Uplink

July • August 2003
Volume 15, Number 4

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

TAXES

# Finding the biggest deadbeats

By Mark Greenblatt
*WBBH-Fort Myers/Naples, Fla.*

Our project began with a search for the top 10 tax delinquents in Charlotte County. With government budgets facing some of their tightest years in state history, my television station wanted to know who wasn't paying their fair share of property taxes in southwest Florida. By the end of the reporting, we discovered a much more important story that could potentially help collect millions of dollars in lost revenue for Charlotte County.

I began by contacting the Charlotte County tax collector and requesting a complete database of all tax certificates issued to people who hadn't paid their real property taxes in the last seven years. We wanted to know who the delinquents were and how much they owed in taxes and penalties.

Within a week of the initial request, I received a comprehensive set of "deadbeat data" from the tax collector on CD-ROM.

*continued on page 18*

SPOTLIGHT: JOBS, PAY & PERKS

# Simple sorts reveal cops' overtime jackpot

By Mark Pazniokas, *The Hartford Courant*

Hartford police complained for years about being underpaid. Starting pay is just $34,554 – lower than in many nearby suburbs. Patrol officers' base pay tops out after five years at $56,095.

So it was startling news when *The Hartford Courant* reported that the median income for the city's police officers actually was $75,000 last year, with 20 percent of the force's 377 of-

*continued on page 17*

**SPOTLIGHT:**

FOR MORE ON JOBS, PAY & PERKS SEE:



FIRST VENTURE

# Learning lessons while tracing pol's spending

By Todd McAdam, *Press & Sun-Bulletin (Binghamton, NY)*

As I walked out of a prominent state senator's office following an interview, he and his spokeswoman handed me a plastic bag. In it were two bars of high-quality chocolate imprinted with his campaign logo, a coffee mug with the same design, a bottle of a spice rub suitable for beef, chicken *or* fish, and a bottle of steak sauce regularly sold at the politician's premier fundraiser.

"You know," I told the spokeswoman, a longtime source, "I can take this only if I do a story on it someday."

She laughed.

Six months later, I sat at my desk and stared at the mug. It was gathering dust on a corner of my desk. I had just returned home from an IRE and NICAR Boot Camp and was eager to prove that I don't believe in junkets – and wouldn't go to Columbia, Mo., for one even if I did.

*continued on page 20*

## Bits & Bytes

### Historical spending data

The IRE and NICAR Database Library has compiled archived data to provide journalists with the most thorough tools for covering federal spending. This historical data allows journalists the unique opportunity to go back in time and track federal spending in their communities across decades. Journalists can use the data to show how different presidential administrations, economic conditions and pivotal events affected spending.

The IRE and NICAR Database Library now has Federal Contracts data covering fiscal years 1979-2002 . The database lists almost all federal contracts worth more than $25,000 and includes the amount, name and location of the company, the agency, the type of work and where the work was performed.

The Database Library has also added Federal Assistance Award Data System (FAADS) covering fiscal years 1983-2002. Journalists can use the FAADS data to determine what federal money has been obligated to individuals, state and local governments, institutions of higher education, profit organizations, nonprofit agencies and others.

For more details on these datasets, including prices and ordering information, go to *www.ire.org/datalibrary/databases* or contact the IRE and NICAR Database Library at 573-884-7711.

*continued on page 4*

# IRE's chief trainer moves on

By David Herzog, *University of Missouri and NICAR*

Ron Nixon, who's trained thousands of journalists across the globe in computer-assisted reporting for IRE and NICAR, is headed back to the newsroom.

After two and a half years on the road, Ron is set to go home to Minneapolis, where he will become the computer-assisted reporting editor on the *Star Tribune* projects team in early September.

As the training director for IRE and NICAR, Ron logged countless miles to help print and broadcast journalists do their jobs better and earned their praise. He's trained in newsrooms, national journalism association events and at events conducted by IRE and NICAR.

Ron joined IRE and NICAR as the director of the Campaign Finance Information Center (*www.campaignfinance.org*) and later moved into the top training post. Before joining IRE and NICAR, Ron was an investigative reporter for *The Roanoke* (Va.) *Times* with years of experience working with local, state and federal campaign finance data.

In his spare time at home, Ron kept his skills sharp by performing database and geographic information system analysis for clients of the IRE and NICAR database library.

Ron's last official day with IRE and NICAR is Aug. 29. Drop him a note at ron@ire.org and join us in wishing him well as he returns to newspaper journalism.

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

# Hands-on CAR training upcoming

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

Here are some of the highlights:

Journalists can get an overview of computer-assisted reporting at Better Watchdog Workshops offered by IRE and the Society for Professional Journalists. Upcoming workshops include Sept. 11 in Tampa, Fla., Oct. 4-5 in Eugene, Ore., and Oct. 25 in State College, Pa.

There are four CAR Boot Camps in Columbia, Mo., next year for journalists who want hands-on training in how to use spreadsheets and database managers to analyze data for high-impact stories. The sessions are Jan. 4-9, March 21-26, May 16-21 and Aug. 1-6.

Also, mark your calendar for the annual CAR conference, March 12-14 in Cincinnati. The conference promises to be great for tips, story ideas and hands-on training.

IRE AWARDS

# CAR gives contest winners boost

By Rachel Schaff, *University of Missouri*

The May/June issue of *Uplink* highlighted the television, special categories and other media winners and finalists of IRE's annual investigative reporting contest that used computer-assisted reporting. Here we feature the newspaper, FOI and Tom Renner award winners and finalists that used CAR.

## Local circulation weeklies

**Finalist:**

*Atlanta Business Chronicle*: "Northern Arc: Road to Riches" by Sarah Rubenstein and Walter Woods. Rubenstein and Woods exposed how the planned construction of a new highway north of Atlanta, called the Northern Arc, would benefit government officials who owned property along the proposed road. The reporters based their investigation on land parcel maps and corresponding Microsoft Excel spreadsheets for the counties through which the Northern Arc would pass. Rubenstein and Woods were named journalists of the year by the Atlanta Press Club and the story won Best Scoop from *Atlanta Magazine*. (Story No. 19576 in the IRE Resource Center)

## Small newspapers

(Circulation less than 100,000)

**Finalists:**

*Charleston* (W.Va.) *Gazette*: "The Long Haul/Broken Promises" by Eric Eyre and Scott Finn. This investigation examined the effects of school consolidation in West Virginia, including the increased amount of time students had to spend traveling to and from school. Eyre and Finn requested bus logs under FOIA for 35 counties. With the information, they built a spreadsheet of 1,569 bus runs, forming the foundation for the story. The series won the Education Writers Association's Fred M. Hechinger Grand Prize for Distinguished Education Reporting. See "Measuring long bus rides for rural children" in the September/October 2002 *Uplink* for a full account of the data analysis. (Story No. 19601)

*Tri-City Herald* (Pasco, Wash.): "Bitter Harvest" by Mike Lee. This story investigated a series of deaths, environmental damage and accidents that were all traced back to one Columbia Basin farm, the largest organic farm in Washington state. Lee showed how the farm's owners ignored environmental regulations, and he exposed the government's failure to take any decisive action. Lee used spreadsheets to track story threads and piece together information from a dozen agencies. The story – which relied on more than 20 public document requests from state and federal agencies — won in the investigative reporting category in Sigma Delta Chi's competition for newspapers with a circulation under 100,000 and first place for agriculture/business reporting in the Inland Northwest SPJ competition. (Story No. 19608)

## Medium newspapers

(100,000 – 250,000 circulation)

**Certificate:**

*The Fresno Bee*: "Last Gasp" by Russell Clemings, Barbara Anderson and Mark Grossi. This special report traced how California's San Joaquin Valley air basin became one of America's most polluted. The reporters used FOI requests to acquire information from many sources, including the U.S. Environmental Protection Agency. They then used Access, Excel and FoxPro to compare current levels of pollution with historical estimates and eventually to understand the growing toxicity of the basin. "Last Gasp" also won first place in the Associated Press News Executives Council's contest for public service reporting; first place in the Best of the West explanatory reporting category; the Meade Clean Air Prize; the Coalition for Clean Air's Corporate Clean Air Leadership Award; and the EPA's Environmental Achievement Award. See the September/October 2003 issue of *The IRE Journal* for details of their investigation. (Story No. 19677)

**Finalists:**

*The Birmingham News*: "The Black Belt: Alabama's Third World" by John Archibald, Jeff Hansen, Carla Crowder, Thomas Spencer and Marie Jones. This piece focused on a small region of Alabama called the "Black Belt." People in this region, mostly poverty-stricken African-Americans, live in conditions much worse than those in the rest of the state. The reporters relied on databases of all land records for 10 Alabama counties as well as census records, mortality and birth records, data from the Alabama Department of Public Health, property tax records, and other state databases. The authors built many of their own databases using Excel and Access. This series also won awards from the Alabama Associated Press Managing Editors, Alabama Press Association and the Medical Association of Alabama. (Story No. 19724)

*Dayton Daily News*: "Deadly Consequences, Ohio's Broken Mental Retardation System" by Tom Beyerlein and Steve Bennish. The *Dayton Daily News* identified 30 cases of neglect within Ohio's mental retardation system by matching death cases against addresses of group homes for the mentally retarded. The reporters used Excel and Access to analyze 400,000 death records and develop their own database of deaths in group homes. (Story No. 19107)

*The Virginian-Pilot*: "Operating Behind Closed Doors" by Liz Szabo. This story illustrates the failures of Virginia's system for protecting patients from dangerous doctors by highlighting the last decade in the

*continued on page 22*

## Bits & Bytes

continued from page 2

### Data analysis

The IRE and NICAR Database Library recently conducted an environmental data analysis for a story that aired on WMAQ-Chicago in May. The analysis identified areas with a large number of polluters under state remediation programs. Using ArcView, the Database Library mapped polluters by ZIP code, by latitude and longitude and, using the Spatial Analyst extension, found the neighborhoods with the highest concentrations of polluters and pollutants. To read the story, go to *www.nbc5.com/news/2203768/detail.html*. For more information on how IRE and NICAR can help your news organization conduct data analysis, contact the Database Library at 573-884-7711.

### Training

Although no IRE and NICAR Boot Camps remain this year, journalists will have many opportunities in 2004 to take advantage of the specialty training that these high-impact CAR seminars offer. Dates for the 2004 Boot Camps in Columbia, Mo., are Jan. 4-9, March 21-26, May 16-21 and Aug. 1-6.

For more details about the Boot Camps, including fellowship and registration information and a typical daily training schedule, go to *www.ire.org/training/bootcamps.html*. For information on other training opportunities available from IRE and NICAR, go to *www.ire.org/training*.

**SPOTLIGHT: JOBS, PAY & PERKS**

# Tallying the bill for unused sick time

By Robert Gebeloff, *The Star-Ledger*

When we use computer-assisted reporting for our government coverage, salary data is often at the top of our list.

We obtain government payrolls and rank them. We compare median salaries of teachers among our school districts. When a government worker gets suspended for official misconduct, we don't report just that. We write that the worker was suspended from a $90,000-a-year job.

But salary data is just the tip of the iceberg in reporting on government workers and their benefits. There's police overtime, government-issue cell phones and credit cards.

And in New Jersey there's the matter of sick and vacation time. Seemingly innocuous, except, as we found, that local government agencies are now on the hook for nearly $1.5 billion to reimburse employees who fail to use their allotted time.

Ever since the Great Depression, civil servants in New Jersey have essentially been guaranteed 15 paid sick days a year. Over time, government agencies began allowing workers to carry unused days over into sick day banks, a practice later codified in union contracts.

It's a progressive benefit intended to protect workers who suffer prolonged illness. But in the modern era, it has become a bonanza for retirees. As public employee unions gained strength in the 1970s, they managed to negotiate mandatory payouts for unused sick and vacation time.

In almost all cases the payouts are made to retirees at their final pay rate, not the salary at the time the days were banked.

The *Star-Ledger* was certainly not the first paper to notice how common it had become for a police chief or schools superintendent to collect a six-figure payout – on top of a generous pension – just for retiring.

In fact, when I was with a regional paper in New Jersey a few years ago, we did a pretty good CAR story that looked at the potential cost of this system.

But nobody had ever been able to report precisely what the statewide impact was, and thus we had a mission.

When it came to gathering data, we had one thing going for us. The state had recently enacted new regulations that required municipalities to annually calculate the value of "compensated absences."

We found that school districts had been making the calculations for years and listed the amount as a long-term liability on their annual audits.

That was the good news. The bad news was that none of this information was available electronically. We wanted a number for 1,177 different agencies throughout the state and it was all on paper. We discovered that all of the paper ends up in one place: Trenton.

So with a research assistant and laptop in tow, I made a series of visits to the state capital. It took us about six days to pull the 1,177 documents out of filing cabinets one by one, find the sick time number on each and enter them into a spreadsheet.

Once we had the numbers it was easy to add it all up and come up with the $1.5 billion total.

More challenging was coming up with

a method for compiling town-by-town statistics. In addition to the state's 566 municipalities, we had to account for school districts and county agencies that collected taxes across town borders.

Since we wanted to measure the impact on the basis of property taxes, we used property tax data to help us divvy up the burden.

For example, employees of the West Essex Regional district had $400,000 in the sick-time bank. Roseland is a town served by the district, and about one quarter of the district's tax base is located in Roseland. Therefore, we assigned one quarter of the obligation to pay West Essex retirees to Roseland residents.

We did the same with the obligation to county employees so at the end of the process, we had one figure for each town that incorporated what was owed to municipal workers, school employees and county employees.

After we had a single number for each town, we wanted to add context.

Obviously, the biggest cities would have the most employees – and likely the highest total obligation.

Yet these towns would also be able to spread their obligation among a larger base of taxpayers. So we needed to come up with a way of quantifying the number. We came up with two.

First, we did a simple per-capita calculation. Then we plugged the number into the town's property tax formula and expressed it in terms of the average homeowner.

We used both in the story. Per capita, we decided, was an easy-to-understand ratio that generally put the total obligation into context. The average homeowner cost, meanwhile, was a more precise measure that told readers "This is exactly what it's going to cost you in coming years."

We found that in a tiny urban enclave called Harrison, the average homeowner was on the hook for nearly $3,100. Overall, the debt was $251 per person in the state's 30 cities, but only $152 per person in the suburbs.

We also found that individual golden parachutes were being proffered all over the state. In one of the audits we pulled, for a tiny rural community called Stillwater, there was a footnote from the auditor explaining that the town's obligation was unclear, pending resolution of a conflict with a soon-to-retire police chief.

Upon closer examination, we found a community where the police force could be counted on one hand, the number of annual crimes on two, and the police chief was walking away with a severance package of $117,587 for sick and vacation time he didn't use.

After the story ran, a reader tipped us off to another case that lead to a follow-up: The police chief in Jersey City had recently been granted a compensation package that provided him with an instant bank of 150 sick days – days he could cash in upon retirement.

That wasn't necessarily unusual, except that the chief had already retired once from the force, as a lieutenant, and cashed in a six-figure payout. Now he was in line to get a second helping of taxpayer generosity.

There were, of course, arguments included in these stories for why such payouts are justified. Unions told us that because these benefits are often negotiated for their rank and file, it means the unions had to give up something else to get such a deal.

Still, the point of our stories was not to pass judgment on the system. Rather, our goal was to tell readers the real cost of such a system – a figure that had been unknown until we completed our project.

Read the story, "Cashing Out," online at *www.nj.com/specialprojects*.

Contact Rob Gebeloff by e-mail at rgebeloff@starledger.com.

# MAPPING IT OUT

*The latest uses of mapping in news reporting.*

## GIS shows close call for Dallas

By John Kelly, *Florida Today*

"How many are dead on the ground?"

I must have asked that question a dozen or so times the morning of Feb. 1, as my colleagues and I scrambled to find out what happened to the shuttle Columbia in the far reaches of the Texas sky. Watching the televised images of giant steel balls, mangled fragments of wing and other heavy metal objects strewn across Texas, it seemed obvious the death toll was going to be higher than the seven astronauts. The question was, "How many?"

The still-dumbfounding answer is zero. NASA got lucky.

NASA seemed to be ignoring how many people on the ground had been in harm's way. It was a fundamental public safety issue. We weren't about to ignore it. We could have just thrown together color from Texas with dueling expert quotes for a "what-a-miracle" story. Instead, we wanted to show in a definitive way the very real risk of flying the space shuttle over populated areas.

We set out to apply the same methods that NASA, the U.S. Air Force and big aerospace contractors use when they try to assess danger to people on the ground during space launches, or when one of their unmanned satellites or rockets eventually plummets back to Earth.

That turned out to be the right approach for two reasons.

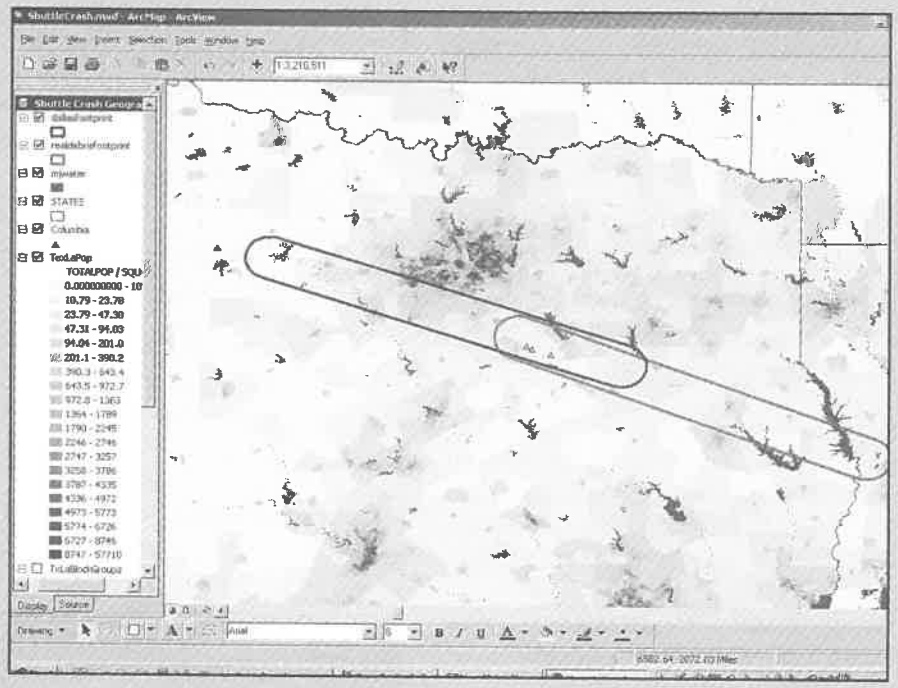
First, my colleague Todd Halvorson and I piled up a mountain of public records showing how carefully the government and aerospace companies make sure that debris from re-entering spacecraft and rockets does not strike land, let alone come down on populated areas. By contrast, Halvorson found NASA had never studied the danger to people on the ground during the very risky re-entry of its space shuttles.

Second, we learned the industry's accepted method for minimizing death and destruction on the ground was to predict the footprint, or area, where the debris would hit, and to do everything possible to control the re-entering space junk so the danger zone would contain few, if any, residents.

I knew we could use ESRI ArcView mapping software to determine the number of people who were in the path of shuttle parts. Using ArcView we layered the debris footprint over U.S. Census population data and determined that about 216,000 people lived in the largely rural area where most of the debris hit. We also used the shuttle's flight path to show the more grim consequences had the shuttle broken up about a minute earlier.

In that case, the storm of metal would have hit Dallas' suburbs. About three times as many people and houses would have been in the path of falling wreckage.

Computer-assisted reporting nerd that I am, I overcomplicated things from the start.



GIS and computer mapping tools were being used in the recovery effort. It was clear search teams were recording the precise coordinates of each piece of shuttle found. NASA was making computer maps.

I went to Stephen F. Austin University, which was helping NASA with the mapping. They made available online ArcView shapefiles of the region of Texas where debris was found. I wanted the shapefiles that were the basis of NASA's nifty handouts, showing every piece of debris as a dot on the Texas map. I had the big idea that I was going to calculate how many people lived a set distance from any one of those dots.

The university asked NASA if they could give me the data. NASA said no. We could FOIA the database, but we did not have time to wait.

I went back to what the university released online. I found a shapefile of the 20-kilometer buffer zone calculated from the centerline of the debris field. The oval was as close as we could get to an "official" debris footprint.

We used a free ArcView 3.x extension called Two-Theme Analyst. I laid the approximated debris footprint over Texas Census data and then used Two-Theme Analyst to approximate the number of people living within the debris oval.

We ran the analysis using Census data at the block, block group and tract level. We got three sets of very similar numbers. In the end, we went with block groups because it worked best for the maps we ultimately published in the newspaper.

We also shifted the footprint northwest along the flight path to get "what-if" numbers for a break up over Dallas-Ft. