Uplink

January • February 2005
Volume 17, Number 1

Published bimonthly by the National Institute for Computer-Assisted Reporting

www.nicar.org

ELECTIONS

Post-vote politics, payoffs

By Andrea Lorenz, *IRE and NICAR*

November has come and gone. Journalists, readers and viewers don't want to hear or see the words "red" and "blue" in the same sentence as the word "state." The 2004 presidential election is long over.

But story ideas abound for enterprising journalists who want to use computer-assisted reporting to find out what happened in the elections and explore possible connections between campaign contributions and political favors.

Power and influence

"Who paid for this election, and what do they want?" asked Ron Campbell, a staff writer at *The Orange County Register*. Journalists can answer these questions with the help of campaign finance data available on the Web or from the IRE and NICAR Database Library. Campaign contributions can explain why some bills pass Congress and other stall in committee.

continued on page 20

SPOTLIGHT: PROBING THE POLICE

Mapping, stats show scope of racial bias

By Andy Lehren, *Dateline NBC*

For a few days in April 2001, it was hard to miss news accounts about riots in Cincinnati.

Daily stories outlined how police officers chased an unarmed 19-year-old black man named Timothy Thomas. It was past midnight in the Over-The-Rhine neighborhood.

A dispatcher called in officers to help pursue Thomas. The voice over the police radio offered a few details: a brief description of Thomas, and that he had 14 outstanding warrants. An officer responding to the call spotted Thomas. He got out of his car and drew his gun, fired and killed Thomas.

Cincinnati rioters protested the police shooting, capturing headlines across the

continued on page 12

SPOTLIGHT:

For more about probing the police see:

- Mapping L.A.'s neglected murder cases, p. 6
- Discovering disorder in St. Louis law enforcement, p. 8
- Nashville police violate hot-pursuit rules, p. 10
- The January-February 2005 *IRE Journal*.

TAX BREAKS

Tallying lost money through farm loophole

By Paula Lavigne, *The Dallas Morning News*

Right across the street from our office building is a field. Every once in a while we'll notice farm equipment rolling onto the property, tending to the growing rows of various crops.

It's prime real estate, almost 110 acres near the intersection of two major highways. The owners are members of one of Dallas' most notable old money families who, among other pursuits, are land developers.

They are not farmers. Yet, by leasing their land to a farmer who reaps a nominal harvest, they get an agriculture tax break that saves them about $545,000 each year in taxes. Their savings is lost money for cities, schools and county governments.

Last year, two *Morning News* county reporters (Ed Housewright and Kevin Krause) and I decided to investigate

continued on page 18

Bits & Bytes

Felon hunters

Matching government databases can produce high-impact reporting. Investigative reporter Steve Chamraz of KCTV-Kansas City recently reported on Missouri felons with hunting licenses who were discovered with database matching, and state lawmakers took note.

KCTV asked the IRE and NICAR Database Library to compare state hunting licenses and felon data tables. Database Library analysts obtained the data from the state corrections and conservation departments, cleaned the data and formatted the tables to make database matching possible. Then the analysts matched the two tables.

The results: nearly 1,000 felons –13 of them convicted killers – in Missouri had hunting licenses. It is against federal law for felons to own or even hold a gun. However, it is not against Missouri state law to sell firearm-hunting permits to convicted felons.

As a result of the report, some state legislators are considering changing the state law. See the report at *www.kctv5.com/Global/category.asp?C=9271&nav=1Puc.*

IRE and NICAR can analyze data for your news stories. Contact Database Library Director Jeff Porter at 573-882-1982 for more information.

Disaster aid

Journalists interested in discovering which businesses or people have received government loans relating to a disaster

continued on page 4

INSIDE NICAR

Politics and elections: An evolving story

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

Covering elections and politics is an ongoing enterprise. The November presidential election is over and the crush of resulting stories has lightened. Still, it's good to remember that some of the best stories coming out of the elections are the ones developing now, the ones you can do any time of the year to examine subjects that are key to understanding our democracy: political power and voter registration, to name just two.

In the most recent election season, journalists from coast to coast fired up their databases and discovered interesting stories about voters registered in more than one locale, political nonprofit contributions and problems with felon-voter purge lists – all great stories. From the stories reported so far, it's clear that even more are waiting to be uncovered.

This issue of *Uplink*, starting with the front page article by IRE and NICAR's Andrea Lorenz, presents resources for journalists who want to stay on the trail through the year ahead.

• • • • • • • • • •

It's that time of year: We're busy getting ready for the CAR Conference March 17-20 in Hollywood, Calif., putting together the best possible schedule of hands-on classes and panels on computer-assisted reporting topics.

For the first time, we're offering a full day (Thursday, March 17) of special sessions tailored for advanced CAR practitioners.

So mark your calendar; we'd love to see you in Hollywood. For more information about the conference or to register, visit *www.ire.org/training/hollywood05.*

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

UPCOMING TRAINING

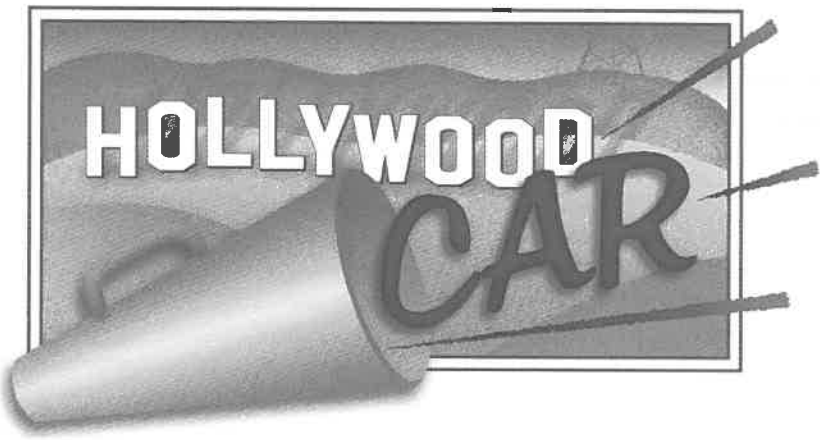
IRE and NICAR have several training opportunities for journalists seeking hands-on instruction in using computer-assisted reporting.

There are three weeklong Boot Camps 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 Camps for this year are March 20-25, May 15-20 and Aug. 7-12.

A special mini-Boot Camp designed for editors will be April 8-10 in Columbia. The training, which places an emphasis on what newsroom managers need to know, will include just enough hands-on work to provide an understanding of CAR.

Journalists interested in learning how to map data for news stories can take advantage of a mini-Boot Camp in using geographic information systems (GIS). The training, Aug. 19-21 in Columbia, will use ArcView 9 GIS. Participants will learn how to create basic maps, work with projections and geocode data.

For more information on these and other IRE and NICAR training events see *www.ire.org/training/otr.html.*



ANNUAL CAR CONFERENCE

Training, networking at CAR gathering

The 2005 Annual Computer-Assisted Reporting Conference will feature an Advanced Day for CAR veterans, special tracks for broadcasters and beginners.

The Advanced Day will be Thursday, March 17, and will be followed by three main days of panels and hands-on training for all levels of experience and expertise. All sessions will be at the Renaissance Hollywood Hotel in Hollywood, Calif.

Keep posted about the details by visiting the conference Web site at *www.ire.org/training/hollywood05*.

Practitioners in CAR will conduct more than 40 panels on nearly every beat and 60 hands-on classes on spreadsheets, database analysis, statistics, mapping and searching for and downloading data from the Internet.

KNBC-Los Angeles and the *Los Angeles Times* are conference hosts. *The Orange County Register* is a sponsor. A local committee of journalists is helping put together panels that will include a regional focus.

The conference Demo Room will offer informal tutorials and small workshops on specific federal and state data and analytic techniques. Top CAR journalists can use the room to share some of the latest techniques and stories in an informal setting.

Among the topics will be using data analysis to report on homeland security, pollution, homicides, natural disasters, elections and campaign finance, inequities in education, exploitation of the poor, the military and the war in Iraq, and immigration border issues.

Advanced topics will include social network analysis, data cleaning, open-source tools, constructing intranets for newsrooms and mining text databases.

For beginning CAR reporters, the conference will include a start-to-finish track and a mini-boot camp of hands-on training. The mini-boot camp will offer a series of classes – open to the first 36 who sign up – that takes novices through data analysis with spreadsheets and database managers, finding data on the Internet and downloading that data into their programs. The classes will be linked to panels that highlight these skills to produce in-depth stories.

Experienced practitioners can learn to hurdle data challenges with advanced programming, sophisticated mapping techniques, scraping data from the Internet and employing more complex statistical methods.

The conference also will have workshops on teaching CAR that will help educators and newsroom trainers. In addition, there will be panels on writing and editing the data-heavy story.

Speakers

These are just some of the expected conference speakers and instructors:

- David Boardman, *The Seattle Times*
- Ziva Branstetter, *Tulsa World*
- Ron Campbell, *The Orange County Register*
- Danielle Cervantes, *The San Diego Union-Tribune*
- Sarah Cohen, *The Washington Post*
- Anthony DeBarros, *USA Today*
- Steve Doig, Arizona State University
- David Donald, IRE and NICAR
- Jaimi Dowdell, IRE and NICAR
- Joe Ellis, KDFW-Dallas/Fort Worth
- Matthew Ericson, *The New York Times*
- Mike Fabey, *Savannah Morning News*
- Larry Gillick Jr., East Carolina University
- Matt Goldberg, KNBC-Los Angeles
- Joel Grover, KNBC-Los Angeles
- Holly Hacker, *The Dallas Morning News*
- Chris Halsne, KIRO-Seattle
- Jeff Harris, KPIX-San Francisco
- Brad Heath, *The Detroit News*
- Tim Henderson, *The Miami Herald*
- David Herzog, NICAR & Missouri School of Journalism
- Sean Holstege, *The Oakland Tribune*
- Brant Houston, IRE and NICAR
- John Jackson, *Los Angeles Times*
- Mark Katches, *The Orange County Register*
- Jennifer LaFleur, *The Dallas Morning News*
- Daniel Lathrop, The Center for Public Integrity
- Andy Lehren, Dateline NBC
- Rick Linsk, *St. Paul Pioneer Press*
- John Maines, *South Florida Sun-Sentinel*
- David Milliron, *The Atlanta Journal-Constitution*
- Ron Nixon, Minneapolis *Star Tribune*
- Dick O'Reilly, Consultant
- Paul Overberg, *USA Today*
- Cheryl Phillips, *The Seattle Times*
- Aron Pilhofer, The Center for Public Integrity
- Jeff Porter, IRE and NICAR
- Carl Prine, *Pittsburgh Tribune-Review*
- Neil Reisner, Miami *Daily Business Review*
- Janet Roberts, *The New York Times*
- Natalya Shulyakovskaya, *The Orange County Register*
- Doug Smith, *Los Angeles Times*
- MaryJo Sylwester, *USA Today*
- Tom Torok, *The New York Times*
- Enric Volante, *Arizona Daily Star*
- Paul Walmsley, IRE and NICAR
- George Watson, *San Bernardino Sun*
- David Wethe, *Fort Worth Star-Telegram*
- Holly Whisenhunt Stephen, WOAI-San Antonio
- James Wilkerson, *The New York Times*
- Margot Williams, *The New York Times*
- Derek Willis, *The Washington Post*
- Debbie Wolfe, *St. Petersburg Times*

Bits & Bytes

continued from page 2

can use the recently updated Small Business Administration (SBA) Disaster Loan database from the Database Library

The database logs every approved loan. Each loan record in the database includes the amount and the related disaster. Loans to businesses also indicate whether the loan has been paid.

Current as of September 2004, the data includes more than 805,500 records of loans and nearly 7,800 disasters.

The (Fort Worth) *Star-Telegram* used the data in the aftermath of tornadoes in Fort Worth to see who received the promised federal assistance. (Story No. 16808). Contact the IRE Resource Center at 573-882-3364 or rescntr@ire.org.

Network news

IRE and NICAR offer journalists resources for using social network analysis (SNA) software to analyze data for news stories at www.ire.org/sna. The resources list includes links to beginning guides to SNA and to recent news stories that have used the software to display connections between political contributors and political nonprofit committees, and between an Islamic charity and groups associated with Osama bin Laden.

The Web site also has a wealth of tipsheets for journalists who want to learn more about SNA, the available software and potential applications to stories.

FIRST VENTURE

Data helps uncover Ohio's dangerous dams

By Ron Regan, WEWS-Cleveland

Our November news report that 450 of Ohio's 1,700 dams are considered unsafe provoked a nearly incredulous response from the state's dam safety chief: "How did you run across that?"

Admittedly, that was pretty obscure information. Considered more closely, it's information that could ultimately save lives and there was only one way for me to get it – by using a database manager to analyze dam inspections data. Before last August, when I attended an IRE and NICAR Boot Camp, I was clueless about how to do that.

For years, computer-assisted reporting remained a mystery. Yet, without those skills, my ability to closely examine the condition of Ohio's dams would have been extremely lacking.

And best of all – this was a report that had started in the Boot Camp lab last summer. In other words, I left a week of training with a story in my pocket. Not bad. (See the March 1996 *Uplink* for guidelines in evaluating dams).

Here's how it went from the classroom to viewers' living rooms: It started with a U.S. Army Corps of Engineers database called the National Inventory of Dams. I first laid eyes on the database during the NICAR training and quickly realized I could turn this from a classroom exercise into a more comprehensive report. So, before leaving the Boot Camp, I purchased the data from the IRE and NICAR Database Library (for more information see *www.nicar.org/data/dams*) and with some assistance from Jeff Porter, the director of the Database Library, was able to whittle down the data for the United States in the Microsoft Access database manager to just Ohio dams.

This is so elementary that CAR veterans will probably snicker at the thought of someone who was so enthralled with entry-level database work. But, if you're reading this, there's a good chance that it is as much of a mystery to you as it was to me.

I wanted to know how many of these dams were near cities, who owns them and when they were last inspected. Access, the database manager we used in the Boot Camp, helped answer these questions.

Think of Access as a big data container, or a filing cabinet. You can use it to store and view data tables, and to write queries that filter your data. I was able to use Structured Query Language (SQL) to write queries that showed me the name of each dam, the year its construction was completed, the most recent inspection date and the hazard level, which indicated how close the dam is to populated areas. In my queries, I asked for dams with a high hazard level.

With a few keystrokes – and I mean very few – I discovered 1,727 dams in Ohio with 462 rated high hazard. Not only that, many of the dams appeared to have gone years without inspection. And, most disturbing, many of them were more than 50 years old.

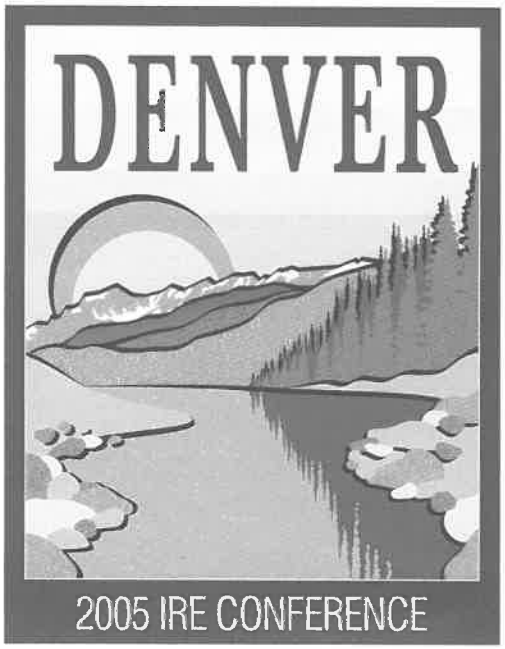
Now, here's the really big lesson in learning CAR: You can run queries in seconds, but that does not replace old-fashioned fact-checking and digging. I obtained paper records of the most recent inspections, using open-records requests. Without checking the paper records, I would have made serious errors.

I also learned that visiting the dams and walking around them pays off. I discovered a charming 85-year-old who lived

continued on page 13

2005 IRE Conference • June 2-5

The best in the business will gather for more than 100 panels, workshops and special presentations about covering public safety, courts, national security, the military, business, education, local government and much more. Visit *www.ire.org/training/denver05* for more information and updates.



Conference Location and Hotel:

Grand Hyatt Denver, 1750 Welton St., Denver, CO 80202

Hotel Reservations:

Please call 800-223-1234 by Friday, May 6, to get the discounted room rate of $119 plus tax. When making your reservation please ask for the Investigative Reporters and Editors room block. If making your reservation online, please visit *http://denverregency.hyatt.com/groupbooking/invr*

If you have hotel or general conference questions, please contact Ev Ruch-Graham, sr. conference coordinator, ev@ire.org or 573-882-8969. If you have registration questions, please contact John Green, membership coordinator, jgreen@ire.org or 573-882-2772.

Host: *The Denver Post*

Sponsors: 9NEWS-Denver
Scripps Howard Foundation on behalf of the *Rocky Mountain News*

NOTICE: You will be listed on the IRE Web site as an attendee. If you don't want to be listed, please e-mail John Green at jgreen@ire.org

REGISTRATION FORM.

To register, please complete this form. Visit our Web site regularly at **www.ire.org** or **call 573-882-2042** for the latest details. *Please write carefully! This information will be used to make your nametag.*

Name: _____

Employer/Affiliation/School: _____

Address: _____

_____

City, State: _____ Zip: _____ – _____

E-mail **(required)**: _____

Home Phone: _____ Office Phone **(required)**: _____

To register, mail this form and a check to IRE, 138 Neff Annex, Missouri School of Journalism, Columbia, MO, 65211. To register by credit card, you must have a VISA, MasterCard or American Express. You may fax your credit card registration to 573-882-5431 or register online at *www.ire.org/training/denver05.*

Cancellations need to be sent via e-mail to jgreen@ire.org. There is a $50 processing fee for all cancellations until June 1, 2005. Refunds will not be given for cancellations after June 1.

Card Number: _____ Expiration Date: _____

Card Holder Name: _____

Card Holder Signature: _____

To attend this conference, you must be an IRE member through July 1, 2005. Memberships are nonrefundable.

MEMBERSHIP:
____ I am a current member of IRE through July 1.
____ **$50** I need to join/renew my U.S. or international ❑ Professional ❑ Academic ❑ Associate or ❑ Retiree membership.
____ **$25** I need to join/renew my student membership.

REGISTRATION:
____ **$150** I would like to register for the main conference days as a ❑ Professional ❑ Academic ❑ Associate or ❑ Retiree member.
____ **$100** I would like to register for the main conference days as a student.

LATE FEE:
____ **$25** for registrations postmarked or faxed after May 17.

CAR DAY – optional: Thursday, June 2
(requires additional fee)
____ **$50** Professional/Academic/Associate members
____ **$35** Student members

BLUES BASH – optional: Thursday, June 2 at 8 p.m.
____ **$15** Advance tickets (specify how many tickets)
Tickets sold June 2 at the conference site will be $20.
Tickets at the door - if available - will be $25.

SPECIAL NEEDS:
❑ Do you require special assistance?
Specify _____
❑ Are you a vegetarian?

TOTAL $ _____

MAPPING IT OUT

The latest uses of mapping in news reporting.

Blocking out unsolved murders

By Doug Smith,
Los Angeles Times

Sometimes a database story starts with an idea. Other times it arises from the data itself. In this case, it was both.

Every once in a while, a reporter walks up to my desk with a diskette or a CD-ROM in hand and asks forlornly, "Can you open this?"

I always try to stop what I'm doing and take a look at their data, knowing that the question really means, "Can you find out if there is anything worthwhile on this disk and tell me what it is?"

Most of the time I give a few tips on using Microsoft Excel, and the reporter goes away with one or two nuggets to put into a story that was largely reported by traditional means. But the result was quite different when reporter Jill Leovy, who was covering homicide, showed up at my desk with a CD. Leovy had a specific question. She wanted to know whether the data on the CD would help resolve her suspicion that the Los Angeles Police Department was deploying too few detectives in South Los Angeles where homicides were most prevalent.

Besides hanging out with detectives and spending nights and weekends with the families of homicide victims, Leovy also had built a back-channel connection in the LAPD's Information Technology Division. Through it she had acquired a database of every homicide investigated by the LAPD since 1988. Even though this was public information, my experience with the LAPD was enough to tell me this was a real find. If I had asked for it six months ago, I'd still be waiting for an interview to talk about what format it should come in.

I quickly looked through the table and had to inform Leovy that it wouldn't answer her question. It would be easy to break down homicides geographically, but there was no information on the detectives' identities. We couldn't calculate caseloads.

Leovy looked disappointed and went away. Then I started getting handwritten notes on my desk. Each one had the name of a police division, a list of years and a number for each year. She was going to each of the 18 police divisions to look through old detective rosters. I started keypunching.

When she was done, I did a simple calculation that showed the caseload for every division for every year. It was obvious that detectives working South Los Angeles had consistently been burdened with more cases than those in the Westside or Valley. This provided the statistical support for Leovy's account of hardworking detectives inundated with cases they seldom could solve.

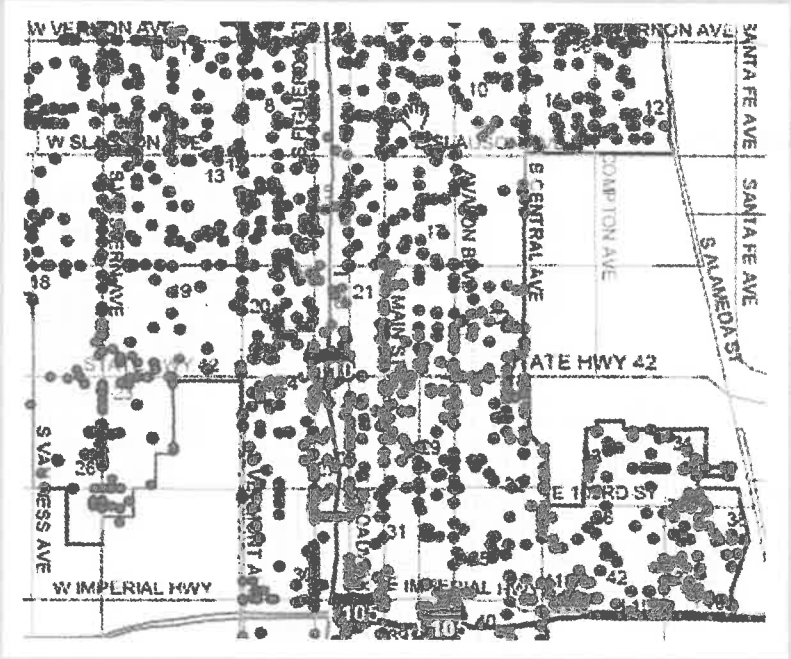
All the while, I was doing my own analysis of the data. The division breakdown showed that homicides were far more frequent in South Los Angeles, but that wasn't news. In 1995, Dick O'Reilly, the then-*Times* director of computer-assisted reporting, did an exhaustive analysis of homicide in Los Angeles County. He established the prevalence of homicides in South Los Angeles and the fact that only about half of homicides are solved. Leovy and I had data covering twice the period in O'Reilly's study, but we had to ask ourselves what was new.

Our concept began to focus on Leovy's experience that the perception and fear of violence are always present in some neighborhoods. I did a little math in my head and realized that with a 50 percent clearance rate, there would be a lot of killers who were never brought to justice roaming the streets of South Los Angeles. We shifted our focus to unsolved homicides, and I turned to mapping to find patterns.

Using ESRI ArcView 8.2, I geocoded the homicides and made a subset of the unsolved ones. Geocoding is one of the most useful tools – it turns street addresses in data tables into point maps. I use it to analyze geographic data for patterns, to present story ideas to editors and to prepare graphics for publication. It doesn't take long to do a first pass that will match 80 percent or more of the addresses in a data table. That's enough to show me whether there is anything worth pursuing.

The first pass yielded startling results. By laying the dots over a street map, I saw that a large area had one or more unsolved homicides on almost every block.

My next step was to produce theme maps to show unsolved homicides as a ratio of population and area of Census tracts. To do this, I had to use ArcView's spatial join function, so I could assign data from one layer to another. These made strong presentation tools when Leovy and I pitched the project to our senior editors. We



got the go-ahead to build a graphic and to pursue the story.

Now that I was working toward publication, it was important to get the highest match rate possible in the geocoding. That meant spending a couple of days using ArcView's interactive geocoding feature to manually match about 2,000 unmatched addresses. In the end, I got about 95 percent of the total.

I also reconsidered the measurements for our story. Not only was I comparing a 12-year accumulation of unsolved killings against the current population, I realized there was little visual or visceral impact in showing homicides as a rate per 10,000 people or per square mile.

It finally hit me that the ideal base was the city block. By shading individual blocks with the number of unsolved homicides that occurred on them, I produced the dramatic pattern that was ultimately published. It was easy to see that a person walking in many parts of South Los Angeles couldn't avoid the scenes of one unsolved homicide after another. No rate was necessary because everyone knows what a block is.

So I switched to using a Census block map and did a spatial join with the homicide points. This added a new field to the block map table with the number of homicides within each block, which I used to create a theme map.

To emphasize the difference between the Valley and South Los Angeles, I gave graphics artist Rebecca Perry blowups of two parts of the city with the individual homicides showing. One had almost no dots while the other had a dense pattern.

I also added an elementary school shapefile and made a table showing the number of unsolved homicides within a quarter mile of each elementary school in the blowups.

The scale was too large to accommodate a complete street grid, which would have delineated the individual blocks. The pattern worked well enough without them except for a few very large blocks in the mountains that run around and through Los Angeles. Some of these are so large they made a single homicide look like a crime wave. We made an executive decision to leave out blocks that are greater than a quarter square mile. We noted this omission in the graphic text.

Once the graphic was done, I went to work on the Web presentation. I wanted readers to be able to see the unsolved homicides in their own neighborhoods. After some experimentation, I broke Los Angeles into map areas to get a small enough scale to show individual dots. Over a couple of days, I made 25 Adobe Acrobat Portable Document Files with the street grid showing lightly in the background. The Web staff turned this into an interactive display that allowed users to click on a master grid of the city and open one of the 25 detailed maps.

There's no way to gauge the impact of these two stories. They generated significant reader response but did not prompt any overt policy change by the LAPD. In fact, the new police chief William Bratton quickly observed on his own that violent crime in South Los Angeles was not getting the attention it deserved. Even before our stories were published, he started tinkering with deployments to beef up the detective strength in overworked bureaus, which meant taking resources away from the San Fernando Valley. This is the kind of reallocation that in the past would have resulted in an outcry from Valley constituents and possibly would have cost the chief his job.

Contact Doug Smith by e-mail at doug.smith@latimes.com

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

SPOTLIGHT: PROBING THE POLICE

Analysis finds 'law in disorder'

By Heather Ratcliffe and Trisha Howard, *St. Louis Post-Dispatch*

As the police department in the small Missouri town of Kinloch crumbled from incompetence, cops began whispering to reporters.

They said many of the other 123 police agencies that patrol the St. Louis area shared the same problems. Resources were spread too thin, causing dangerously inconsistent law enforcement.

The *St. Louis Post-Dispatch* began an investigation to discover why. What we found was far more significant than inconsistency.

In flagrant violation of police licensing laws, untrained auxiliary officers were patrolling alone and answering calls for help. There were troubled officers shuffling amoung poor departments willing to take whomever they can get.

There were broken-down patrol cars and one department with no two-way radios. There were procedures so loose that one small-town cop shot a motorist by mistake and then fired again when the man fled for his life.

There were budgets so tight that police refrained from calling in license checks to a dispatching service. There were departments whose officers' conduct at crime scenes would shock anyone who has watched even a single episode of CSI.

During the yearlong investigation we amassed an enormous amount of data through public information requests. Crime rates, budgets, officer salaries and other statistics were compiled into spreadsheets and compared in graphics topped by a regional map that gave a clear overview of the problem.

Joined by photographers Andrew Cutraro and David Carson, we spent weeks on patrol with more than two dozen police departments, riding along day and night to document the quality of law enforcement at street level.

The end result was a series of articles titled "The Law in Disorder," available on the Web at *www.stltoday.com/police.*

Casting the net

Research began in October 2002 when we requested records from 62 police departments in St. Louis County. Several months later the scope of the story expanded to include the six-county area around St. Louis.

In the spring of 2003, we mailed open records requests to more than 140 police departments seeking such information as their annual budgets, staff lists, salaries paid to chiefs and starting officers, and policies for using deadly force and pursuing fleeing suspects.

Police departments with volunteer officers were asked for additional information. We asked to review three months of duty logs, arrest reports and traffic tickets. We were looking for evidence of uncertified officers enforcing laws and therefore violating state regulations.

Some police departments responded in a few days. Others took as long as three months. At least three departments never complied with the FOIA requests after repeated telephone calls and faxes.

Most of the departments provided the information free of charge. Some mailed it to the newspaper, and others required reporters to pick it up in person and show identification. Some departments charged up to $1 a page in research and copying fees.

None of the departments provided these documents electronically.

One acting police chief told a reporter that he had never heard of his state's open records law. When pressed, he said his staff didn't have time to put together the requested statistics.

He changed his tune when we called the city attorney.

For those departments that never responded, we enlisted the help of *Post-Dispatch* reporters who cover those cities to get the information. They succeeded, demonstrating the value of teamwork in the newsroom.

Data challenges

We discovered inconsistencies in record keeping in the two states in the St. Louis area – Missouri and Illinois. While Missouri officials would not release details of an officer's work history, Illinois could not provide a complete list of every certified police officer in the state.

State and federal court records proved to be extremely valuable in our research. We relied mostly on civil lawsuits to document cases of officer misconduct because police departments would not release such information.

We gathered Uniform Crime Report statistics electronically from the Missouri Department of Public Safety and the FBI. Reporters downloaded the Illinois crime data from the state's Web site.

We quickly learned the flaws of relying too much on self-reported crime statistics. We found that police departments counted their crimes differently and some failed to submit their numbers each year.

We decided to compare only violent crimes to get a fairer representation about the dangers in each community.

We also averaged the crime rate over three years to get a more accurate account.

We used spreadsheets in Microsoft Excel to track the vast amount of information we gathered.

We entered the 5,900 officers into a spreadsheet so we could search the names, track their certification status and make notes on each officer whom we found to be accused of misconduct.

Reporters used the budgets to calculate statistics such as cost per capita, cost per call and cost per officer.

Spreadsheets also helped reporters organize the status of the department's response to record requests.

The information was presented in a series of graphics that included a color-coded map of the per-capita costs. Those, paired with photos of shoddy equipment and poor tactics, provided a compelling presentation.

Reaction from the public was immediate. We received hundreds of phone calls, e-mails and letters expressing outrage about the quality of policing in St. Louis.

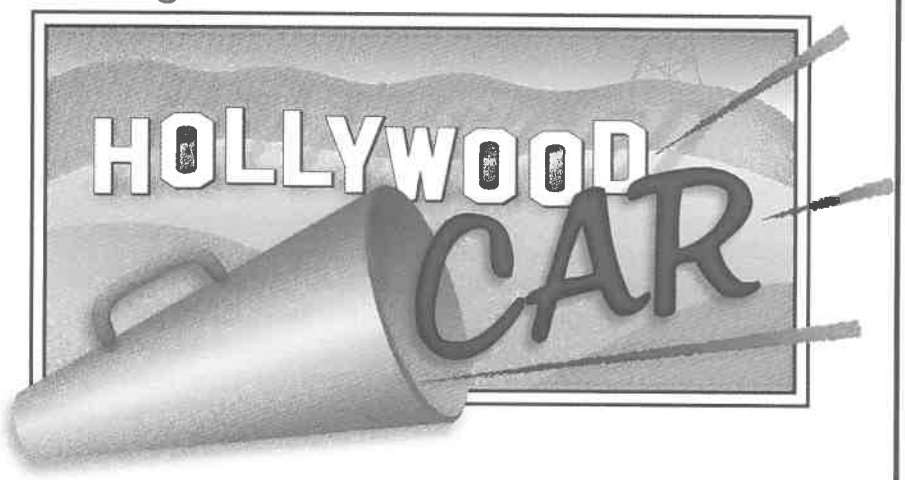
The St. Louis County Prosecuting Attorney launched an investigation into the misuse of volunteer officers. The Missouri State Auditor began a probe into the state program that investigates police misconduct.

Months after the series was published, the newspaper still receives requests from residents and public officials wanting more information. Many have requested additional copies of the stories and graphics to use in their own work.

Contact Heather Ratcliffe by e-mail at hratcliffe@post-dispatch.com and Trisha Howard at thoward@post-dispatch.com.

2005 Computer-Assisted Reporting Conference

Hollywood, Calif. • March 17-20



Don't miss out!

Panels are planned on every newsroom beat and topic in the news. Key topic areas include local government, business, crime, education, health, environment and more. Issues will range from getting the most out of census numbers to following up campaign finance data to understanding federal contracts and international data. Panels will also focus on homeland security, military and infrastructure issues.

PLUS!

Exhibitors, networking, software demonstrations and special sessions for beginners, educators, broadcasters and advanced users.

For CAR beginners, select panels will be coordinated with specific hands-on classes. The combination will give particiapnts a mini-boot camp experience with plenty of opportunity to apply what they learn.

Hands on classes will cover basic to advanced skills in spreadsheets, database managers, using the Internet, building your own database, mapping, statistics, social network analysis, SQL and more.

Hosts:

KNBC-Los Angeles

Los Angeles Times

Sponsor:

The Orange County Register

Conference Hotel:

Renaissance Hollywood Hotel • 1755 N. Highland Ave. • Hollywood, CA 90028

Hotel reservations: **Call 800-468-3571 by March 1**, to get the discounted room rate of $164 plus tax. When making your reservation please ask for the IRE/2005 Annual CAR Conference room block.

Presented by Investigative Reporters and Editors, Inc. and the National Institute for Computer-Assisted Reporting

For more information, registration and the latest schedule, visit

www.ire.org/training/hollywood05

SPOTLIGHT: PROBING THE POLICE

Data shows cops ignore hot-pursuit rules

By Ian Demsky and Christian Bottorff, *The (Nashville) Tennessean*

Just because a police department says it followed all the rules doesn't make it true.

After a presumed hand-to-hand drug transaction in January, Nashville police chased a fleeing suspect through residential neighborhoods into downtown.

At a busy intersection near a high school that was about to dismiss students, the suspect ran a red light and crashed into a car. His car flipped over, and he was killed. The motorist who was hit escaped with minor injuries.

After that chase, we wanted to learn the rules the police have for chases and determine how often they break the rules.

The policy was very clear: Officers may chase only if allowing a suspect to flee "would present an imminent danger to human life or cause serious injury." The guidelines say a suspect's erratic driving, speed and attempts to evade are to be "minor considerations" in deciding whether to continue the chase. In other words, just because a suspect flees, the police do not have license to chase him.

Considering the case in which the suspect crashed downtown, it did not seem there was an imminent danger to life or potential to cause serious injury from the drug transaction.

The department insisted the chase was within policy, but we decided to examine all pursuits from 2003.

We found that 67 percent of the pursuits appeared to have violated the written policy.

Additionally, we found that pursuits of suspects wanted for minor traffic violations led to the most accidents and injuries and were more likely to exceed the speed limit by 30 mph than during other chases.

Building a database

We started with the original paper records kept by the police department. First, we photocopied all 229 of the pursuit report forms and attached narratives generated after each pursuit. The officer is required to fill out the form after each pursuit as a liability precaution to ensure the pursuit met guidelines. The reports are then reviewed by several supervisors.

We started with the original paper records kept by the police department

We created a Microsoft Access database with the fields we wanted to track and entered the data from the paper records. We created fields to track the distance of the chase, the reason it started and the way it ended. Checkboxes indicated whether police, suspects or bystanders were injured or had their cars or property damaged. We also had fields for the speed of the pursuit, traffic conditions and whether the pursuit went through a residential or school zone.

It took a few long nights, but as it turned out, that was the best method we could have chosen.

If we had requested an electronic database from the department, we wouldn't have found the story. The database they kept of chases is entered from the same paper records we used, but it didn't match the paper records. The department's database showed far fewer fast chases and far more slow chases than the records indicated. They couldn't account for this discrepancy. The database lacked the officers' narratives, which we used to correct inconsistencies with the paper forms.

Also, the department failed to fill out the reports consistently. For example, if officers attempted to stop a person for weaving while driving, one form might be marked "traffic," one "misdemeanor" for reckless driving and one "DUI." Or it might have several reasons marked.

We knew that to determine which pursuits appeared to violate the policy, we'd have to come up with our own system to keep track of how the pursuits started. We also wanted to be able to demonstrate that we had given the department the benefit of the doubt whenever we could. For example, if a person was chased for outstanding warrants for which the nature wasn't listed, we assumed the warrants were for violent offenses and categorized it under our category of major life-threatening crimes.

Using information from the narratives and long discussions, we categorized the chases based upon what the suspect was wanted for when the chase began. So if the police discovered that the car had been stolen during a chase for a traffic violation, we counted the chases under "traffic."

Our categories were:

Minor traffic violation – headlight out, speeding, improper tags, running a stop sign, etc.

Major traffic violation –reckless driving, DUI or any indication by the officer that

he or she felt a need to get the suspect off the street to protect the public.

Minor, non-life threatening crime – trespassing, breaking into cars, soliciting a prostitute and drug transactions.

Major, life-threatening crimes

Stolen vehicles: We made this a separate category because, although on the surface it was a property crime, it seemed to fall in a gray area and was often linked with other crimes.

In many cases, the form itself did not match the narratives or was missing information, and in those cases we relied on the narratives to fill in the gaps.

The forms had a checkbox to indicate whether the pursuit had complied with policy. Only seven times did that box say a pursuit had failed to meet the department's policy. One supervisor wrote that he knew the chase didn't comply, but because the suspect was later found to have committed other crimes, he decided it was justified.

The pursuit reporting forms lacked critical detail regarding the duration, distance and top speed of the pursuit. Instead, the form had checkboxes for ranges. For example, for the duration, the officers had the choice of checking boxes for less than one minute, one to five minutes, six to 10 minutes, and so on. So we selected text formats for these fields, and as a result, we could not run the calculations we would have liked.

Also, the way the department recorded how a chase ended was even more confusing. We could not easily tell whether any property was damaged or anyone was hurt.

In several cases, the records indicated there were injuries in the pursuit, but according to the narratives, these happened during scuffles between police and suspects afterward. We did not include these injuries in our calculations.

continued on page 22

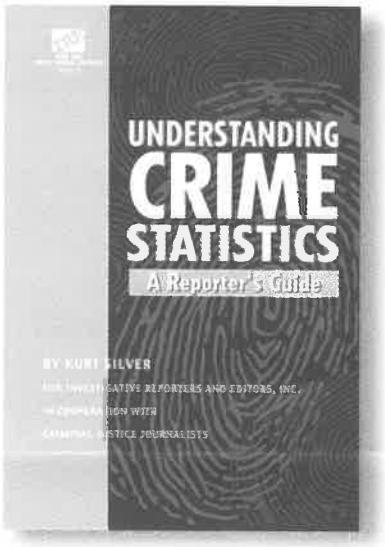
REQUIRED READING For Your Newsroom

Understanding Crime Statistics

A Reporter's Guide, by Kurt Silver

For Investigative Reporters and Editors, Inc.
in Cooperation with Criminal Justice Journalists

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.



BOOK CHAPTERS

Crime by the Numbers
Uniform Crime Reports
National Crime Victimization Survey
National Incident Based Reporting System
Other Major Statistical Sources
Writing the Crime Statistics Story
Database Analysis of Crime Statistics
Web Sites
Resources:Tipsheets and Stories

ORDER TODAY!

Call 573-882-3364 with your VISA, MasterCard or American Express

-OR-

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

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

2005 CAR Training

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.

Boot Camps:
Columbia, Mo.
• March 20-25
• May 15-20
• Aug. 7-12

Mapping Data for News Stories Boot Camps:
Columbia, Mo.
Uncover news stories by mapping data with geographic information system software.
• Aug. 19-21

Boot Camps for Editors:
Columbia, Mo.
Tailored to the needs of newsroom managers.
• April 8-10

Advanced CAR Statistics:
Tempe, Ariz. (ASU)
Aimed at strengthing the skills of reporters who want to move beyond basic CAR.
• Feb. 18-20

For more information:
www.ire.org/training/bootcamps.html

Racial bias

continued from page 1

nation. Fellow Dateline NBC producer Jason Samuels and I were asked to look into the circumstances surrounding these events. After collecting traffic ticketing data and going through court records, I came back to the dispatcher's brief call on the radio: 14 outstanding warrants. How does a 19-year-old rack up so many warrants?

That question led us on a lengthy investigation that culminated with an hour-long report called "A Pattern of Suspicion." With correspondent John Larson, we examined disparities in how police across the United States deal with blacks and whites.

Despite its national scope, the story focused on something simple: a traffic ticket. That's because Thomas' 14 outstanding warrants all stemmed from traffic tickets. Almost all were for the tickets themselves, and the last two were for fleeing from officers for not paying the traffic tickets.

There was also something intriguing about the kinds of traffic tickets that Thomas got. They were not for the usual things, like speeding or running a stop sign. Almost every citation was for an offense that is difficult or almost impossible for an officer to know about without first stopping someone. The list includes driving with a passenger not wearing a seatbelt and driving without a license. Only one ticket was for a moving violation.

Police, prosecutors and criminologists have another way of looking at these tickets. Some, like speeding far above the speed limit or running a stop sign, are also known as low-discretion tickets. These are moving violations, and the offense is obvious even at a distance. Police don't have to make too many decisions before chasing those drivers.

Then there high-discretion stops for driving without a seatbelt or having windows tinted too dark. Notwithstanding special crackdowns, these non-moving violations are harder to detect, and are often ignored by officers.

These are usually ignored unless an officer is looking for something else. In the war on drugs, police and others told us it's perfectly legal for an officer to pull over drivers for these high-discretion violations in the hope of catching them at something worse.

Advocates charge this is essentially racial profiling, violating the equal protection clause of the 14th Amendment and also inflaming already strained relations between police and minority communities.

With Cincinnati's ticketing database, we were able to see Thomas was not alone in getting high-discretion non-moving violations. We obtained the data, with more than 300,000 entries, through a state open records request. It included fields such as the time, street, violation, the driver's name and demographic information like race.

By calculating odds ratios, we could see that blacks were at least three times as likely as whites to get cited for driving without a seatbelt or not having a driver's license.

This kind of ratio is known as an internal benchmark. It relies on information just within the ticketing database. You compare the rate for which blacks versus whites get a high-discretion ticket, like driving without a seatbelt, and divide that by the rate for which blacks versus whites get low-discretion tickets like speeding. For those interested in learning more about odds ratios, Sage publishes several handy books, and Phil Meyer includes a valuable critique in his latest edition of "Precision Journalism."

One advantage of internal benchmarks is that it is an accepted standard in criminal justice circles, and it does not involve population estimation. It is as precise as possible.

Comparing ticketing versus overall population is known as an external benchmark, and among police, its soundness

is more debatable.

We looked at the data other ways. By mapping violations in ArcView, geocoding the locations where tickets were issued, we could see the disparities were not particular to Over-The-Rhine. In predominately white, low-crime neighborhoods, blacks were still ticketed far more often for non-moving violations versus moving violations.

We collected various documents and reports that were useful for starting our data shopping list. A General Accounting Office (now known as the Government Accountability Office) report, among others, outlined which cities with at least 100,000 people across the nation are collecting ticketing and stop data.

We sought data from as many city police departments as possible from coast to coast. This included police departments that had never had anyone request the information – that includes journalists as well as advocates and academic researchers.

The problems we encountered are unfortunately typical: places like San Diego took months and months in replying; Denver rejected our initial request but later turned over the data, and St. Louis charged by far the most. In the end, we had more than 4 million traffic tickets and stops in more than a dozen cities.

One hurdle we faced was that almost no city kept the data exactly like another. In some places, it was impossible to tease out moving versus non-moving violations; those cities were excluded from our analysis. Other police departments recorded stops, and noted the general category of the violation, rather than the precise violation code, that led to the stop. In those cases, we could include the data because the categorizations offered at least a window into that department. As can be expected, this kind of categorization was limiting because it prevented us from examining variations within a grouping – when a police department was pulling over blacks at a higher rate for, say, equipment violations, was the rate even more disproportionate for seatbelt offenses?

In any city that included geographic information we mapped out where the violations occurred. We cross-checked both high-crime and low-crime neighborhoods. In the end, we mapped more than 1 million addresses in six cities, a process that accelerated with programming that improved standardization.

We performed a logistic regression in SPSS to examine the effects of other demographic data on the disparities involving non-moving violations. Logistic regression tests the probability of an event, when there are only two possible outcomes. The results are often reported as odds or odds ratios. This underscored how race was a key factor by allowing us to see essentially how the odds of someone being black is cited for these violations, versus the impact of a person's sex or age.

We constantly re-checked the data and talked to criminal justice experts. We gave each police department more than two months to respond to a written summary of our findings. We wanted to ensure that our work was accurate, and that every police agency had an opportunity to talk with us about what was in the numbers.

The work we did along the way proved extremely useful in these discussions. When one department cited a Bureau of Justice Statistics survey, we could point out that the questionnaire categorized stops in a way that made it impossible to detect moving versus non-moving violations. When another department cited a NHTSA study that blacks did not wear seatbelts about a quarter of the time, we noted that the rate was virtually the same for whites.

Only one police department offered a geographical argument: Richmond said the ticketing disparity was a result of heavy policing in high-crime neighborhoods, which are often black. When we pointed out that the disparity passed through to low-crime, predominately white sections, it responded with "No comment."

For other journalists doing this kind of story, it's important to get the data, talk with sources and get familiar with the subject of ticketing disparities.

Aside from Dateline NBC's investigation, we helped affiliates with local stories. This allowed us to get out more information than what could fit in our piece. We also gave the data and worked with *The Denver Post*, which has an affiliation with NBC. It published a front-page story on the day we aired. And we posted additional reporting on the Dateline Web site at *www.msnbc.msn.com/id/4703833*.

Contact Andy Lehren at andy.lehren@nbc.com. He is a former IRE and NICAR Database Library Administrator.

Dams

continued from page 4

just down the street from the most dangerous dam in Ohio. In addition, I spent weeks talking with nationally recognized experts on dam safety to ensure I had interpreted the data correctly. I contacted Ohio officials early in my reporting, explaining what I was doing. I relied on them for assistance.

Most state dam inspection programs are seriously under funded and under staffed. People in those agencies recognized that a story about years-long budgetary neglect could ultimately help them and those living downstream from the dams.

The Boot Camp gave me the skills to look at data and understand the story possibilities. Also IRE and NICAR provided support over the phone after the Boot Camp. I am hardly a CAR expert, far from it. But, I am challenged by it. After 30 years in print and television news, this is the new horizon.

You can read an account of the broadcast report on the Web at *www.newsnet5.com/news/3911212/detail.html.*

Contact Ron Regan by e-mail at regan@wews.com.

OPEN SOURCE

Setting up, stocking your data warehouse

By Aron Pilhofer, *The Center for Public Integrity*

If you followed the instructions in the November-December 2004 *Uplink*, a MySQL server should be running on your system or network, though you would hardly know it. The only evidence you'll see is an entry for mysqld or mysqld-nt in the Windows Task Manager processes tab.

So how do you run MySQL and work with data?

MySQL is different from a standalone database manager like Microsoft Access. The programs (or clients) that you use to connect to the database engine (or server) are separate.

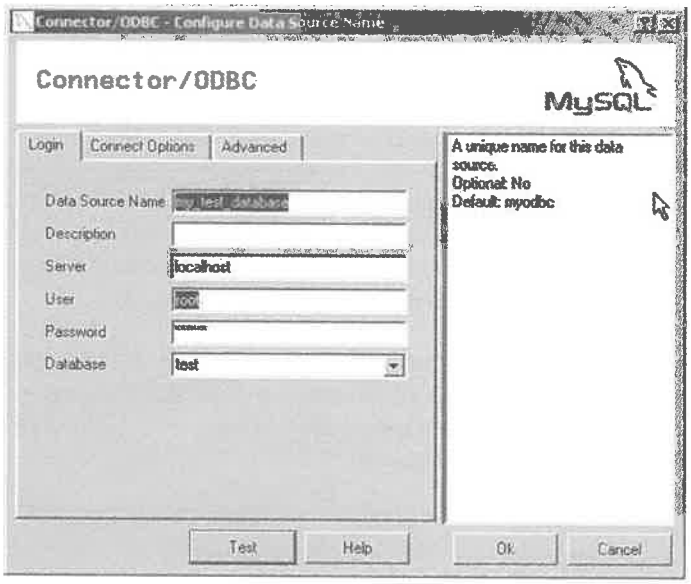
Technically speaking, any software interface used to connect to a database is a client – that includes Web pages or even the DOS command line. In order to set up our newsroom database system, we want something a little more specialized: a client to help manage our server, data and system users.

Some database servers come packaged with an official client (Enterprise Manager, for example, in the case of Microsoft SQL Server), but not MySQL. Due to the explosion of interest in MySQL, there are plenty of excellent options.

Many of the best are free, open source or very inexpensive shareware. If you have the money, Navicat (*www.navicat.com*) is worth it every penny of its $100 price for its unparalleled speed, stability, import filters and feature set.

MySQL will import ASCII text files at mind-bending speeds

We'll continue building our system with a pair of newer no-cost entries – MySQL Administrator and MySQL Query Browser – from the company that makes MySQL. The former program is designed for administering the database and the latter for running queries. Download both from *http://dev.mysql.com*.



After downloading, install the programs and accept the defaults. Launch MySQL Administrator. Enter "root" for the user name and then the password that you selected during the initial installation of MySQL.

If the host field is empty, enter "localhost" if you are working on the computer that's running MySQL. If you're not, type the IP address of the server. The port field should default to 3306, but if it's empty you can enter that. Click OK and you should be greeted with a page showing some basic information about your MySQL installation.

You can use this interface for most of the tasks necessary to keep your MySQL data warehouse up and running. Many of the functions available through the interface are helpful in enhancing the program's performance. That's outside the scope of this article, but you can learn more by reading the reference manual at *http://dev.mysql.com/doc/administrator/en.*

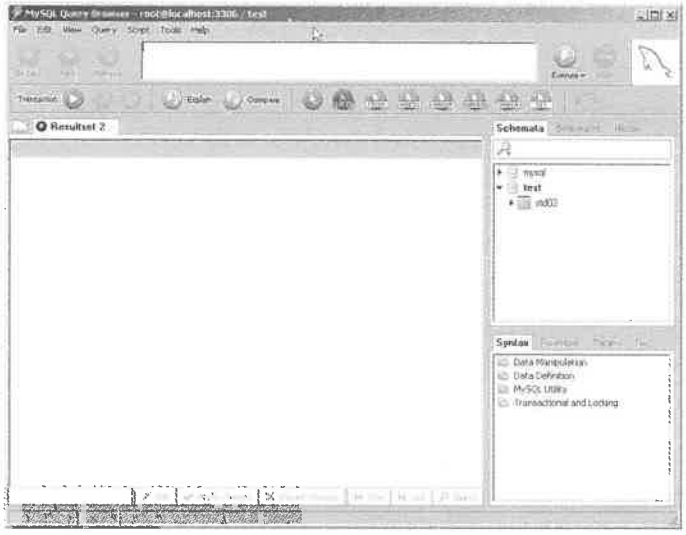
Using this interface, you can do just about anything you need to do to keep your MySQL data library up and running. Most of the functions are designed to help tweak MySQL and get the most performance possible, and much of that is beyond the scope of this article. MySQL Administrator is extremely well documented: *http://dev.mysql.com/doc/administrator/en.*

Security first

One of the big advantages of moving your newsroom data from standalone Access or Microsoft FoxPro databases into a centralized MySQL database system is you can better regulate security. As newsrooms gather more and more and more reporters start doing their own data work, maintaining your newsroom database collection can become a logistical nightmare. More importantly, when anyone can open, change, delete, update and edit data, it becomes impossible to ensure the integrity of the data.

MySQL, like all database servers, allows you to control who has access to which databases, and what they can do with the data. You can give certain users access to only certain databases, specific tables within those databases, or even particular fields.

You can use MySQL Administrator to easily create, manage and administer user accounts by pointing and clicking. (See Chapter 9 of the user manual for



detailed instructions). You can also perform these functions with the MySQL command line (see *http://dev.mysql.com/doc/mysql/en/User_Account_Management.html*). You can create a user-permissions scheme to reflect the expertise groupings among your newsroom staff and assigned shared users names to these groups.

Connecting users

Once you have your user scheme in place, you can connect users with a client program designed for end users. There are a number of excellent client programs available; Query Browser (above) is good for making a more robust connection to the database, though the program is still in early development. In addition, users must run queries using Structure Query Language. (For other options, read "Putting a useful face on MySQL" in the May-June 2003 *Uplink*).

To connect users using one of the many MySQL-specific clients, follow the same procedure you did with Administrator. Enter the host name of the server (usually "localhost," or an IP address), as well as the user name and password – that's it.

But the easiest client for users initially is one you probably already have on your newsroom computers: Access. After going through a simple setup, users can link to data tables stored in MySQL and then write and save queries on those tables just like any other Access table. The only thing users cannot do from Access – and this is a problem with Open Database Connectivity (ODBC), not MySQL – is create or alter a table in any way. That is an operation that must be done through Query Browser or another MySQL-specific client.

To use Access as a front end, first you will need to make sure you have the proper drivers installed for each user who wants to connect to MySQL using Access. Here's how:

• Open Windows Control Panel and go to Administrative tools\Data sources.

• Click on the drivers tab and look for MySQL ODBC driver. If it is there, skip the next two steps.

• If the driver is not there, download it from: *www.mysql.com/products/connector/odbc*.

• Install the driver on all PCs that will use Access to query data stored in MySQL.

Then you will need to configure the ODBC driver:

• Open Windows Control Panel and go to Administrative tools\Data sources.

• Double-click on the Data sources icon and then select either the user (data source name) or system DSN tab (either will work, though you may not be able to save system DSN connections, depending on Windows security settings.

• Click Add and then find the MySQL driver, and double click on it.

• In the dialog box that opens, give your connection a name that your users will be able to understand. Enter the IP address of the server, or "localhost" if MySQL is running on the local machine. Enter the name of the MySQL database that you want to connect to. By default, MySQL installs two databases: MySQL and Test. MySQL is the database that contains all the user, password and login information – never, ever, ever give anyone access to this database. Choose Test.

• Enter the user name and password you would like this user to log in under (one of the accounts we created earlier) and click the test data source button to ensure the connection works. If it doesn't, you might have to go back to Administrator and grant that user account specific privileges to that database, done under the "Schema Privileges" tab in the user accounts section. Once you've successfully connected, your users can begin using tables within the database.

Now the user can use Access' File\Get external data menu item to link to or import the MySQL tables. For example, to import:

• Select Import from the menu.
• Select ODBC databases in the Files of type list that appears in the dialog box that opens.
• Navigate to the DSN list and select the desired data source name

Data in

Now that we've connected to the database, we can put data in it.

There are a number of ways to get data into a MySQL database. The easiest is to export tables from Access, for data tables with 500,000 or fewer records:

• Right click on the table in Access and select Export.
• Choose ODBC as the file export type
• Click OK in the next dialog box
• Select the DSN into which you wish to import

Many of the clients available for download or purchase have data import fil-

continued on page 23

Tech tip...

Learning to play the Access numbers game

By MaryJo Sylwester, *USA Today*

You've created a table in Microsoft Access, and you set the data type as number. But every time you type "1.6," it puts "2" in the field. And every time you type "10.3," it puts "10." So you look at the field properties box in the design view of the table and it says that the decimal place is set to auto. So, doesn't that mean it will automatically accommodate however many decimal places you need?

Unfortunately, no.

Access doesn't play that nice. So here's a primer you can clip out and hang up next to your computer for the next time you're pulling your hair out over number values in Access.

When you create a field in an Access table — particularly for numeric values — there are at least two things you need to set.

• Data Type: Text, numeric, date/time, memo, etc.

• Field Properties: More nitty-gritty stuff, such as how big the number can be and number of decimal places.

You set the data type first. Notice that the field properties section (the gray area below the list of fields) will slightly change depending on what you picked. Figure 1 shows what it looks like for Number.

Pulling down on the "Field Size" drop down list will reveal a list of options. (See figure 2.)

- **Byte**
- **Integer**
- **Long Integer (default)**
- **Single**
- **Double**
- **Replication ID**
- **Decimal**

So which to choose? Read on for various scenarios and guidance.

Scenario One

You need a field to hold whole numbers (no decimals) that will be used for calculations.

Things to consider: How big will the numbers get? Does the size of the database matter?

Options:

• Byte: Stores positive integers (whole numbers) between 1 and 255. Examples: ranking on a limited scale (such as 1 to 10); day of the week (such as 1 for Sunday); day of the month.

• Integer: Stores whole numbers between -32,768 and 32,768. Examples: four-digit years; some sports stats, such as number of touchdowns.

• Long integer: Stores even larger integers, between roughly -2 billion and 2 billion. Use when the numbers would be too large to fit into any of the categories above.

How to choose: If hard drive space on your computer is an issue, use the smallest option that is suitable. If that is not an issue, go with long integer to be sure you'll have plenty of space.

Scenario Two

You have a field with numbers that need one or more decimal places for some or all of the records.

Things to consider: How many decimal places do you need?

Options:

• Single: Use for numbers with up to three decimal places.

• Double: Use for numbers with more than three decimal places.

• Decimal: Use for really big numbers with way more decimal places than we will ever need as journalists.

How to decide: Unless you are a rocket scientist or you're calculating pi, you'll be fine with Single. Keep in mind that Single requires 4 bytes, while Double uses 8. That means that any calculations you do on that field will go twice as fast if you use Single rather than Double. The speed difference may be indiscern-

Figure 1

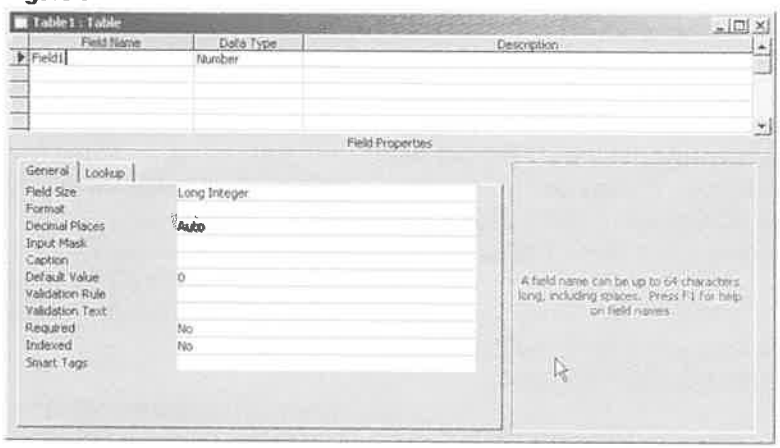
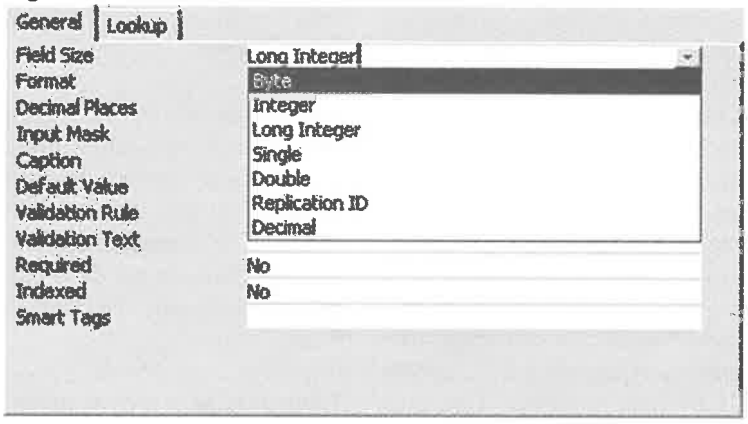


Figure 2



ible to your naked eye, but in this day and age, every second counts.

Specifying number of decimal places: Unlike Excel, the number of decimal places displayed is the same as the number that is stored. So reducing the number of decimal places means your number will forever be rounded — including when you do calculations. In the field properties box, the Decimal Places option will be set to "Auto" when you choose Single or Double. That will accommodate however many decimal places your data has when you import or type in a number. If you want to limit it to two or three, for example, change the "Auto" to one of the numbers in the pull-down list. (See figure 3.)

A word of warning: If you're importing data that contain decimals, you must be sure the field it is going into can accommodate decimals. (If importing a text file, you can set the Data Type to Single or Double in the Import Wizard). If the data instead goes into a plain integer field with no decimal places, you'll lose the decimals.

Scenario Three

You have a field of numbers – no decimal points – and you do not want to do any calculations.

This might be ZIP codes, a unique identifier, or another code, particularly if it has leading zeros.

Things to consider: If this same field appears in another table in your database (especially if it is a unique identifier), be sure that you use the same data type in both tables.

Nothing to decide here: Simply set the Data Type to Text. Then adjust the "Field size" to accommodate the largest possible number of digits.

Scenario Four

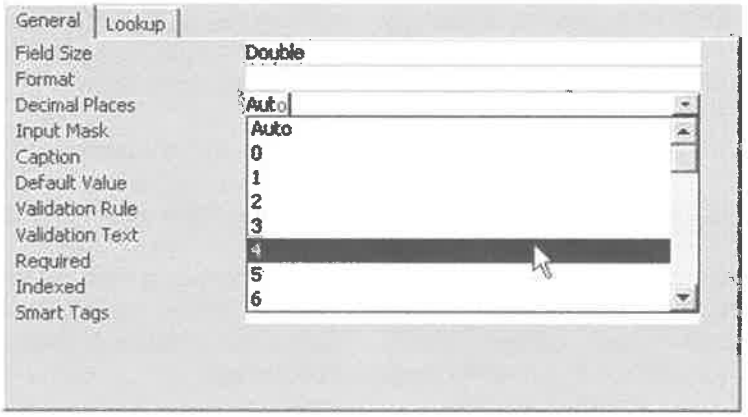
You have special numeric values: dollar values or percentages.

Access makes working with dollars and percentages a bit easier by having pre-set Data Types or Formats that show the symbols as part of the number. Of course, you can store a dollar value as a Number (Data Type) and either Long Integer or Single (Field size) — it just won't have the $ sign and commas.

Setting as Currency: Under Data Type choose Currency and in the field properties box you can set Decimal Places or leave it at Auto (which will give you two).

Setting as Percentage: If you have a field that contains percentages in decimal format (such as 0.50 for 50 percent), then you can change it so that it displays as 50.00%. Set the Data Type as Number and in the field properties box set the Format to Percent. It doesn't matter what you choose for field size because it will automatically set two decimal places.

Figure 3



Contact MaryJo Sylwester by e-mail at msylwest@usatoday.com. She is a former IRE and NICAR Database Library Administrator.

For more tips like this, don't miss Sylwester and Ron Campbell of The Orange County Register speaking about "CAR in everyday life: Tricks of the data trade to get you through the simplest tasks" at the CAR Conference in Hollywood.

Tax break

continued from page 1

just how much tax money schools, cities, and counties lost through these agricultural exemptions.

We also wanted to know which landowners profited the most from the tax break. Were they really farm families for whom the exemption was created almost 30 years ago?

Or were they corporations, developers and land speculators looking to save millions in taxes by leasing the land for farming?

Our goal was straightforward. We wanted to measure the taxable value of the agriculturally designated land and its market value.

Then we had to calculate taxes based on the agricultural value versus the market value to find the total revenue lost to exempt property.

Land records

Our data came from property records from our four main counties: Dallas, Denton, Collin, and Tarrant. In Texas, entities called appraisal districts assess properties for the various taxing entities.

The Tarrant Appraisal District had all of its data available for free download on its Web site. The Denton district sent its data to us on CD. We had previously paid for Collin County data, but the appraisal district didn't charge us for this year's update, downloaded via FTP. Dallas district officials initially said they were going to charge us for our download, but never did.

The format of each district's data varied. Dallas' was in Microsoft SQL Server format. Collin's was in dBASE. Denton's was comma delimited text. Tarrant's initially came with special translation software, but it was too cumbersome so the technician there ended up sending a comma delimited file.

The number of records ranged from about 270,000 for Collin to about 900,000 for Dallas, with a maximum of 30 fields.

The format and content of property records vary greatly from state to state, and even within Texas the format was often different among the neighboring districts.

For example, the field titled "appraised" in one database means the taxable value whereas it means market value in another.

I talked to every database manager before coming up with my methodology to make sure I was using the right categories. After I finished, I outlined my step-by-step process and e-mailed it to them asking them to try to poke holes in it. They responded that my methods were sound.

The calculations were simple. Using Microsoft FoxPro, I created four separate smaller tables from the original large datasets (more than 1 million records in one case). I chose records in which the amount in the agricultural value field was greater than zero.

I created a Microsoft Access database and imported the smaller files so I and the other reporters could work on them.

Cutting losses

For each table, I ran a sum query on the market and agriculture (or taxable) value fields.

The difference between market value and agriculture taxable value provided the value of taxable property exempt or "lost" to the taxing districts. I put these LOST numbers in a separate column. But it's not the amount of lost tax revenue yet. That takes another step.

Each property record in an appraisal district database includes one or several fields to indicate that property's taxing entities. One district had them all crammed into one field separated by commas: CAD, CCS, PRK, JSO. The number of entities varied from one to five, and they were not always in the same order.

Another district had fields labeled COUNTY, CITY, SCHOOL, SPEC1, SPEC2, SPEC3, SPEC4, SPEC5 and the appropriate codes were in each field. One had six columns.

I did a GROUP BY on each entity field and a SUM on the newly created LOST field. That gave me the total value lost by each entity.

I had previously downloaded Excel or text files from the districts' Web sites with each taxing jurisdiction's tax rate. I plugged my LOST totals into those files. The tax rate is per $100 in taxable value, so I divided my LOST totals by $100 and multiplied them by the tax rate for each entity.

The result: The total amount of tax revenue each taxing entity lost this year because of the agriculture tax break.

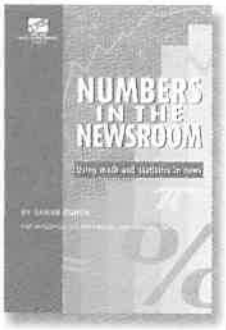
One district in a fast-growing suburb north of Dallas was losing more than $20 million in revenue. Speculators and businesses with valuable land were paying a pittance in taxes because they

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, MasterCard or American Express
-OR-
Visit our Web site at www.ire.org/store
for online ordering or order form downloads

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

either ran a farming operation on it themselves or leased it to a farmer to keep it agriculturally active.

After we saw who the big losers were, we wanted to find the big winners.

Winning big

I started to calculate this by doing a GROUP BY on the taxpayer name field and a SUM on the LOST field. Of course, companies tend to spell their names several different ways. If I grouped by Electronic Data Systems, I missed all the entries for EDS.

I also grouped by taxpayer ID and by the taxpayer address field. After taking three or four swipes at this, I felt confident that I had grouped as many of the largest companies and landowners together as possible.

I struggled to definitively say that a certain proportion of property owners claiming the exemption were big businesses or land speculators. But filtering out company names with a Co., Corp., Inc., etc., wasn't really scientific, plus some of those entries were actually small farm families that had incorporated for tax purposes.

So, I dumped my taxpayer file that I recently sorted into Excel.

My intent was to find an easy way to prove that most – or more than half – the value of the ag exemption was held by major corporations or land speculators.

I started by sorting descending by the companies with the largest amount of taxable value deducted because of the exemption. I did a rolling sum of those values until the total sum equaled more than half the total sum of the exemption values. There were few enough rows that I could go through the landowners one by one to determine if they were farmers or businesses/speculators.

Here's how I did that: Say my last column was J, and J was the difference between market and taxable value. I created another column, Column K, in which I calculated what percentage of Column J's total value was represented by cell J2, the sum of cells J2 and J3, the sum of cells J2, J3, and J4, etc. I was looking for the point at which all the property owners who owned 51 percent of the total value of agriculture land savings were listed.

Through some calculations, and some eyeballing, I determined that most of the value was attributable to large corporations and property development companies. "Most" was good enough for my purposes at this point.

We also decided to map the parcels receiving the tax break using ESRI ArcView 8.3, and the parcel shapefiles the districts gave us for free. That became a nice four-county map showing the outlying suburban land basically engulfed in low-tax, high-value property.

That was the CAR part. The rest involved all three of us interviewing property owners, farmers, appraisers, agriculture industry trade representatives, business representatives, and school, city and county officials.

Despite our numbers showing that big corporations were benefiting from the tax break in the short term, they argued they paid back the value through the rollback. The rollback provision makes landowners pay five years of back taxes plus interest when the land's use changes and it's no longer agricultural. However, they could keep any benefit beyond five years.

Farmers and ranchers were the most vocal proponents of the system. Even though they're not benefiting directly from the tax break, they are able to lease cheap land and keep farming in an area where vacant farmland is disappearing.

You can read the story on the Web at *www.dallasnews.com/s/dws/dn/latestnews/stories/120804dnmetcashcows.312ff.html.*

Contact Paula Lavigne by e-mail at plavigne@dallasnews.com.

readme.txt

Tips for working with property tax data from Paula Lavigne, *The Dallas Morning News*

1. Get the codebooks for your data and have the appraisers explain which codes are still used and what the fields mean.

2. Find out how the system records subdivided land. Does the master parcel's ID number go away, or is it attached to one of the smaller parcels? And how does it measure the change in value when a large parcel dissolves into several smaller ones?

3. If you're measuring the change in value over time, find the best way to join data tables containing other years of data. Watch out for duplicate or missing records, especially when dealing with subdivided commercial or residential lots or with properties that cross city and county lines.

4. Be wary of owner names; different spellings or styles are common. If you want to group or sort by property owner, see if there's a property owner ID or a taxpayer billing address. Sometimes those are more reliable.

5. Be suspicious of numbers that seem too high or too low, and use with medians rather than averages. Errors, exemptions, new construction and other factors can distort your numbers. And make sure you know the difference between how the appraisers calculate the market value and the taxable value.

Voting

continued from page 1

In 2001, after the previous presidential election, *The Wall Street Journal* investigated campaign, inaugural and recount fund contributions by industry. A yearlong series of stories by the *Journal* showed the GOP's top 10 contributors. *Journal* reporters tracked the policy interests of the contributors and found the administration and Congress supported (or opposed) legislation in line with the interests of contributors.

Tom Hamburger, who was the *Journal's* money and politics reporter at the time, says once the data analysis by The Center for Responsive Politics was complete, *Journal* staffers on every Washington beat were involved in the reporting. Says Hamburger, analyzing contributions to the administration and legislators shows the public how the government works and affects lives.

The problem with donation stories, Campbell says, comes from the necessity of joining three databases together. For example, to find out who gave money to which candidates in California, reporters need Federal Election Campaign data, 527 contribution data available on the IRS Web site and data from the California secretary of state.

Checking changes

Campaign data also can help journalists look at structural changes in politics and provide context. "Why is [billionaire George] Soros writing all these checks?" Derek Willis of *The Washington Post* asks. "It has to do with structural changes."

Changes in tactics, laws and procedures affect donors, parties and committees. In October, *The Washington Post* published an article detailing the results of the paper's analysis of 527 data maintained by The Center for Public Integrity. They found six out of 10 donors to 527 groups are billionaires. The McCain-Feingold campaign finance law prohibited soft money contributions to political parties. *The Post* used data to show how the 2002 law affected the process. To make up for the new constraints, donors gave more money to state parties.

Some of the structural changes are so minute, Willis says, that reporters have to understand campaign finance law as well as political science professors and campaign lawyers do. He points to a radio ad that aired before the election. The campaign ad criticized a proposal by Democratic candidate Sen. John Kerry. At one point, words were added to include liberals in Congress. Willis says doing this allowed the campaign to share a cost of the ad with the party.

Changes that might seem insignificant to people are deliberate. "It's got to be the data," Willis says. "You can't see them in meetings making decisions about what to do, but you can see where the money goes."

Structural changes made before the election might not be apparent until afterward. Now is a good time to figure out first what the changes are, and next how did they affect the election. It's best to have earlier data, Willis says. Compare contributions, results and policy to the last election. If power changed hands, fundraising might change. Who shifts allegiances? Check out the new-

readme.txt

IRE and NICAR offer numerous resources for journalists who want to improve their coverage of elections and politics by using computer-assisted reporting.

Journalists can purchase Federal Election Commission campaign contribution data for the 1991-2004 election cycles from the IRE and NICAR Database Library. The data can be purchased by two-year cycle or by a weekly update subscription. The database contains information about candidates and committees, and contributions by individuals and political action committees (PACs).

For more information about the FEC data see *www.ire.org/datalibrary/databases/feccc*. Contact the Database Library at 573-884-7711 or 573-884-7332.

Working with campaign finance data can be challenging. The IRE Beat Book, "Unstacking the Deck: A Reporter's Guide to Campaign Finance," provides detailed information about working with state and federal data. The book is written by veteran journalists Michael A. Weber, Aron Pilhofer and Derek Willis, and is available from IRE. For more information or to order see *www.ire.org/store/books/campaign.html*.

The IRE Resource Center has tipsheets and news stories for journalists who want to learn more about campaign finance and elections reporting.

Willis offers guidance for reporting on political nonprofit organizations (527 committees) in tipsheet No. 2210.

Tom Hamburger of *The Wall Street Journal* provides a list of Web sites useful in covering money and politics (Tipsheet No. 1913).

Wendell Cochran of American University in Washington, D.C., provides detailed advice for working with the FEC's online data (Tipsheet No. 2018).

Staff writers at *The Wall Street Journal* examined how contributors to President Bush and the GOP looked for political favors in 2001, the year after Bush's election. (Story No. 18798)

To order tipsheets and stories contact the Resource Center at 573-882-3364 or rescntr@ire.org.

comers in the legislature. "Okay, they got in – now what?" Willis asks. Know who has campaign debt.

Know not only how much money was spent, but know where the campaign put money. Willis suggests creating timelines in Excel or Access of when and how campaigns spent contributions. Even if the end result doesn't make it into the paper, knowing this information will help during the next campaign.

Examining voters

Journalists can also use voter registration data to check up on registration and voting patterns.

The *Sarasota Herald-Tribune*, like a number of other newsrooms, has a database of historical election results and Census data. Adding the 2004 results would yield a database that could show shifts in demographics and an increase in voter registration. "We could spend a few months with that database we have," says Matt Doig, a reporter at the paper. Florida reporters could follow up on absentee ballot stories by using geographic information system software and absentee-ballot data to find out how many "snow birds" voted and where they live the rest of the year, he says.

And of course, any analysis done for the 2004 elections will add context to 2006 and 2008 federal elections. If you took a breather from the red and blue state madness, now is a good time to get back; the data is waiting.

Contact Andrea Lorenz at arlp54@mizzou.edu.

Journalists reporting on campaign finance and elections can find data and other useful resources at these Web sites:

IRE's Campaign Finance Information Center (*www.campaignfinance.org*) offers links to recent campaign finance stories.

The Federal Election Commission makes its campaign finance data available for download (*www.fec.gov/disclosure.shtml*) and provides images of financial reports.

Journalists can search for individual contributions to political nonprofit groups (527 committees) at the Center for Public Integrity Web site (*www.publicintegrity.org/527/db.aspx?act=main*).

The Center for Responsive Politics (*www.opensecrets.org*) and PoliticalMoneyLine (*www.fecinfo.com*) also provide federal campaign contribution information on line. The Institute for Money in State Politics provides information about state campaign finance at *www.followthemoney.org*.

readme.txt

Extra! Extra! on the IRE Web site (*www.ire.org/extraextra*) carries summaries of and links to recent election and campaign stories that used computer-assisted reporting. Here are some examples:

The Seattle Times analyzed voter registration data and found that nearly 130 felons in two counties voted illegally in the 2004 general election. *The Times* also found that election officials had failed to purge felons from the voting rolls or had failed to check whether newly registered voters had felony records.

A *South Florida Sun-Sentinel* analysis of Florida's 2004 election results found that counties with touch-screen voting were more prone to error. The newspaper reported that voters using touch-screen voting machines were 50 percent more likely to make an error, compared to voters who used a voting machine that used paper ballots.

The Boston Globe Ideas section carried an analysis of the national presidential vote and examined the regions where President Bush and challenger Sen. John Kerry drew their support. The reporting was based on continuing geographic information system (GIS) data analysis by *CommonWealth Magazine*.

Just before the election, *The Miami Herald* reported that an overwhelmed clemency system was blocking felons from having their voting rights restored. The analysis found that nearly 40,000 felons released from custody from 2001 to 2003 were unable to vote.

REQUIRED READING For Your Newsroom

Unstacking the Deck:
A Reporter's Guide to Campaign Finance
by Michael A. Weber, Aron Pilhofer and Derek Willis

This invaluable guide for pursuing stories about the impact of money on elections, political parties and candidates at the federal, state and local levels is packed with story tips, resources and Web sites.

ORDER TODAY!
Call 573-882-3364
with your VISA, MasterCard or American Express
-OR-
Visit our Web site at **www.ire.org/store**
for online ordering or order form downloads

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

Pursuit

continued from page 11

The data the department later gave us did.

Two things we found extremely helpful: a memo field in which we could note brief highlights, interesting examples or questions we had and unique identification numbers for each chase, which we also wrote on our copies of the paper records. Those ID numbers allowed us to easily move between the data and the records. (They also helped us reconstruct our data when we deleted something from a query and accidentally killed information in the underlying table.)

We also kept track of the gender and race, finding that 63 percent of those pursued were black, but our editors decided not to a purse a racial angle on the story because they thought it would take away the punch from our story about policy violations.

Later, we used ESRI ArcView 3.2 geographic information system to map the starting locations and times of the pursuits. We saw that the chases occurred more often at night in areas with higher-than- average black populations. The police explained that this was because those locations had the highest concentration of officers.

About a week before the story ran, we presented our methods and findings to the police department. At first, the recently hired chief and support personnel came back with their own data that said far fewer chases violated policy. But their data turned out to be inaccurate because of department data-entry errors. After some discussion, they agreed that according to the way we classified our data, it showed that two-thirds of the pursuits didn't meet the threshold of imminent danger listed in the policy.

Our reporting – from photocopying the records with a portable copier to writing the story – took about two weeks.

After the story ran, the new police chief announced changes to the pursuit policy. He said he had been looking at it before our analysis.

The new policy didn't change the justification for chases, but it gives supervisors the responsibility to authorize each pursuit and grant continuing permission as the conditions of the pursuit change.

In the months since that report came out, we hear the change when we listen to the scanner. An officer will radio in that he or she is behind a car that's not stopping. The supervisor will ask what the suspect is wanted for. Many times, if it's minor, the supervisor will tell them to stop. They are also quick to stop pursuits if the speeds start getting too high or traffic conditions change.

Contact Ian Demsky by e-mail at idemsky@tennessean.com and Christian Bottorff at cbottorff@tennessean.com. Demsky is a former analyst for the IRE and NICAR Database Library.

readme.txt

IRE and NICAR offer resources for journalists who want to use data to probe law enforcement and crime.

"Understanding Crime Statistics: A Reporter's Guide," by Kurt Silver is an IRE Beat Book that provides detailed guidance for analyzing crime report databases. The book covers use of the FBI's Uniform Crime Report data and the National Crime Victimization Survey. Order the book online at *www.ire.org/store/books/crime.html* or call 573-882-3364.

Journalists can purchase Uniform Crime Report and U.S. campus crime data from the Database Library. UCR data includes data on index crimes – murder, non-negligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny-theft, motor-vehicle theft and arson – by law enforcement agency, and detailed information about homicides. For more information see *www.nicar.org/data/fbi*. The Database Library has the data for 1980, 1990 and 1993-2002.

The campus crime data from the U.S. Department of Education contains **information about alleged criminal incidents reported to campus or local law enforcement agencies. The data covers 2001-02 and includes data about arrests that result in campus disciplinary proceedings. For more information see *www.ire.org/datalibrary/databases/campus*.**

Contact the Database Library at 573-884-7711 or 573-884-7332 to order the data.

Tipsheet No. 1196 by journalists from *The Philadelphia Inquirer* provides advice for using the Uniform Crime Report data to examine whether police are underreporting rapes.

John Archibald of *The Birmingham News* lists resources for journalists who want to use crime and justice data for looking at law enforcement (Tipsheet No. 2056).

Karisa King of the *San Antonio Express-News* provides guidance for examining traffic stop data to detect possible racial disparities. (Tipsheet No. 2094).

Order tipsheets by contacting the IRE Resource Center at 573-882-3364 or rescntr@ire.org.