Uplink

November • December 2002
Volume 14, Number 6

Published bimonthly by the National Institute for Computer-Assisted Reporting
www.nicar.org

DATA ACCESS

# GIS data carries high cost

By Jaimi Dowdell,
*IRE and NICAR*

For three years, David Washburn of *The San Diego Union-Tribune* has been itching to get his hands on a file from SanGIS, an organization that handles San Diego County's GIS data. Although the GIS data is available to the public, neither Washburn nor his newspaper is ready to fork over $15,000 for one year of data.

"I think my county is ripping me off and the taxpayers off how they are handling GIS," Washburn said.

Within the past five years, use of GIS software and data has exploded in newsrooms. Many journalists using computer-assisted reporting have found the programs essential and use them in their daily reporting. While many government agencies have posted GIS data files on their Web sites for free download, other agencies haven't been as open.

*continued on page 17*

**SPOTLIGHT: CRIME & JUSTICE**

# Mapping finds bars, DUI wrecks linked

By Rose Ciotta and Karl Stark, *The Philadelphia Inquirer*

Drunken driving was old news. It had been written about and rallied against for years. Yet, this year, for the first time in nearly a decade, drunken driving became a more deadly crime than murder. After years of decline, drunken driving's death toll was inching up nationally and in New Jersey, and holding steady in Pennsylvania.

Those facts gave us ample reason to investigate drunken driving in the Philadelphia region. But, we faced the challenge of finding a fresh approach. Mapping and density analysis gave us the tools to dig deeper and come up with new information for our series, "Loaded for Trouble," which ran in June.

From researchers we learned of cutting-edge work that found a link between the concentrations of liquor licenses and the locations of alcohol-related accidents.

We set out to test the connection in the Philadelphia region. Our first problem was getting the data. In Pennsylvania, accident locations are a secret. Attorneys for the state cite a law that protects the state from liability claims. In addition, Pennsylvania's Department of Transportation several years ago removed from its public data on accidents just about every field of inter-

*continued on page 16*

## SPOTLIGHT:

This issue showcases some of the best uses of CAR to dig into crime and justice. Reporters discovered that the failure of some Texas judges to grant speedy trials led to the release of thousands of suspects, that there was racial inequity in Pittsburgh juries, and the correlation between drunken driving accidents and liquor license locations.

# Data boosts crime stories

By Megan Christensen,
*IRE and NICAR*

For the resourceful journalist using computer-assisted reporting, there are as many great criminal justice stories waiting to be written as there are Law & Order reruns on cable.

Stories about local and national crime trends often begin with the FBI's Uniform Crime Reports. The data is compiled from voluntary reports submitted by city, county, state and national law enforcement agencies. Roughly 95 percent of all police agencies report

*continued on page 14*

## Bits & Bytes

*Recent updates from the Database Library*

**Vehicle accidents (updated)**
How many people died in accidents on U.S. roadways last year? What are some of the more dangerous highways? Questions like these can be answered with the Fatal Accident Reporting System (FARS) database from the U.S. Department of Transportation. Records are available from 1975, when the DOT started tracking fatal accidents, through 2001. The database has three main tables to track by person, vehicle or accident. With this database, it's possible to find out many details, such as how fast a vehicle was traveling, weather and road conditions, and whether the driver had prior convictions for DUI or speeding.

Another database deals specifically with truck crashes, whether or not anyone died. The DOT Truck Accidents database has commercial carrier accidents as reported to state and local police from 1988 through 2001. The main table has details on the crash, such as number of injuries and vehicles involved, plus the name and location of the carrier. The database also has the truck's DOT census number, which makes it possible to link the crash database to truck inspection or truck census databases, also available from the Database library.

**Aviation data (updated)**
The Aircraft Registration Database has information on more than 320,000 planes, in-

*continued on page 4*

NICAR UPDATE

# Another member of the family moves on

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

As some of you know, Aron Pilhofer, who directed IRE's Campaign Finance Information Center, has taken a job as database editor with the Center for Public Integrity in Washington, D.C.

For more than a year, Aron had been busy with IRE and NICAR projects. He revamped the CFIC Web site and launched its two online databases where IRE members can search for federal political campaign contributions and simply click on the contributor's name to see whether the contributor received federal contracts.

Aron revived *Tracker*, the center's quarterly publication that covers campaign finance reporting, and trained journalists across the country in how to analyze campaign finance data for news reporting.

Aron served as NICAR's in-house open source software guru, promoting the virtues of free software tools with all kinds of strange names – Python, Mandrake, MySQL, TUTOS. When Aron spoke at the annual CAR conference in Philadelphia this year about "World-Class CAR on 99 Cents a Day," he wasn't kidding.

Starting with the next issue of Uplink, Aron will contribute a regular feature focusing on open source software that journalists can use for their work. For starters, he'll write about his experience with OpenOffice, the open source alternative to Microsoft Office that's been garnering some interest on the NICAR-L listserv.

Best wishes in your return to the newsroom, Aron.

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

## Upcoming CAR, census training

IRE and NICAR has numerous training opportunities in the coming months for journalists who want to learn computer-assisted reporting and data analysis for reporting on Census 2000. For a complete list of training events visit *www.ire.org/training/otr.html*.

Here are a few of the many training highlights:

The Jan. 5-10 computer-assisted reporting boot camp in Columbia, Mo., provides hands-on training in how to use spreadsheets and database managers to analyze data for high impact stories.

The Jan. 11-12 Census reporting workshop in Columbia, Mo., will offer a track for beginners and another for journalists with intermediate census reporting skills. Paul Overberg of *USA Today* and Steve Doig of Arizona State University will lead the classes. Hands-on sessions are included.

A Feb. 21 census workshop at Arizona State University in Tempe is intended to help beginners learn how to find stories in the data that's already been released from Census 2000. It offers a good introduction for beat reporters, graphic artists or assignment editors who need a focused, concrete introduction to demographic data and computer-assisted reporting.

CENSUS 2000

# Covering income, commuting with SF3

By Jeff Porter, *IRE and NICAR*

The flood of Census data is over. For now.

U.S. Census Bureau servers, like journalists across the U.S., worked overtime for two months as the Summary File 3 became available, state by state, in August and September.

IRE's Census-L listserv buzzed with questions and stories. Journalists communicated by code, it seemed. Sample question on the Census-L: "Are there numbers for median household income by race for 1990? I'm looking in STF3 and not seeing such a table. Anyone?" Answer: "Table P82."

Incredibly, most Census-L subscribers understood that answer. It refers to a specific table released after the 1990 Census.

The SF3 data came through as a series of tables, derived from the Census Bureau's April 2000 long-form survey, sent to about one of six households. With that, the agency took the responses and crafted summaries, released in 76 data files for each state, plus a file containing geographic information.

And while the data can – and will – keep reporters busy for months to come, there are stories aplenty to examine and learn from.

## The gender gap

*The Detroit News*, for example, looked at the slow improvement in the "gender gap" – the difference between what men make and what women make in the same jobs.

Brad Heath conducted data analysis for the newspaper. A big challenge, he said, was not to go too far.

"The biggest problem was that the data stopped just shy of giving me the smoking gun I was looking for," he said. "The data the Census Bureau provided show inequality to be sure, but they can't get you far enough to show actual bias."

Still, *The Detroit News* could report that while every state has made progress in closing the gender gap from 1989 to 1999, Michigan lagged. In that state, women who worked full time made 67 cents for every dollar earned by men, a figure that budged less than 5 cents in the 1990s. Nationwide, women who worked full time in 1999 earned 73 cents for each dollar men earned.

Like almost any database, though, the Census data didn't answer the question: Why?

"The census made clear that women and men work at different jobs and have different educations, both of which clearly play a role in the earnings disparity," Heath said. "But there wasn't enough there to come up with a solid answer to why Michigan ranks so low in pay for women. That was frustrating. And because the Census Bureau changed some of its job categories in the past decade, it wasn't really possible to show how well women are making inroads into higher-paying fields here."

Complaints about comparing 1990 and 2000 data items were common headaches expressed on Census-L traffic for more than a year.

## Rising black income

Dwight Worley of *The Journal News* of White Plains, N.Y., encountered challenges in reporting on the rise of the black middle class population in New York in general and *The Journal News*' circulation area in particular.

One worry: error margins. What might look like a big change or trend could turn out to be statistically insignificant, depending the error margins or how much of the data is imputed – that is, estimated – by the Census Bureau.

"So you have to be careful – especially when you're comparing data over several decades," Worley said. "That's why we stuck with using state and county-level data, which have lower error margins than town or place level data."

To make those several-decades comparisons, Worley looked at income, education, employment, poverty and home ownership factors, from the SF3 in 2000 and from either Summary Files 3 or 4 from earlier years. He perused PDF files on the bureau's Web site for 1990 SF4 information. Going back to 1980 and 1970 was even more difficult. He finally went to government document libraries in New York City and Westchester County to go through books and microfiche images and just plugged in numbers by hand.

Then there's the definition of a household vs. a family. Here's how the Census Bureau divides the two groups: A household includes all of the people who occupy a housing unit. A family includes a householder and others, living in the same household, who are related to the householder by birth, marriage or adoption. Households, then, may include more members than do families. And not all households contain families, since a household may be composed of a group of unrelated people or of one person living alone.

There are tables that break down the number of workers by family type and the aggregate income of each type, but not by race. Household income is broken down by race, though, and counts everyone – not just those who constitute a family.

But, Worley said, even that can leave questions. "If there are, on average, four workers in a black household now when there used to be only two in 1970 then that could seriously alter the analysis of rising incomes. In that instance,

*continued on page 4*

## Bits & Bytes

continued from page 2

cluding those owned by businesses, individuals and government. Current through May 2002, the database allows journalists to find the owner of an aircraft by looking up the registration number (commonly called the "N number"). This comes in handy for covering plane crashes, especially if someone can identify the N number on the plane's tail. The data also provides the plane's serial number, which is useful when searching the FAA Service Difficulty Reports database, also available from NICAR, for reported mechanical problems.

Find out about accidents and other incidents involving airplanes in the United States with the FAA Accidents and Incidents database, from 1990 through August 2002. The database contains reports of crashes, collisions, deaths, injuries, major mechanical problems and other events, from small private planes to large commercial carriers. Another database available from NICAR has accidents and incidents from 1973 to 1989.

The Database Library offers other transportation databases, including vehicle recalls and complaints; accidents involving hazardous materials and the NASA Air Safety Reporting System.

To order or find out more about these databases, visit *www.ire.org/datalibrary* or call 573-884-7711.

# Income

continued from page 3

household income is rising, but it could be because there are more workers in the house, not necessarily because individual blacks are earning more money."

### Race, ethnicity, money

*The Miami Herald* found a rise in black income in Florida. That newspaper used median income by race and ethnicity to make a comparison down to the level of places that have a population of at least 10,000.

The Census Bureau provided a 1990 table of those places and median income.

Said *The Miami Herald*'s Tim Henderson: "That was fine as far as it went, and showed us that there was a large increase in black income in Florida, but it wasn't reflected in our urban county (Miami-Dade) — only in our suburban county, Broward. There, black income soared and reached 81 percent of white non-Hispanic income, better than anywhere else in the state."

One of the problems of using only places and counties, he said, is that it could mask changes in areas that weren't marked by municipal boundaries.

"We could see this clearly when we mapped tracts by black income, screening for tracts with a significant black population [at least 10 percent]," he said. "A new regional area of black affluence stood out clearly at the western border of the two counties, and the ethnic makeup of those tracts made it clear that Jamaicans were playing a role in that new affluence. African-Americans also dominated in the adjoining Dade County tracts."

Those points made a story about the region's ethnic chemistry — a positive note in light of the recurrent criticism that the region has no black middle class to provide hope and support for the poverty-stricken inner cities, Henderson said.

### Commuting with CAR

Any movement from the middle of a city means an increase in the number of commuters and the time they spend on the roads. Matt Waite followed commuters in Pasco County, Fla., for the *St. Petersburg Times*.

Economic demographics play a large part in the commuter game.

"Pasco County is made up of three distinct areas: west, central and east," he said. "The east is the same as it was 40 years ago — rural. Central is the rapidly developing area of $150,000-plus homes. West is where all the retirees moved in the '60s and '70s, and they are all now dying off, leaving their $40,000 to $60,000 homes behind. The people moving in are 20- or 30-something, with kids."

Those moves, of course, mean an interesting mix of commuters.

"You have a population of people buying low-cost homes who have to commute long distances, which is counterintuitive," Waite said. "You expect more well-off people driving a ways to their downtown offices. But you don't expect to find an area with a $23,000 median household income driving an average of 30 minutes to work every day."

And, like the gender gap or income stories, the data can't tell the whole story. In Waite's case, the big question is where the commuter is going. The data can only tell where the commuter lives.

Now journalists are awaiting the arrival of still another data release – called Public Use Microdata Sample, or PUMS. Beginning with a release tentatively set for December, PUMS data is a sample of the actual responses. With that, journalists can go beyond the Census Bureau's cross-tabs and build their own summary tables.

Contact Jeff Porter by e-mail at jeff@nicar.org.

ELEVATOR SAFETY

# Access forms help build inspections database

By Mike Sherry, *CQ Weekly*

Elevators have become as much a part of everyday life as cars.

But while journalists investigate many facets of vehicle safety, such as fatal accidents and faulty equipment, there is precious little reporting about elevators – even though they carry millions of people every year.

So, when I came across the Missouri Elevator Safety Board on the state Web site, it piqued my interest. If the state was worried enough about elevator safety to appoint a board to oversee it, then perhaps a story was there.

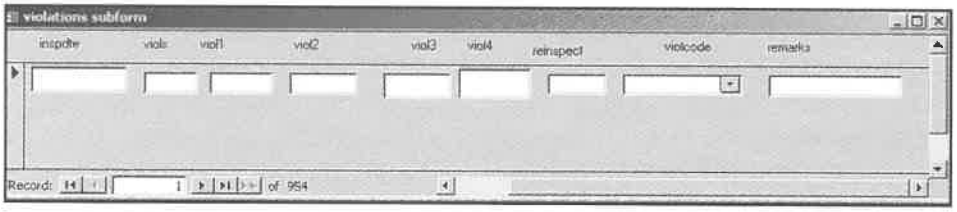
The result was a computer-assisted reporting story – written as part of my master's degree final project at the University of Missouri School of Journalism – that ran July 7 in the *St. Louis Post-Dispatch*. Given that the United States has an estimated 700,000 elevators, this is a story that can be replicated in just about any community.

Some of the unique challenges of this story forced me to learn important CAR lessons that others can benefit from.

## Designing databases

Using Access 97 and Access 2000, I grappled with database design for a relatively small database. It had many tables, but the largest one only had about 20,000 records. Everyone needs to think about database design when information must be entered from paper records.

The state Elevator Safety Unit had an extensive Access database that included a lot of data, such as inspection dates and elevator locations. But the unit was so understaffed that the staff had no time to enter all the violations uncovered by inspectors. The violations were detailed in color-coded folders (with a different color for every month) that contained more than 10,000 reports. I took a random sample of about 260 reports that had an average of about four violations. This gave me a good idea of the types of problems inspectors found.




*A form created in Microsoft Access was used for entering violations data from elevator inspections.*

Next, I created an Access form for data entry using a Tech Tip in the April 2000 *Uplink* as a guide. (*Uplink* articles from 1990 to the present are available to IRE members at *www.ire.org/resourcecenter*.)

The problem, however, was that inspectors did not record violations consistently. Some inspectors referenced particular sections of the code book, some just jotted down notes, and some did both. I needed consistency, however, so I could use Access to group by and count the records to identify the most common violations.

## Learning lookups

So I created a lookup table, where I coded each type of violation with a number. For instance, machine room door problems got a code of "3." I linked my data-entry form to that table, which gave me a drop down menu each time I needed to enter a particular violation.

After running the queries, I found that many elevators had serious problems like broken phones. Machine room doors that didn't close and lock on their own were also hazards because someone could walk in and be hurt or damage equipment that could harm passengers. This was the first time anyone – including agency staff – evaluated the seriousness of the problems.

Setting up the violation codes solved one quandary, but another remained: how to link the violations table with the activity table that the Elevator Safety Unit had created to keep track of inspection dates and other information.

Each elevator had a unique state ID number, which I entered in my violations table. The problem, however, was that each elevator often appeared more than once in the activity table – mainly because the state entered a record when that elevator was registered, then created separate records for the subsequent inspections. Meanwhile, each one of my violation records included the equipment ID, so if an elevator had four violations,

*continued on page 8*

# MAPPING IT OUT

*The latest uses of mapping in news reporting.*

## Mapping the census

The U.S. Census Bureau's latest big data release – long-form questionnaire results reported down to the tract or block group – is keeping journalists busy as they report the latest economic, social and demographic trends. (See article on page 3) Many of the journalists have been using geographic information system (GIS) programs to uncover patterns and changes. Here are some recent examples from newsrooms across the country:

### Language barriers

Ron Campbell at *The Orange County Register* used ArcView 3.2 and Census long-form data to find areas with the highest concentrations of residents over 14 who did not speak English fluently. The map he generated (Map 1) displayed the tracts with the highest non-English speaking rates using darker shades.

One Santa Ana neighborhood – with 57 percent of its population over age 14 not speaking English – had the highest rate. Another reporter visited the neighborhood and found residents who routinely spoke Spanish at home, church and inside stores.

A demographer quoted in the story noted that the neighborhood, like some others, is the point of arrival for new immigrants who find it easier to adapt by living among people who speak the same tongue.

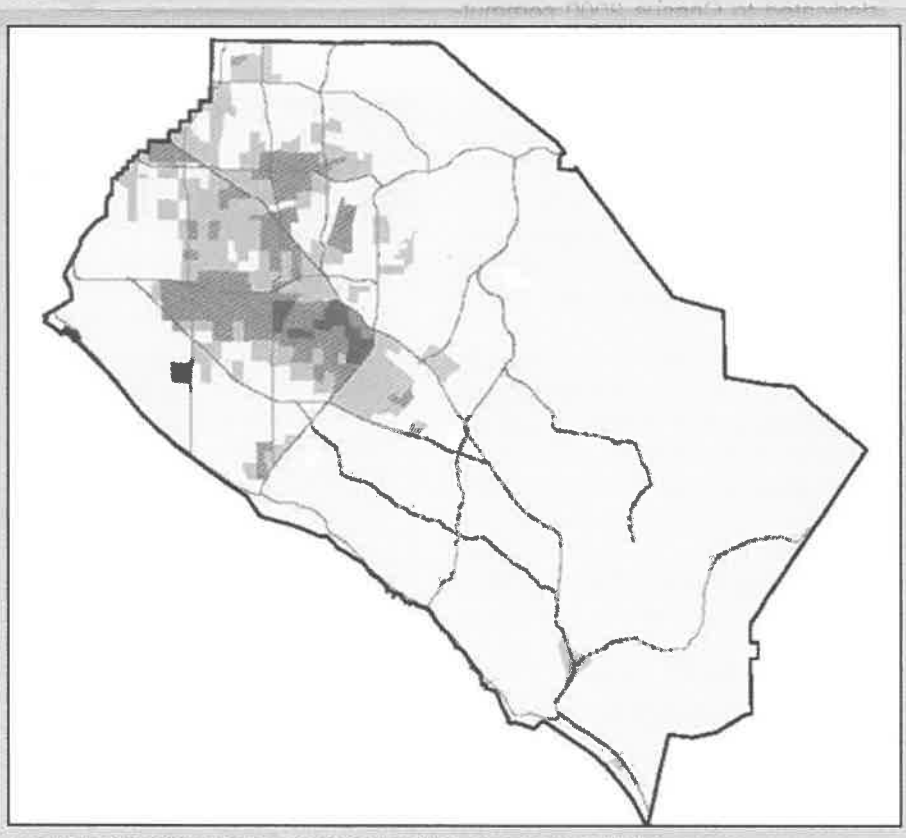
Some residents described the tension between generations: Parents wanted their children to speak Spanish, while their children wanted to learn English.

### Rich and poor

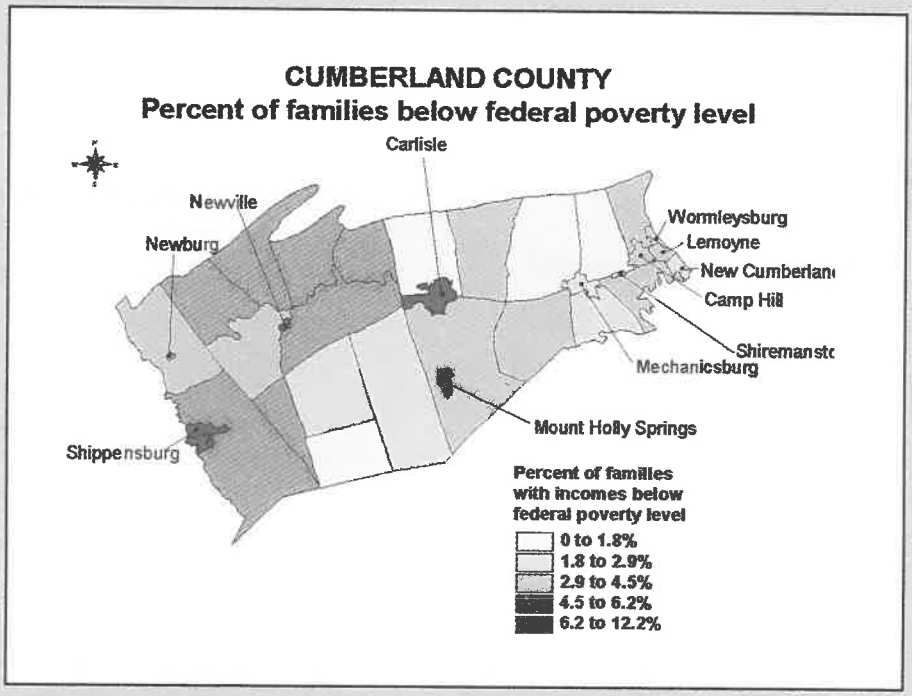
Mapping helped Garry Lenton, the CAR coordinator at *The* (Harrisburg, Pa.) *Patriot-News,* find pockets of poverty in Cumberland County. The county has one of the highest median incomes in the Harrisburg region, yet it had the third-highest poverty rate increase among Pennsylvania's 67 counties.

Lenton wanted to learn the extent of poverty in the county's municipalities. After posting a request for advice to the NICAR-L listserv, Lenton settled on mapping the percentage of people meeting federal poverty guidelines by community and running the numbers of families in poverty by municipality within the newspaper's graphic. (Map 2)

Mapping by percent revealed that some of the sparsely populated communities, mainly in the rural western portion of the county, suffered the highest poverty rates. Reporters for



**Map 1:** *The Orange County Register mapped long-form data by census tract to highlight areas where residents did not speak English fluently.*




**Map 2:** *GIS helped The* (Harrisburg, Pa.) *Patriot-News see pockets of poverty inside a county with a high median income.*

*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.***

the paper later found that a boom in the number of low-income housing units had helped fuel the increase.

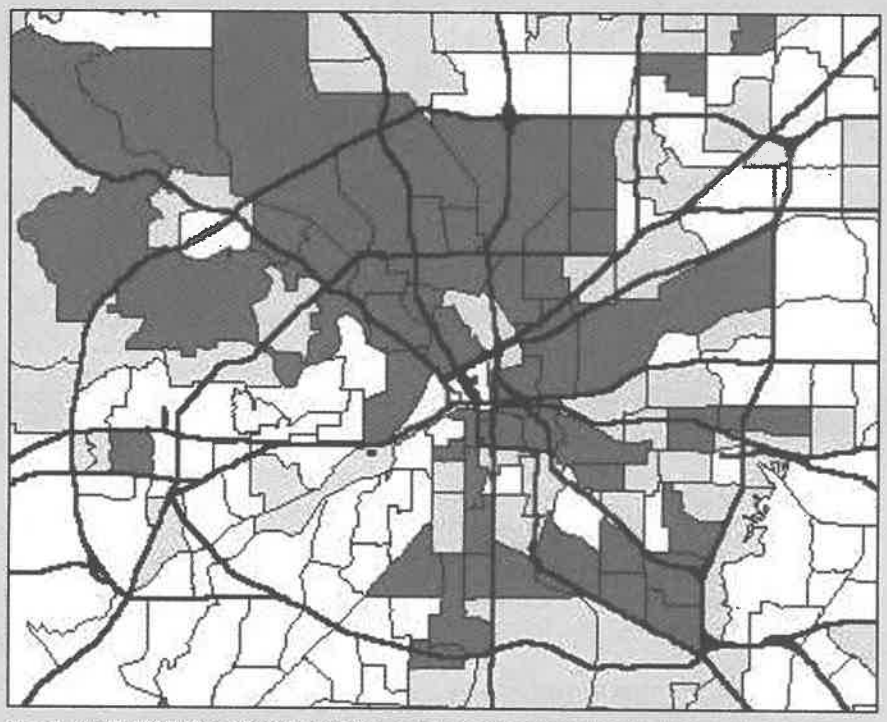
for a "Your Commute" page report dedicated to Census 2000 commuting statistics for the newspaper's area.

## Commuting patterns

Journalists have written plenty of stories showing how affluent commuters are willing to spend more time on the road so they can have big houses in the suburbs. Using GIS, journalists at the *Fort Worth Star-Telegram* discovered another pattern.

