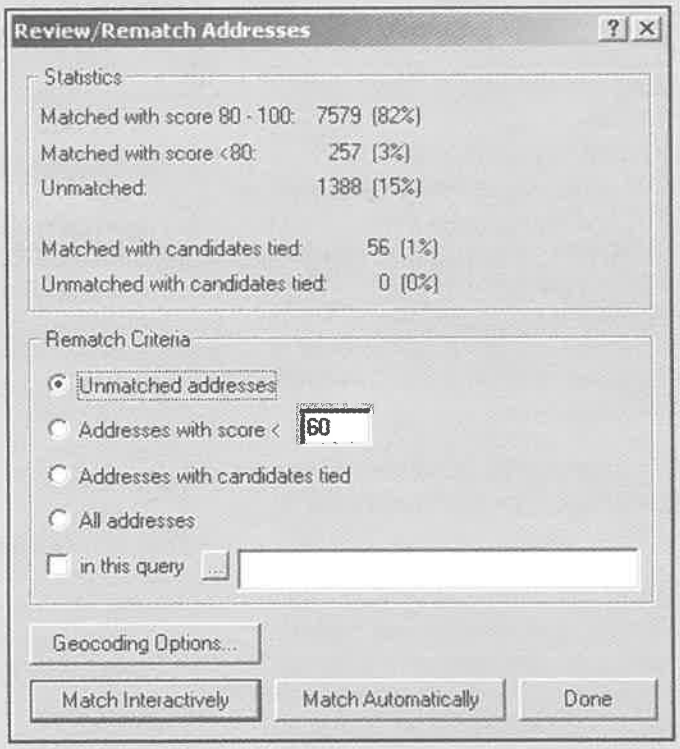
You can also tailor your results with the Spelling Sensitivity and Minimum Candidate Score options. You can lower the Spelling Sensitivity (the default is 80 out of 100) if lots of addresses are misspelled. The Minimum Candidate Score determines how closely an address matched the street file to even be considered a match.

There's no easy way to determine which settings to use – just experiment with your data. For both the molesters and day-care centers, we compared the number of valid matches we got using a minimum match score of 60 versus one of 80. We lost 257 centers using a score of 80 instead of 60. That sounds like a lot, but it was only 3 percent of all day-care centers. Similarly, we lost 843 molesters by changing a minimum score of 60 to 80 – that's 5 percent of all offenders. Using that tougher standard, we still mapped 82 percent of all day-care centers and 69 percent of all molesters in the original address files. Based on those results, we came up with 605 cases of molesters living near daycare centers.

By tightening the geocoding criteria, we erred on the conservative side in our analysis. Had we used a lower minimum match score we could have picked up more valid cases of molesters living near day-care centers – but with so many cases to check, we could have ended up with bad matches too. Even with our higher match criteria, we spot-checked matches on the Web using MapQuest and other sources to make sure we had valid matches.

See the stories on the Web at *www.kirotv.com/investigations/2825037/detail.html.*

Contact Holly Hacker by e-mail at holly@ire.org.



Attacks

continued from page 1

Year after year, the reports kept coming.

I lacked the proof but got the feeling that the number of vignettes – and the problem – was increasing. So I put the idea on a story list, not as an idea for a major project but just a suggestion for a single story that someone else might consider. Years later, during a meeting to discuss project ideas in early 2002, projects reporter Mike Wagner convinced *Daily News* editors to pursue the idea as a major project, asserting that most Americans would be shocked.

Following that February 2002 meeting, I started filing the first of more than 75 federal Freedom of Information Act requests and appeals with the Peace Corps, including one for the agency's main database of assaults, the Assault Notification Surveillance System. The database contains records for nearly 3,000 attacks on volunteers since 1990. Two months later, I appealed for lack of proper response, and in May the agency released the first version of the database at no cost.

What we got was virtually worthless.

The Peace Corps denied us the unique number assigned to the assault, the country of the assault, report date, time, day and date of assault, type of place where the assault occurred, a brief description of the attack, whether local authorities were notified, and whether the victim was medically evacuated.

All this information, according to the Peace Corps, was being withheld to "protect the privacy interest of individuals." Since we weren't demanding the names of the victims and since just about every law enforcement agency in America releases the same information, I appealed and challenged the agency to defend how such things as the name of a country would violate someone's personal privacy.

In the meantime, we continued to pursue the story, by interviewing hundreds of people in 11 countries and collecting thousands of pages of paper records, many of them overseas.

Our newspaper sued the Peace Corps in federal court in November 2002, and the agency suddenly became much more cooperative.

Underreporting also appeared to be a special problem for the Peace Corps.

By July 2003, we had a copy of the database with nearly all of the fields. We also were able to get two other databases: the Epidemiologic Surveillance System, containing monthly reports of burglaries, alcohol-related incidences, pregnancies, STD cases, and other problems encountered by volunteers; and the Deaths in Service, which provided details about how more than 200 volunteers had died since 1962. I analyzed the data using Paradox database manager.

Our analysis and other records found:

• The number of all assaults increased more than 200 percent 1990-2002, while the number of volunteers increased about 29 percent. We found a 300 percent increase in aggravated assaults, the most serious type of physical assaults. In 2002, there was an assault or robbery of a volunteer every 23 hours.

• Many assaults were linked to a decades-old practice the agency refuses to change: Sending young volunteers – the majority of them women – just out of college to live alone with virtually no supervision in remote regions of some of the most dangerous countries in the world. Volunteers frequently arrive at their sites fresh out of training without adequate housing or a job that keeps them busy. Some turn to drinking, using drugs, traveling to unsafe areas or engaging in other activities that put them in danger.

• Volunteers who were alone frequently were the targets of assaults, and young women appeared the most vulnerable. In 62 percent of the more than 2,900 assaults, the victim was identified as being alone, and in 59 percent of all assault cases, the victim was identified as a woman in her 20s.

• Alcohol was identified as a factor in nearly one in three assaults since 1999, although the assailant was the sole person drinking in some of those cases. Alcohol was also linked to nearly one in six deaths since 1962.

• Families of dead volunteers were unaware of crucial details in the databases and other records we obtained. One mother didn't know her daughter had attempted suicide with aspirin in Ecuador before lapsing into a coma and dying. The brother of another volunteer said the family had never been told that his 22-year-old sister had written a resignation letter the day she disappeared in Africa or that the Peace Corps suspected suicide. Her body was found five days later floating in a river, and the family was led to believe that a crocodile had killed her.

As helpful as the databases proved to be, there were a number of problems with the way the agency collected information.

The Peace Corps failed to collect worldwide crime statistics until 1990, 28 years after it first sent volunteers overseas. Certain crimes still weren't specifically identified in the assault database. Abductions, kidnappings and even murder usually were recorded as assaults in the data.

Some crimes that should have appeared in the data weren't there at all. We couldn't find two known murders in the database.

In January 2001, weeks before a volunteer disappeared in Bolivia, Mike Mulvaney and another volunteer were abducted in a taxi, the second such abduction of a volunteer within 11 months in that country. Mulvaney was beaten during the incident but managed to escape. Both the Peace Corps country director in Bolivia at the time and the agency's worldwide security director confirmed that Mulvaney spoke to them about the abduction, but the incident wasn't recorded anywhere on the assault database. A written response from the Peace Corps says, "Not having a record of this alleged assault was not an error." Mulvaney, it says, should also have reported it to the medical office.

Underreporting also appeared to be a special problem for the Peace Corps, since volunteers risked being moved out of the country or even out of the Peace Corps by simply making an assault report. Government auditors estimated that underreporting was as high as 50 percent.

When we asked Elliot Jaspin of the Cox Newspapers Washington Bureau, and founder of what is now NICAR, to review our findings before publication, he found something even more troubling: The crime rate the agency was making available to Congress and the public was based on incidents, not the number of victims. After three women were raped in El Salvador in 1996, the incident was recorded as a single rape.

About one in every five assaults involved more than one volunteer; but it was impossible to determine what happened to the other volunteers in each case by using the data we had.

At a meeting in Washington, Jaspin, reporter Mei-Ling Hopgood and projects editor John Erickson and I confronted Peace Corps officials, but they insisted their data collection methods were the best way to make volunteers safer. They also said the U.S. General Accounting Office, during an earlier audit, had validated the Peace Corps' data collection methods. The GAO later said it was unaware that the Peace Corps failed to count all victims.

Problems with the way the agency collected and interpreted data was one of several reasons the extent of dangers facing Peace Corps volunteers had been disguised for decades. Another reason is that the attacks occurred thousands of miles away. Still another reason is that the agency made little effort to publicize them or deliberately kept people from finding out – while emphasizing only the positive aspects of Peace Corps service.

Contact Russell Carollo by e-mail at rcarollo@coxohio.com.

readme.txt

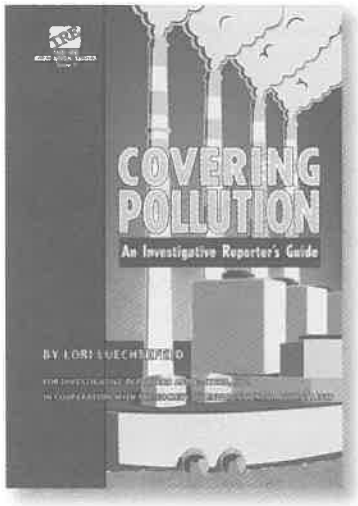
The *Dayton Daily News* series "Casualties of Peace" won a certificate in the 2003 IRE Awards for medium newspapers (100,000-250,000 circulation).

To order a copy of the story and a contest entry questionnaire that details the reporting, contact the IRE Resource Center at 573-882-3364 and ask for Story No. 20926.

For more information about other winners and finalists in the IRE Awards, visit the IRE Web site at *www.ire.org/contest/past/03winners.html*.

The 2004 edition of Winning Investigations details how winners and finalists in the 2003 IRE Awards reported their stories. To order, go to *www.ire.org/store*.

COMING SOON!



COVERING POLLUTION
An Investigative Reporter's Guide

by IRE
in cooperation with SEJ

This primer gives an overview of useful resources reporters can use for local investigations into environmental pollution. Its main focus is to show how to get to the heart of an investigation quickly and without waiting months for FOI requests to be fulfilled.

Filled with examples and references to stories, tipsheets and other resources available from IRE and SEJ.

Watch the IRE Web site for information on ordering.

www.ire.org

Arrests

continued from page 1

ity children had swelled over eight years. In 1995, 60 percent of all juvenile arrests in the city involved white children. By 2003, 60 percent of all juvenile arrests involved black children, who make up 18 percent of all kids in the city.

My story, "The Color of Justice," didn't start out as an examination of the racial composition of juvenile arrests.

I had just been assigned to cover city hall when the city council member who represents central Columbia was considering whether to push for a curfew law. One of the opponents said the law would be enforced primarily in the central city, where the city's black population is concentrated, and act as a sort of "Jim Crow law."

From my previous reporting experience, I knew many police departments use computers to map where crimes and arrests occur. Sure enough, Columbia possessed the same technology. I asked the police to map the general locations of curfew-hour arrests. It turned out the curfew opponent was right on the mark – most curfew-hour arrests occurred in the central city.

I negotiated for as many fields as I could get.

But when I showed the map to my editor, he wasn't convinced there was a story yet. That's when my lessons of the NICAR boot camp kicked in. I knew that I needed to talk to the keeper of the database at the police department to determine what other information was available.

I negotiated for as many fields as I could get. Because juvenile records are protected, I had little leverage. What I received from the city on CD-ROM was messy.

I could have sworn that I requested the data in delimited text format, but when I opened the file, I found that the data ran together and had inconsistent breaks. Appropriately, the city's computer technician labeled the CD "Datadump."

The IRE and NICAR Data Library wrote a Perl program that cleaned the file for the most part, though there were still some problems. Later when I sought additional data, the city's computer tech produced the data in an Excel spreadsheet, which he could have done from the start.

Many fields in the more complete data set were inconsistently filled, such as the officer, whether a weapon had been involved and time of the arrest. However, I found that the data included the following columns of information: uniform crime reporting code, age, year, sex and race. That's what I stayed with.

I was dealing with about 15,000 records, so it was easy for me to import the data from Excel to Access for querying. I used Access to count the number of arrests by race and year; soon I had totals for 1995 through 2003. I then moved those numbers to an Excel spreadsheet, where I created a line graph that showed a dramatic shift toward more black arrests than white arrests.

Now all I had to do was get the police department to explain why.

I showed my analysis and the curfew-hour arrest map to the police chief and his captain and they had no explanation.

"You can make numbers do anything you want them to," the captain replied.

My editor prompted me to find other cities and groups that have addressed the racial imbalance in arrests and prisons and I eventually learned about how Portland, Ore., Santa Cruz, Calif., and Seattle had reduced juvenile arrests and incarcerations.

A local community outreach center put me in touch with a family that had run-ins with the police and probation department. One of the children ended up serving as a focal point for my story.

After the story was published, the disproportionate minority arrests emerged as an issue in the mayoral election.

Here are some lessons I learned reporting the story:

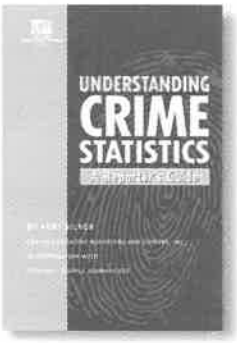
• Nearly 70 percent of all police departments track and report their arrests by race to the federal government. It's still uncertain whether all departments will be required to report the racial composition of their arrests, as the Juvenile Justice and Delinquency Prevention Act evolves. However, because courts are funded by a separate pool of money than police

REQUIRED READING For Your Newsroom

Understanding Crime Statistics

A Reporter's Guide, by Kurt Silver

Know the data and reports available to add perspective and authority to your crime reporting. Learn how to use the data and avoid common pitfalls while writing responsible crime statistics stories.



ORDER TODAY!

Call 573-882-3364 with your VISA or MasterCard

-OR-

Visit our Web site at **www.ire.org/store** for online ordering or order form downloads

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

agencies, police don't have the same incentive to report arrest statistics as court offices do with inmates.

• Because police organizations are hierarchical, you'll have to start somewhere at the top and find your way to the department's database guru.

• If you ask the police whether they keep computerized records for the arrest of each child, and then ask to see a record layout (which shows the fields in the database, such as name, date of birth, race, location of arrest and charge), the police will often point you in the direction of the person who keeps the database.

• Be willing to negotiate for fields you want after you see which are available. Because juveniles' identities are highly protected under the law, you will have little leverage for negotiating.

• Ask for as many years of data as possible so you can observe trends. I had eight years of data and was able to see the shift toward the arrests of black children.

• Consider the mapping possibilities. During my reporting, I learned that the police department uses mapping to mark locations of crimes and target patrols. While police may not divulge the specific locations of arrests, they might reveal the general block group or area of arrest. I simply asked police to map juvenile arrests by block group.

• When assembling juvenile justice data, check out the U.S. Justice Department's Office of Juvenile Justice and Delinquency Prevention Web site at *www.ojjdp.ncjrs.org*. The site has national statistics on juvenile arrests, which can be compared on a county-by-county basis.

You can read "The Color of Justice" on the Web at *http://archive.columbiatribune.com/2004/Feb/20040208Feat051.asp.*

Contact Dave Moore by e-mail at dmoore@tribmail.com.

Interviewing the Interviewers

In the latest addition to The IRE Collection, some of the most renowned interviewers in investigative journalism share their experiences, techniques and advice. The book includes chapters on dealing with sensitive issues, cross-cultural interviewing, confrontational interviews, handling whistleblowers, technical interviews and finding and cultivating sources. Each chapter includes "Tips from the Pros," with specific, time-tested techniques. The book also includes a list of tipsheets and tapes taken from IRE conferences to provide additional help.

IRE members: $12 each
Nonmembers: $18 each

Plus media-rate postage: $4 for the first book, $2 for each additional book

E-mail questions to rescntr@nicar.org

ORDER NOW!

By PHONE

Call 573-882-3364 with your VISA or MasterCard

By MAIL

Send your check to IRE Resource Center: 138 Neff Annex Missouri School of Journalism Columbia, MO 65211

By WEB

Visit www.ire.org/store for online ordering or to download an order form

Cash

continued from page 3

spective campaigns, making deadline reporting easier than ever. Also, electronically filed forms include expenditures made by the committee, the FTP data does not.

Electronic filings are viewable online or they may be downloaded in a comma-delimited format, which can be easily imported into a spreadsheet or database manager.

(For more information about working with the electronically filed data, see the July-August 2002 *Uplink*. For further reference see *www.claptrap.org/ire/ElectronicFilingandthe FECPart1.pdf* and *www.claptrap.org/ire/Electronic Filing and the FEC Part 2.pdf*).

In contrast, the FTP data is much harder to work with because, among other things, the files can be quite large and require quite a bit of cleanup. (The IRE and NICAR Database Library provides a cleaned version of this database to journalists at a nominal cost. See *www.ire.org/datalibrary/databases/feccc* for more information).

The biggest drawback to using the FTP data is the lag time, which can range for a few days for electronically filed reports to as much as a month for those Senate filings that must be manually entered.

On the other hand, the data is the official electronic record. Electronic filings are considered preliminary because FEC officials have not examined them. Furthermore, the FTP data is designed to be maintained over the long term. Each record in the database is uniquely identified, which allows the data to be updated correctly when filings are later amended (a frequent occurrence).

NICAR has collected literally dozens of tipsheets covering campaign finance data. You can search for tipsheets on the Web at *www.ire.org/resourcecenter/initial-search-tipsheets.html*.

Other resources

In addition to raw campaign finance data, there are some other sources journalists should know about. Under the new law, funds used to pay for what are called "electioneering communications" must be fully disclosed within 24 hours, even if the group sponsoring is a nonfederal entity or even an individual.

Electioneering communications are a subset of what has come to be known as "issue ads" – commercials that are virtually indistinguishable from regular campaign ads, but do not specifically call for the election or defeat of a candidate. These ads proliferated over the past decade because they could be bought with unregulated soft money. The new law tries to reign in some issue ads by restricting some groups from airing them within 30 days of a primary election or 60 days of a general election, and requires full disclosure from other groups or individuals who buy such ads.

These filings are posted on the FEC Web site at *www.fec.gov/pages/electioneering.html*.

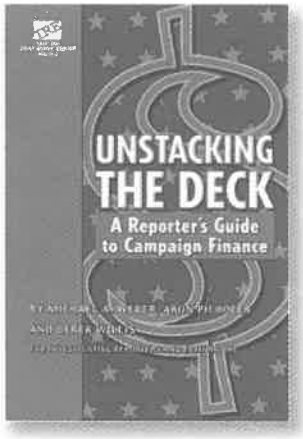
The FEC also has launched a searchable database of enforcement activities, which can be found at *http://eqs.sdrdc.com/eqs/searcheqs*. The interface allows users to search for cases in any number of ways – by respondent, by case number or even a fuzzy word-matching scheme.

Unfortunately, the system only includes newer cases – those made public since January 2002. But it includes all the documents associated with the case in a searchable format, even if they were not specifically submitted in that format. For more information about the system and its limits, see *http://eqs.sdrdc.com/eqs/eqs_info.html*.

Political nonprofits

A more recent development has been the rise of political nonprofit organizations, which are generally (though not entirely accurately) referred to as Section 527 organizations. The name refers to the section of the federal tax code under which these groups and

Unstacking the Deck:
A Reporter's Guide To Campaign Finance



The latest Beat Book from Investigative Reporters and Editors, Inc., is a guide to navigating the language and practices of campaign finance.

Journalists will find it invaluable for pursuing stories about the impact of money on elections, political parties and candidates at the federal, state and local levels.

- Understand the loopholes in soft money restrictions.
- Learn about the increased use of nonprofits to funnel money.
- Find out how to track where candidates spend their money.
- Learn how to find and read useful documents and data.
- Packed with resources, Web sites and story tips!

ORDER NOW!

By phone: Call 573-882-3364 with your VISA or MasterCard
By Web: Visit www.ire.org/store to order online or download an order form.
By mail: Send your check to IRE, 138 Neff Annex, Missouri School of Journalism, Columbia, MO 65211
IRE Members: $15 each • Nonmembers: $25 each
Plus postage: First class — $4 for the first book, $2 for each additional book

all political committees achieve tax-exempt status. But the difference between 527s and other political committees is enormous.

For one, the 527s don't file with the FEC or any state elections regulatory agency (unless that state's law specifically requires it). Prior to 2000, they didn't submit any financial disclosure documents of any kind. Since then, they have filed with the Internal Revenue Service, and those forms are available at *http://eforms.irs.gov*.

A few words of caution: because of widespread confusion about which committees are required to file, there are literally tens of thousands of erroneous reports on the IRS site, most of them from state and local candidates who already file with their state elections authorities. The Center reviewed these filings and found about 500 committees that fit the definition of a true 527 out of more than 20,000 filed with the IRS.

For about the past year, most of these committees have been required to file electronically with the IRS. Those filings can be downloaded in a comma-delimited format from the IRS Web site. The download page includes a very detailed document that explains the contents and layout of the data file, which is a bit confusing.

The database we created at The Center for Public Integrity is searchable and includes data back to 2000. (See *www.publicintegrity.org/527*). Other groups have tracked 527 activities. The Center for Responsive Politics (*www.crp.org*) also has posted a searchable database of filings from last year forward and includes the largest 527 organizations active in federal elections. PoliticalMoneyLine (at *http://politicalmoneyline.com*) has tracked 527 activity for several years, but details of its database are restricted to subscribers and its information is not as comprehensive as that offered by the Center for Responsive Politics or The Center for Public Integrity.

Contact Aron Pilhofer by e-mail at apilhofer@publicintegrity.org.

State

continued from page 3

sylvania, Illinois and Rhode Island are examples of the latter. Typically, such files can be batch-downloaded if you know the filenames.

States that offer HTML versions of their data that can be put into a database with a little bit of work include Arizona, North Dakota, North Carolina and Nebraska. In some instances, you can simply save the HTML and use Access' import utility, which will convert HTML tables into Access pretty easily. In other cases, it may make sense to develop a script that retrieves the pages or automatically "scrapes" the information from them.

Other states offer what can be called "pseudo-Web disclosure," providing only summary information but not the itemized records.

It should come as no surprise that such practices are wildly inconsistent in terms of the information provided and whether electronic records match paper filings. Maine, for example, offers text filings on its Web site purporting to contain campaign finance information, but they very often are missing thousands of records from party committees. Before revamping its system, Indiana offered downloads that included all filings by certain committees, meaning that amended records were simply piled on top of the original data with the potential for significant overlap.

Expenditure data, while often overlooked, can be a valuable resource. But it can have pitfalls, particularly when it comes to the issue of "subvendors," the breakdown of payments made by credit card, for example. Check with your state to see if it requires such itemization on its reports and whether any data available makes it easy to separate the lump sum payments from the itemized records (in most cases, you want the latter).

Here are some tips for handling state campaign finance data:

• Know what the files, and thus the data, are supposed to look like. It makes sense to grab a copy of a blank report, even if you get all-electronic data.

• Remember that a number of states have some form of public financing, and those records often appear in the general contributions. While it may be technically accurate to list the state government among the largest contributors to a candidate, you probably will want to remove those records from your query results.

• If you are interested in the activities of political parties be sure to check your state's laws to see what kinds of accounts parties can maintain. A handful of states permit parties to have operating accounts for political use and administrative accounts for overhead and generic expenses. Sometimes these administrative accounts have little or no disclosure. South Carolina's parties will begin disclosing contributions to their administrative accounts at the end of the year. Other parties have building funds or housekeeping accounts that file separate reports, so be sure to check or you might be missing some of the picture.

• Don't forget the small stuff: states treat small-dollar contributions very differently. Some states require lump sum reporting of contributions below a certain threshold, while others, like Nebraska, don't require any reporting of contributions that are not itemized. In smaller states, such figures could seriously affect your numbers.

Contact Derek Willis by e-mail at derek@thescoop.org.

He added that the story helped to dismiss the myths of voter turnout. "In this case, reality didn't mirror stereotypical beliefs."

During election season, Rom uses CAR to look historically at voter turnout by mapping several elections over time.

Campaign finance reporting is a staple of the election season. Remember to look for the unusual, like Alan C. Miller of the *Los Angeles Times* did a few years ago, when he focused on young contributors.

Starting with the Federal Election Commission records, independent researcher Dwight Morris, hired by the newspaper, looked for donors who were identified as "students."

After eliminating donors who were listed as students, but turned out to be housewives or professionals who went back to school, the analysis found nearly 9,000 student donations totaling $7.5 million. Using online newspaper searches, voting and motor vehicle records and available birth records, Miller and Morris were able to specify the age of each donor.

Their list included a 15-year-old who donated $20,000 to the Democratic Congressional Campaign Committee and a 10-year-old who gave $1,000 to Republican presidential candidate Lamar Alexander. (You can read about the reporting of the story in the July-August 1999 *Uplink*).

If the last presidential election is a precursor, then 2004 promises to be interesting and full of opportunities for new CAR techniques and stories.

Contact Megan Clarke by e-mail at megan@ire.org.

readme.txt

IRE and NICAR's Campaign Finance Information Center has more resources for covering the election. See *www.campaignfinance.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: David Donald, ddonald@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

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

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

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

Uplink

Uplink Info

A newsletter of the National Institute for Computer-Assisted Reporting

Editor
Brant Houston
brant@ire.org

Managing Editor
David Herzog
dherzog@nicar.org

Asst. Managing Editor
Jeff Porter
jeff@nicar.org

Art Director
Lisa Triefenbach

Copy Editor
Pia Christensen

Contributing Editors
Christina Caron
Megan Clarke
Jaimi Dowdell
Holly Hacker
Andrea Lorenz
Tim Ragones

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

Director of Publications
Len Bruzzese
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.