CAR specialist Jeff Claassen analyzed long-form commuting pattern data by tract with ArcView 3.2 to identify areas where a higher percentage of commuters left before 7 a.m. for work. The newspaper found that people living in low-income inner city neighborhoods had some of the longest and earliest commutes (Map 3).

The *Star-Telegram* used its findings



**Map 3:** *The Fort Worth Star-Telegram mapped commute times by census tract and found that residents of some inner-city neighborhoods posted long drive times (represented by darker shades).*

# Access

*continued from page 5*

that ID would be in four records. I needed to modify the tables so I could avoid a many-to-many join.

I needed a unique identifier to identify each inspection, so I created one in both tables. Then I ran update queries that combined the equipment ID with the inspection date.

Creating a violation code table and unique IDs may seem like old hat to CAR veterans, but for people like me – who have not built a lot of databases – these were good skills to learn. They apply to any sort of story that requires building a database or guarding against many-to-many joins.

## Finding the story

But what about the story? I found an agency that was underfunded and understaffed. The database analysis showed that the state had inspected only about 40 percent of the elevators in Missouri, even though the program was created nearly a decade ago. Additionally, of those that were inspected, about 13 percent went more than year before they were re-inspected – letting potentially harmful deficiencies persist.

**The database analysis showed that the state had inspected only 40 percent of the elevators in Missouri.**

States handle elevator inspections differently. Some make this a state function while others leave it up to municipalities. No matter who is doing it, you will probably find an agency that lacks the manpower to thoroughly inspect elevators annually. Getting a sense of the inspection budget and inspection staff is a starting point for this story.

Additionally, plan on reading lots of lawsuits – in both state and federal courts. It's a good bet that the enforcement agency lacks great information on accidents, but lawsuits against elevator companies are a treasure trove.

Contact Mike Sherry, a former data analyst in the IRE and NICAR database library and now a reporter for *CQ Weekly* by e-mail at MSherry@cq.com

## readme.txt

Mike Sherry's story about elevator safety can be ordered from the IRE Resource Center. Call 573-882-3364 or e-mail rescntr@nicar.org and ask for Story No. 19663.

Missouri's state auditor released a report in November finding that the Division of Fire Safety has not established procedures to identify unregistered elevators and to report registration violations for enforcement. See the report at *www.auditor.state.mo.us/press/2002-110.htm.*

A three-part Boston Globe series examined declining safety standards for elevators and escalators in Massachusetts and in the rest of the nation. The Globe reported that lax inspection standards and a lack of federal supervision helped result in an alarming frequency of crippling accidents and death. The investigation prompted one public official in charge of elevator safety to resign, and spurred the Massachusetts Public Safety Commissioner to order an overhaul of the inspection system. (Story No. 11001)

The IRE Resource Center provides stories and tipsheets to journalists. Search the story and tipsheet databases at *www.ire.org/resourcecenter.* Call 573-882-3364 or e-mail rescntr@nicar.org to order copies.

**SPOTLIGHT: CRIME & JUSTICE**

# Data shows judges set suspects free

By Joe Ellis and Brian Collister, *KMOL San Antonio*

Thousands of accused criminals in Bexar County, Texas, are off the hook because the county justice system couldn't get their cases to court fast enough.

We broke the story after four months of investigating, but fought more than a year for the data. No other news organization in San Antonio had investigated the Bexar County Court system with a database of its court records. And some county officials wanted to make sure we didn't either.

The battle began in December 2000. We asked for all Bexar County criminal court records to be provided electronically. We needed a complete database to see how judges run their courts.

At first, the county's Information Services Department quoted a $17,000 cost for our request, citing costly programming since there was no existing program for the job. After gathering advice from the NICAR-L listserv, researching the cost, and seeking help from the state's General Services Commission, we negotiated the cost down to about $2,000. But the county folks still dragged their feet.

Via the Texas Open Records Law, we monitored the correspondence between county officials regarding our request and bothered them constantly. Agreed deadlines for providing the data were not met. So finally, we contacted the company that set up the county's computer records system and hired one of its employees to write and run the program for $1,000. Within weeks, we had a "flat file" we imported in to an Access database.

## Querying the data

Our first objective was to see how tough – or lenient – the criminal court judges were on crimes like DWI, drugs, assaults or sex offenses. However, when we looked at the "COURT DISPOSITION" field, we kept noticing cases "dismissed for lack of speedy trial." That got our attention. So we isolated all the "DSMD-SPEEDY TRIAL" cases, and we counted them by the "DISPOSITION YEAR" field we created from the "DISPOSITION DATE" field in the data table.

We noticed a dramatic increase from the 607 cases dismissed for "lack of a speedy trial" between 1996 and 1998 to the 5,200 cases dismissed for the same reason from 1999 to 2001. We decided to focus on the latter three-year period and filtered out those 5,200 records and put them in a separate table.

Through further querying we calculated how long these cases were on file. Some were 15, 10, and five years old. But it seemed strange that many would be less than two years, or even one year old. Law experts agreed.

We then started looking through hundreds of individual case files, looking for answers, and people to help with our story.

In those files we found victims like Virginia, who had been beaten by her husband once and had a restraining order against him when he assaulted her the second time.

She didn't know, until we told her, that a criminal court judge had thrown out the case against her now ex-husband because he didn't get a speedy trial.

A "count" query of the "OFFENSE DESCRIPTION" field of our data table, showed more than 400 other assault cases that didn't get to court fast enough.

We also found 734 DWI-related cases where the driver was never prosecuted, including the driver who hit Myrna Ellison.

"I assumed something was done," Ellison said.

Nothing was done, though. Even though it was the driver's second DWI arrest, Judge M'Liss Christian dismissed the case.

Christian wouldn't agree to an interview, but when we caught up with her outside the courthouse, she claimed the misdemeanor cases she was tossing didn't have victims.

The judge obviously never met Myrna and Virginia.

Overall, we found Judge Tim Johnson, a judge with 14 years on the bench, tossed out the most cases with 1,528 speedy trial dismissals. Fourth-year judge Al Alonso had 1,057 dismissals, followed by seventh-year judge Karen Crouch with 920, and Christian with 802 dismissals.

None of these judges would agree to an on-camera interview, but Johnson, sent a statement on behalf of all county court judges: "The cases dismissed in Bexar County are dismissed under appropriate circumstances with the proper procedures and are justified both by statutory and case law."

But we found the courthouse crisis unique to Bexar County.

We checked with the Office of Court Administration in Austin and its records show: Harris (Houston), Dallas (Dallas), and Travis (Austin) County Courts *combined* report only 129 cases dismissed for "lack of a speedy trial" over the past three fiscal years. Bexar County courts report dismissing close to 4,200 cases on speedy trial grounds over the same period.

## Tallying the cost

Courthouse sources explained that the judges had created a "speedy trial docket" in 1999, an aggressive dismissal procedure. On the docket, judges choose cases they want dis-

*continued on page 10*

# Judges

*continued from page 9*

missed and assign them to attorneys who help get rid of them, sometimes hundreds at a time.

For taxpayers, we found another concern.

We obtained electronic payment records from the Bexar County Auditor's Office. This data details how much money each court-appointed attorney made in each court, case by case, for the past five years. With the "CASE NUMBER" field as the unique identifier, we matched the payment records with the speedy trial dismissals for the three-year period. We soon realized Bexar County Criminal Court judges were using thousands and thousands of dollars in public funds to get rid of cases never even heard in court.

By joining and comparing the court records data table and the attorney payment records table we discovered more than $100,000 in taxpayer money spent by judges appointing criminal cases on the "speedy trial docket." Christian was tops in this category, spending $28,076. Alonso spent $21,446, while Crouch spent $17,623 on appointed attorneys.

We found that basically the judges bypass the district attorney's office and handpick cases they want dismissed. They also handpick attorneys to help dismiss them. One lawyer, who asked not to be identified in our story, described the dismissal process.

Because of our investigation the county commissioners met to propose plans to stop judges from spending public funds dismissing cases and letting accused criminals go free. The commissioners created a task force to come up with some recommendations. The suggestions include: court coordinators and the County Warrants Division work together to make sure warrants are served; judges and the district attorney checking caseloads in criminal courts to make sure cases are heard in a timely manner; and fugitives being picked up when getting a marriage license, car registration, etc.

Also, local interest groups filed complaints with the Texas Commission on Judicial Conduct asking them to investigate and publicly reprimand the judges. The commission will not confirm or deny any investigation in progress.

Without computer-assisted reporting, the problem may not have been uncovered. We wouldn't suggest that other reporters will find the same problem in their community, but we would suggest obtaining similar data. We've produced several stories from this data and found it to be a valuable resource in the newsroom for enterprise reporting and daily use.

The key is doing one story at a time and staying focused. And when requesting and negotiating for data, never back down, matter how long the process may take.

Contact Joe Ellis by e-mail at jellis@kmol.com

Contact Brian Collister by e-mail at bcollister@kmol.com

## readme.txt

Search IRE's story and tipsheet databases at *www.ire.org/resourcecenter*. Call 573-882-3364 to order copies.

Tipsheet No. 1377 from Alden Loury of *The Chicago Reporter*, advises on covering jury selection.

Tips about covering criminal courts are included in Tipsheet No. 1435 from John Sullivan of *The* (Raleigh) *News & Observer*.

Looking into prosecutorial misconduct? Tipsheet No. 1677 from *Chicago Tribune's* Ken Armstrong and No. 1678 from Steve Weinberg, former executive director of IRE, can provide direction.

SPOTLIGHT: CRIME & JUSTICE

# Analysis finds juries lack color

By Mark Houser, *(Pittsburgh) Tribune-Review*

Blacks have as much right as whites to serve on a jury, according to a U.S. Supreme Court ruling that's more than a century old. You might not know it from looking in an Allegheny County jury room.

The recent *Tribune-Review* story "A jury of peers?" used surveying and GIS analysis to show how a system that's supposedly designed to randomly pick a fair cross-section of the community for jury duty consistently overlooks blacks.

Residents of the county's overwhelmingly white neighborhoods are twice as likely to get called for jury duty as residents of black neighborhoods, the story showed. While blacks make up 11 percent of the county's adult population, they make up only 4 percent of those serving on jury panels.

As a result, black defendants – who outnumber white defendants in our courts – often have their fates decided by all-white juries.

It's possible that courtrooms in your area have similar problems. Just take a look inside jury duty rooms and see whether the racial makeup of the jurors matches the demographics from your county's census numbers. Here's how you can use CAR to answer some key questions.

## Surveying for race

Pick a week or two and go to the jury room every morning before the cases start. Ask the court administrators how many jurors they have that day. Then count the number of black jurors. (When I was unsure, I briefly interviewed the person and asked his race.)

When I did this over two weeks I had counted 1,031 jurors, 4 percent of them black. That gave me a very small margin of error, plus or minus 1.2 percent at a 95 percent confidence interval.

Pittsburgh has almost no Hispanics, so I did not count them. If you want to look at Hispanics in the jury room, you have to ask everyone his or her ethnicity. Or you can substitute mapping analysis.

## Mapping juror lists

Another way to test for fairness is to map the addresses of people called for jury service and see whether residents of minority neighborhoods are as likely to be called for jury duty. You'll need to get the court's jury array – the list of people called to serve each day – along with their addresses and the dates served. Try to get a least a year; I got 18 months of potential jurors in a database.

Even in privacy-mad Pennsylvania (it has the fourth-worst public records laws in the United States, according to the Better Government Association, a nonpartisan watchdog group based in Chicago) court administrators must release the jury arrays.

Since the arrays document everyone called for jury duty, not just the people who showed up, the list is a better measure of how well the court administrators are summoning a "fair cross-section of the community," as the Supreme Court ruling requires.

Once I had the arrays, I needed to remove duplicate names. Our county court jury arrays are made public 30 days before jury duty, so there are always people on it who later are granted a postponement. Those people then show up months later on another array when they're rescheduled for duty. To find them, I sorted by last names and street names, then made a separate column with an "=IF" test to see if a cell repeated the one above it. (For details see sidebar on page 12).

With the duplicates removed from the data, I was ready to geocode those addresses and create a point map showing the residences of the potential jurors. ArcView 3.2, using Census TIGER street files, matched about 91 percent of the addresses off the bat. I got the match rate up to 99 percent by fixing mistakes in addresses, using the StreetMap geocoding extension and manually plotting particularly troublesome addresses.

Once I had created the shapefile of 45,000 juror points, I overlaid it onto a census block group layer with Census 2000 demographics. After that, I opened the attribute table for the potential jurors and block group layers and performed a spatial join to append the demographic data to the juror layer. Then I used the Summarize function to count the potential juror points inside each block group.

Using the Census 2000 race demographics for block groups, I created an Excel spreadsheet with two key numbers for each: the percent of the adult population that is black, and the number of jurors per 1,000 adults population, which I called the "jury service ratio."

The average jury service ratio in block groups that were at least 98 percent white was 53 per 1,000. The average ratio in block groups that were at least 50 percent black was 26 per 1,000. There was my lede: Black neighborhood residents were half as likely to be called for jury duty. I selected 98 percent white to make sure I excluded suburban neighborhoods where a significant number of blacks also live. I calculated similar numbers for the county's 130 municipalities, so I could refer to them in stories and graphics, because nobody knows what block group he lives in.

I calculated them for most of Pittsburgh's neighborhoods. Like many cities, Pittsburgh has several

*continued on page 12*

# Juries

*continued from page 11*

black and white enclaves. Instead of saying that in Pittsburgh 32 of 1,000 adults got called for jury duty, it was more telling to say that 20 did in (black) Larimer, and 57 did in (white) Regent Square.

If you have clearly defined Hispanic neighborhoods, you might be able to use GIS analysis to make up for difficulties in using jury room surveys to identify Latinos.

## Explaining patterns

If you find the similar results, the next step is to determine why. This was the most difficult part of the story, and I only was able to suggest some possibilities.

Most jury pools are picked by computer from voter rolls and driver's licenses lists. Many people say, "Aha! Blacks don't vote as often as whites." Maybe. But according to a Census Bureau study of the 2000 election, blacks register at almost the same rate as whites. "Motor Voter" laws prevent election authorities in Pennsylvania from regularly purging voter rolls, so the lists are useless in trying to see if black neighborhoods have fewer registered voters.

License records in Pennsylvania lack a racial designation for drivers. And the state Department of Transportation is hyper-secretive about those records so it is impossible to check if there are fewer licenses per adult population in black neighborhoods.

I found a very likely culprit for the jury pool discrepancies by looking elsewhere. The jury commission sends "qualifying questionnaires" to the people its computer picks to make sure they will be able to serve. A third of those never come back. The commission ignores the non-responses, even though at least half of them are returned because of a wrong address – in other words, the commission fails keep its address records current.

We know from the census that blacks in the county are more likely to rent. Renters move more often. Ergo, blacks are more likely to be missed by the jury commission because of a recent move.

The commission does an even worse job finding people in the neighborhood around the University of Pittsburgh. Residents there are mainly students, and they move every year. Only 10 adults per 1,000 in the Pitt neighborhood got called for jury duty.

Is that another story? Not likely. The Supreme Court has ruled that people can't challenge a jury room for being disproportionately old.

Contact Mark Houser by e-mail at mhouser@tribweb.com.

# Checking for duplicates in Excel

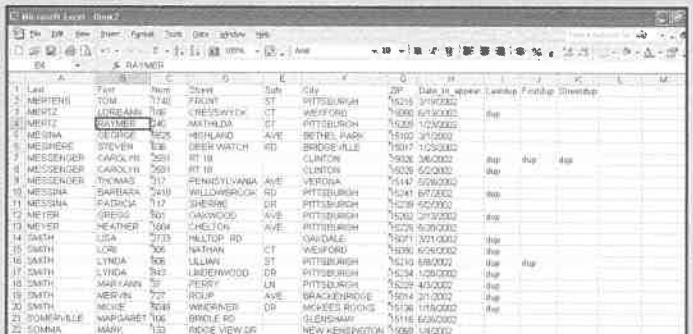
To find duplicates in the list of people called to jury duty, I sorted the list by last name, first name, and street name. Then I ran an "=IF" test on all three to see if the value in one cell was the same as in the cell below it.

This statement shows you if the last name is duplicated in the cell below for the spreadsheet pictured here:

```
=IF(A2=B2,"dup","")
```

If there is a duplication, Excel inserts the word "dup" the cell. I calculated a similar formula for first names and streets. Then by filtering I could find the triple "dups" and delete them, leaving only one record for the person. I checked double "dups" and if they appeared to be true duplicates, I deleted them.

In this example Carolyn Messenger is a triple "dup." She was initially summoned in March 2002 and then summoned again two months later. That means she postponed her first jury duty summons, because you can't serve twice in a year. I deleted that row.



Lynda Smith has two addresses. I can't know if she's a duplicate from this list, or if there are two Lynda Smiths, so I let both records stay.

Robert Sommer shows that it's important to use first names too. There are two Sommers from Galway Drive, but one's Robert and one's Roselyn. Both stayed.

CAR INVESTIGATIONS

# Elder care homes ripe for probes

By Stephanie Kang, *IRE and NICAR*

As America's elderly population explodes, journalists use database reporting to explore the state of assisted living centers and nursing homes around the country.

At just more than 20 years old, the assisted living industry is still relatively new. Arising out of dissatisfaction with nursing homes, assisted living centers cater to the elderly with financial means and little need of intensive nursing care. Yet because they are regulated by states, not by the federal government as most nursing homes are, monitoring assisted living care can become difficult. Assisted living centers and nursing homes have come under fire for lax regulation, understaffing, and patient abuse and neglect.

Months of anecdotal stories from state inspectors prompted *Milwaukee Journal Sentinel* reporter Mary Zahn to investigate the conditions of Wisconsin assisted living centers. A six-month investigation resulted in "Overwhelmed and Broken Down: Caring for the Elderly and Disabled," published in August 2001. Zahn used Wisconsin's open records law to obtain a database of state inspections and citations against nursing homes and assisted living centers. The series found that hundreds of elderly and disabled people in assisted living facilities had been injured and put at risk. In addition, 10,000 of the state's disabled or elderly residents had been forced to wait for months or even years to move into an assisted living facility. (Story No. 18778 in the IRE Resource Center)

Reporters for *The Oregonian* built and used existing databases while investigating the assisted living industry, focusing on Keren Wilson and the company she created, Assisted Living Concepts. For a series that ran in June 2001, they built several databases to track investigations of centers run by the company, entering information from investigative reports. They also tailored a complaints database that they got from the Oregon Senior and Disabled Services Division. An analysis of the data showed that Assisted Living Concepts had a track record that was worse than the state average. (Story No. 18972)

In "Who Will Care For Them? Ohio's Elder Crisis" the *Dayton Daily News* examined how Ohio cares for its elderly. The *Daily News* analyzed more than a half dozen state regulatory databases with FoxPro and SPSS. Some of the databases had never been publicly available in Ohio and the reporters had to negotiate with health department officials to obtain them.

The *Daily News* discovered a system "dangerously under-scrutinized," where the elderly often are forced to choose between costly assisted living care centers and state-funded nursing homes. The *Daily News* also found that state inspection personnel reductions and a 70 percent decrease in citations for serious health and safety violations threatened to minimize the state's influence on a sub-par facility (Story No. 15917)

In the questionnaire that accompanies their stories in the IRE Resource Center, the reporters advise: "Data is virtually limitless. Don't get overwhelmed. Use it to answer your questions, not for fishing expeditions."

## Finding data

For starters, federal law mandates that state regulators keep registries of aides working in every assisted living center and nursing home.

Nursing Home Compare data (*www.medicare.gov/nhcompare/home.asp*) allows you to check individual nursing homes in your area or download information to build your own database. The Online Survey Certification and Reporting file (OSCAR) at *www.longtermcareinfo.com/crg/about_oscar.html* is an administrative database of the Health Care Financing Administration. OSCAR provides survey and data. That means every institutional health-care provider in the United States that is certified to provide services under either Medicare or Medicaid (or both) is listed in OSCAR. Because different types of providers have to report different information during the survey and certification process, OSCAR results can vary.

Journalists also have access to complaint investigation forms. In Missouri, the Social Services Department Division on Aging regulates nursing homes and maintains a "Statement of Deficiencies and Plan of Correction" record. Most state health departments will maintain a complaints and enforcement database.

Look for datasets that lay out Medicaid and Medicare eligibility for nursing homes and assisted living centers, as well as a costs database, also from the Health Department.

These departments usually administer inspections and give licenses for adult care centers, adult residential care homes, intermediate care and skilled nursing centers. These datasets may include facility surveys, and provide operator information and resident demographics.

For details about the stories by the *Milwaukee Journal Sentinel* and *The Oregonian*, see the feature package about investigating assisted living centers in the November/December 2002 issue of *The IRE Journal*. Also included are tips and sources for covering assisted living and examples of other elder housing stories that used CAR. Order a copy through the IRE Resource Center by calling 573-882-3364 or e-mailing rescntr@nicar.org.

Contact Stephanie Kang
by e-mail at Stephanie@nicar.org

# Crime

*continued from page 1*

their data, which outlines crime totals and details homicides. The database also contains information about law enforcement agencies and officers, and demographics of people arrested for major crimes.

A useful companion is the National Crime Victimization Survey, a Department of Justice report based on a household survey designed to estimate the incidence of crimes such as theft robbery, assault and burglary. Journalists can use the data to estimate crimes that may not have been reported to law enforcement. Be aware that the crime reporting categories differ. Kurt Silver compares the uses of FBI data and the NCVS in "Understanding Crime Statistics," available from the IRE Resource Center.

Reporters have also used CAR to follow suspects through the courts. Journalists at *The Chicago Reporter* created a database of judicial appeals to help track a police unit's homicide convictions. A neighborhood coalition claimed the police regularly relied on faulty eyewitnesses. The database was used to write the 2000 series called "Seeking Justice."

Electronic court records give reporters the power to evaluate outcomes in the justice system. Although most court documents are open unless a judge says otherwise, some state and federal courts have resisted providing full electronic access to data. *The* (Wilmington, Del.) *News-Journal* asked for 10 years of felony and misdemeanor data in 1997 to investigate repeat offenders. The paper won limited data after dropping identifying details from the request and promising not to use the data to find individuals. The paper still had to sue to get information on repeat offender status. A recent court ruling upheld that request but denied access to records for arrests that did not end in convictions.

Some state records laws address electronic court records explicitly. See how your access law treats courts in Tapping Official Secrets, a Reporters Committee for the Freedom of the Press publication at *www.rcfp.org/tapping/index.cgi*.

## More data sources

For federal statistics, reports on criminal justice issues and other background material, the best starting point is the U.S. Department of Justice's Bureau of Justice Statistics Web page. (*www.ojp.usdoj.gov/bjs*) It links to other enforcement agency sites, in case you don't find what you need here.

The Sourcebook of Criminal Justice Statistics is published annually and available on the Justice Department page. To keep up with changes in sourcebook data before the yearly edition is published, see *www.albany.edu/sourcebook*. The site provides downloadable Lotus (.wk1) spreadsheets for many reports.

The Federal Justice Statistics Program, a part of the BJS, compiles comprehensive information about individuals entering and leaving the criminal justice system. It reconciles and strips identifying material from agency records, including the U.S. Marshals Service, Executive Office for U.S. Attorneys, the Administrative Office of U.S. Courts, U.S. Sentencing Commission and the Federal Bureau of Prisons. Pretrial services, appeals and probation data are included.

A public version of the files is provided in downloadable files. The standard analysis files, are grouped into cohorts – individuals in the system in a given fiscal year. Each year has a cohort file for several agencies' records. The Federal Justice Statistics Program provides a linking index file for combining agency data or linking between agencies, and SAS statistical program files designed to aid data analysis.

For faster, more focused searches, try the online query option or look for summary reports. Summary tables from the Compendium of Federal Justice Statistics can be downloaded as Acrobat files or spreadsheets.

When sorting through government reports, be careful how you choose or compare statistics. Be sure to read descriptions of definitions, sources and methods used to create or analyze the numbers. The Federal Justice Statistics Program generates many spin-off reports, but some government publications are based on periodic surveys or other statistical reporting.

A growing number of courts offer online docket listings with court schedules and litigant name searches. (See a list of state access systems at *www.ncsc.dni.us/NCSC/TIS/TIS99/pubacs99/PublicAccesslinks.htm*)

PACER (*http://pacer.psc.uscourts.gov*) acts as a central federal docket for U.S. district courts, bankruptcy court and appeals. Users pay 7 cents per page for online access to case, party and docket information. Dial-up services can be purchased for the few courts not available on the Web.

Contact Megan Christensen by e-mail at megan@nicar.org.

# Tech tip...

## Stitching points, data when GIS address matches fail

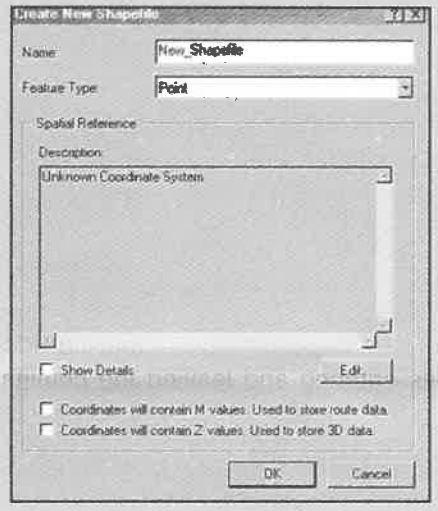
By James E. Wilkerson, *The (Allentown, Pa.) Morning Call*

One of our reporters recently wanted to get a visual picture of shootings around housing projects over a few months and came to me, because he know that I am using ArcView for mapping. What sounded like a simple project – plotting a list of shooting locations onto a street map of Allentown – turned out much more difficult than imagined, mostly because the location data was inconsistent.

Instead of having an exact street address, the shooting location field often contained an intersection or general block on which the shooting occurred. Geocoding didn't seem to be an option, at least not for the precise placement wanted. But because we were looking at only a couple dozen incidents, I decided to map them by hand.

First, I added a map shapefile of city streets that the city's GIS department provided. This street map – like many created by municipal agencies – is more often updated and corrected than the TIGER mapping files provided by the U.S. Census Bureau. The data did not include housing project boundaries, so reporter Keith Herbert and I drove to them and jotted down the streets around the perimeter of the projects. Later, in ArcView 8, we created a new layer for the polygon boundary.

Using computerized crime data provided by the city police department, I created a spreadsheet containing the incidents to map as points. Each record had basic information about the shooting and a unique identifier that I created, something needed to create a point map. I saved that to a dBASE file and added it as a table to the project.




Then I got stuck trying to figure out how to join the crimes to points on the city map. After posting a query on the NICAR and ESRI listservs, several people suggested that I create a new point file with no data and join it to the list of shootings.

### Here are the steps:

With streets and project layers in a view, I created a new, separate shapefile. In ArcView 8, you do that in the ArcCatalog application. Right click on the drive location where you want to store it and choose New/ Shapefile. Name the file and make sure that it is a point file.

After adding the point shapefile to a view in ArcMap, open its attribute table and you will see that it is empty, except for three fields.

Second, I opened the dBASE file of shooting incidents and arranged the table window and map window so I could see both. With the new shapefile opened for editing, I used the Create New Feature pointer tool. For each shooting location listed in the dBASE file, I created a point on the map that was as close as possible to the location described. Each time I clicked on the map to add a new point, ArcMap created a new record in the shapefile.

To avoid confusion, I worked with one point at a time. I copied the unique ID number of the shooting into the ID field of my new point shapefile, periodically saving the file.

In a couple of hours, I had a new map of about two dozen shooting locations, which I then joined to the file of shooting data on the common ID field for further analysis. The dBASE file and new shapefile could also be merged into one file, if you prefer, but I kept mine separate.

Unfortunately, the story never made it to print as we originally envisioned it, but the map gave us the ability to add new analysis to the subject, which was essential in deciding whether the story was worth pursuing.

Contact James E. Wilkerson by e-mail at james.wilkerson@mcall.com.

# Bars

*continued from page 1*

est, including the date of the accident, whether alcohol was involved and whether there were any injuries. State lawyers cited another law that requires the to withhold the identities of accident victims.

In appealing these rulings, *The Inquirer* succeeded in getting PennDOT to release data at the town level, which allowed an analysis of trends in accidents by town and county over a five-year period. That data, however, did not identify any roads. We still did not know the worst areas in the region for alcohol-related crashes.

For that, we had to rely on backdoor methods. Working through sources within the department, we were able to work around PennDOT's secrecy policy. We obtained data that located accidents on state roads in five-mile road segments. The table also included the total for a five-year period of alcohol-related accidents and fatalities. With that information, we were able to calculate the road segments with the highest per mile rate of alcohol-related accidents. That analysis gave us our first look at the worst roads for alcohol-related accidents. Once we had this data, we appealed to the state to release the latest year's data. By publication, we had road specific accident data for the most recent five-year period.

## Creating 'hot spots'

Using ArcView 8.1, we mapped the accident data and revealed for the first time roads with above-average accident rates. We focused the story on state Route 3, a 25-mile stretch that extends from Philadelphia into its western suburbs that included a segment with the highest accident rate at 22 per mile.

Working from paper records that we had accumulated from police and other sources, we found alcohol-related accidents that had occurred on Route 3 in which people had been killed. Through our reporting, we also located other accidents that had occurred on the very roads that police knew had high alcohol-related crash rates and where bars were concentrated. In some cases, they were located in towns where police made relatively few arrests for drunken driving. The tragedy of drunken driving was told through the stories of the victims who had been killed and injured.


Next, we obtained liquor license data from the state's Liquor Control Board. In this case, using ArcView's Spatial Analyst extension worked much better than producing an ArcView map. Mapping the 4,000 establishments by plotting each location produced a map of little value. Likewise, a shade map of license counts by town was also misleading since it would have incorrectly given the impression that licenses were evenly distributed across a town when in reality they were generally clumped in older, established neighborhoods.

We wanted to generate a map showing the liquor license "hot spots." The best tool turned out to be Spatial Analyst, using the actual locations of the point data to generate the "hot spots" in just the area of the map where the liquor licenses were concentrated, giving a much more accurate rendition.

The result was a map that displayed concentrations of liquor establishments from two to six or more per square mile. The more liquor establishments, the darker the ring.

When layered with the accident data, we were able to identify the same link between crashes and liquor establishments that researchers had come up with in California. While crashes have many causes – road design, weather, speed – those involving alcohol happened most where bars and restaurants are concentrated.

This link was especially important in Pennsylvania. A new law makes it possible to sell liquor licenses within counties. Previously, a population-based quota had restricted licenses to the towns in which they were already located, resulting in a concentration of drinking establishments in older, more densely populated neighborhoods.

Our analysis accounted for traffic volume. We found that some of the busiest roads did not have the highest accident rates. Parts of Interstate 95 had the most traffic in the region – but not nearly the highest rates of alcohol-related crashes. We also learned that using traffic count data was flawed because traffic counts are usually taken during the daytime, while alcohol-related crashes were more likely to occur between 8 p.m. and 4 a.m.

Because we were new to density analysis, we also consulted several experts including Paul J. Gruenewald of the Pacific Institute for Research and Evaluation (*www.pire.org*) who had done research in California studies.

For those of you who want to know more about how Spatial Analyst works, here's a brief summary from ESRI's "Using ArcGIS Spatial Analyst:" Spatial Analyst calculates density by spreading point values over a surface. A density value is calculated for each cell in the output raster.

Points that fall within the search area are summed and then divided by the search area size to get each cell's density value. We elected to calculate density using the kernel calculations because points at the center of the raster cell's search area are weighted more heavily than those near the edge. The result is a smoother distribution of values.

## Mapping put online

For online, Matthew Ericson, senior editor for Web technology, designed an interactive map using ESRI's ArcIMS Internet map server, which allowed readers to zoom in on the most dangerous roads in the region and see the total number of alcohol-related crashes, the number of alcohol-related fatalities and the per mile crash rate.

The Web site (*www.philly.com/mld/inquirer/news/special_packages/dui*) included a number of interactive tools that allowed our readers to go beyond what we printed in the newspaper and see how DUI affects their lives. A blood alcohol calculator lets readers enter their weight and gender to see how many drinks it takes to reach various levels of impairment.

In addition to the accident/liquor license analysis, the project included an analysis of police arrest data by town. Calculating a per 10,000 arrest rate and dividing the results into quintiles allowed us to compare police department performance. Readers could also look up their town in a searchable database to find detailed information on the number of alcohol-related crashes by town and by county, and arrests by town, by county, by arrest rate and per police officer. The project included the analysis of other data from courts in two states and Philadelphia and campaign contributions.

Contact Rose Ciotta by e-mail at ricotta@phillynews.com.

Contact Karl Stark by e-mail at kstark@phillynews.com.

## readme.txt

### Stories

Many recent stories on crime and justice that used computer-assisted reporting can be found online and ordered from the IRE Resource Center by calling 573-882-3364 or e-mailing rescntr@nicar.org.

**Recent stories include:**

• "Justice Undone," by *The* (Baltimore) *Sun*. The paper investigated more than 1,400 murders between 1997 and 2001 and found that two-thirds of the time, either no one was arrested, or the people who were arrested went free or were sent to jail for short periods of time on lesser charges.

• "Prosecutors' conviction rate falls," by *The Wichita Eagle*. *The Eagle* analyzed more than 14,000 court cases in Sedgwick County District Court and found a decline in conviction rates. Meanwhile, not guilty verdicts in felony trials more than tripled from 1999 to 2001. Sources attributed the changes to charging some people with crimes more serious than the evidence supported, and to inexperience among prosecutors.

• "Who's Watching the Deputies? Patronage at work in the sheriff's department," by the *St. Louis Post-Dispatch*. Reporters found that some deputies in the St. Louis sheriff's department who didn't show up for work got paid anyway. The paper also found five men working as deputies, even though they had been rejected when they applied for security guard licenses. (Available as Story No.18987 in the IRE Resource Center.)

# GIS

*continued from page 1*

Other journalists are running into the same problems Washburn has experienced. The privatization of government functions and amendments to public records laws has hampered access to GIS data. He said that because SanGIS was initially funded by taxpayer money and backed by the county it should be more reasonable with its prices.

"I totally think they are breaking the law by doing this because it involves taxpayer money," said Washburn, an investigative reporter. SanGIS contends that because it is now self-sustaining, its prices are reasonable and reflect the time and effort it takes to digitize the county layers.

Journalists are finding that, along with charging high costs for GIS data, the departments have basically cornered the market. While some private businesses such as real estate or law firms may be able to afford the data, most news organizations lack the money.

## Laws increase costs

While the GIS costs in San Diego County are one extreme, journalists have encountered a spectrum of difficulties and costs when attempting to obtain GIS data from government agencies. The availability and cost of GIS data depends on the location, public records laws and who's reviewing the request for data.

Part of the reason Dan Keating, database editor at *The Washington Post*, first began using GIS data was because it was so easily accessible from county GIS departments in Florida, where he had worked before. "Both Miami-Dade and Broward counties gave me CDs with dozens and dozens of local GIS layers for practically nothing, no more than $50," Keating said.

That changed once he left *The Miami Herald* and began working in

*continued on page 18*

# GIS

*continued from page 17*

Washington, D.C. While costs for data from the District and Virginia counties were reasonable, he said, obtaining GIS data from agencies in Maryland proved to be a nightmare. "The state has an exemption to its public records law for GIS, saying that it can be sold for market cost. So layers from a county can cost thousands of dollars," Keating said.

Maurice Tamman, staff writer at *The Atlanta Journal-Constitution*, has had troubles dealing with GIS laws. While working on a story requiring parcel maps, Tamman requested GIS data from Forsyth County and was quoted a price of more than $10,000 for the maps he was seeking.

A recently enacted Georgia law gives agencies the authority to charge for GIS data based upon the actual recovery and development time involved with the GIS. Tamman said the county follows this new law to the letter.

Meanwhile, he was able to receive the same type of GIS data from a neighboring county absolutely free. "It is interesting to me, because traditionally we view any data produced by the government as being fundamentally free," Tamman said. "But GIS has been placed on a pedestal because it has some intrinsic market value."

## Fighting some battles

Some reporters have gone head-to-head with government agencies challenging fees based on outrageous standards. While many of these journalists bear the battle scars, some have the GIS data to show for it as well.

Paula Sullivan, city hall reporter and computer-assisted reporting coordinator at *The* (Tacoma, Wash.) *News Tribune*, went around in circles trying to negotiate with Pierce County for some GIS data. The agency's going rate was $60 per shapefile, when all they had to do to produce the files was copy ones already created.

"State law does not allow them to do this, so we complained to the county executive and – instead of trying to negotiate a reasonable price – they gave it to us free just that once," Sullivan said.

Since that initial dispute, she said they have received other files from the county. "However, the department continues to charge the $60 rate to others and makes them sign all sorts of forms, including a data-sharing agreement that we have always refused to sign."

> **"...GIS has been placed on a pedestal because it has some intrinsic market value."**

While some journalists have had to fight government agencies to lower costs for GIS data, others have found GIS departments to be friendly, helpful, and even willing to provide free data.

Although Sullivan had to convince one Washington county to give her data, she said dealing with the city of Tacoma is a different story. This city posts many of its most-accessed shapefiles on its Web site, free to any user with an Internet connection.

With the recent explosion in the use of GIS, more state, local and federal agencies have been putting GIS data on their Web sites for free download.

## More experiences

John McCarthy, a reporter at *Florida Today*, had a good experience with one state agency.

McCarthy requested GIS layers of Florida's coastline to analyze how vulnerable those areas would be to hurricanes. Florida's Emergency Operations Center provided the data quickly. "Their GIS specialist spent quite a bit of time on the phone with me working out bugs and was very cheerful about it," he said.

Other journalists have had to give in just a little to receive the shapefiles and layers that they needed. Maricopa County, Ariz., was willing to give Enric Volante, CAR coordinator and database reporter at the *Arizona Daily Star*, GIS data for free, but only if he signed a liability waiver.

"First time in my many years of journalism that I ever signed a liability waiver in order to get a public record," Volante said. "I also filed a formal protest, but they didn't seem to understand what all the fuss was about."

John Hill, systems editor at *The* (Vancouver, Wash.) *Columbian*, has received just about all of the GIS data a newsroom could need from the Clark County GIS department. The newspaper paid $1,200 per year for two CDs that include school district boundaries and soil contours. "While it does seem like a lot of money, when we have a GIS project going it's all money well spent," Hill said.

While many journalists have been able to get their hands on the data they need, there are still some like Tamman and Washburn who have been completely shut off from the data because of the high prices GIS departments demand.

"It is certainly putting a roadblock to somebody trying to do their job," Tamman said. "I couldn't do some of those stories I wanted to do. Whether it'll stop me from trying, though, that's another story."

Contact Jaimi Dowdell by e-mail at jaimi@nicar.org

# IRE and NICAR Services

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

## Programs and Services

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

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

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

**On-the-Road Training**: As a top promoter of journalism education, IRE offers loads of training opportunities throughout the year. Possibilities range from national conferences and regional workshops to weeklong boot camps and on-site newsroom training. Costs are on a sliding scale and fellowships are available to many of the events.
**Contact**: Ron Nixon,
ron@nicar.org, 573-882-2042

## Publications

***The IRE Journal***: Published six times a year. Contains journalist profiles, how-to stories, reviews, investigative ideas and backgrounding tips. *The Journal* also provides members with the latest news on upcoming events and training opportunities from IRE and NICAR.
**Contact**: Len Bruzzese,
len@ire.org, 573-882-2042

***Uplink***: Monthly newsletter by IRE and NICAR on computer-assisted reporting. Often, *Uplink* stories are written after reporters have had particular success using data to investigate stories. The columns include valuable information on advanced database techniques as well as success stories written by newly trained CAR reporters.
**Contact**: David Herzog,
dherzog@nicar.org, 573-882-2127

**Reporter.org**: A collection of Web-based resources for journalists, journalism educators and others. Discounted Web hosting and services such as mailing list management and site development are provided to other nonprofit journalism organizations.
**Contact**: Ted Peterson,
ted@nicar.org, 573-884-7321

## For information on:

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

**Membership and subscriptions**:
John Green,
jgreen@ire.org, 573-882-2772

**Conferences and Boot Camps**:
Ev Ruch-Graham,
ev@ire.org, 573-882-8969

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

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

# Computer - Assisted Reporting Boot Camps

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

- Jan. 12-17 – Columbia, Mo.
- March 23-28 – Columbia, Mo.
- May 18-23 – Columbia, Mo.
- Aug. 3-8 – Columbia, Mo.

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

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

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

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

– Anonymous (from seminar evaluation)

*"Well worth the money!"*

– Anonymous (from seminar evaluation)

**More information is available at www.ire.org/training**

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

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

Ira Chinoy
Philip Merrill College of Journalism
1117 Journalism BLDG
University of Maryland
College Park MD 20742

**Uplink Info**

*A newsletter of the National Institute for Computer-Assisted Reporting*

NICAR is a joint program of Investigative Reporters and Editors Inc. and the Missouri School of Journalism.

NICAR services include supplying journalists with government databases, training programs, tipsheets and data analysis.

**Editorial**
573-884-7711

**Subscriptions**
573-882-2772

*Postmaster: Please send address changes to IRE-NICAR.*

**Uplink Address:**
IRE-NICAR, 138 Neff Annex
Missouri School of Journalism
Columbia, MO 65211

**Subscriptions**
IRE members $40, nonmembers $60

**Subscription Administrator**
John Green
jgreen@ire.org

**Advertising Coordinator**
Pia Christensen
pia@ire.org

**Director of Publications**
Len Bruzzese
len@ire.org

**Editor**
Brant Houston
brant@ire.org

**Managing Editor**
David Herzog
dherzog@nicar.org

**Asst. Managing Editor**
Jeff Porter
jeff@nicar.org

**Art Director**
Lisa Triefenbach

**Copy Editor**
Pia Christensen

**Contributing Editors**
Megan Christensen
Jaimi Dowdell
Holly Hacker
Stephanie Kang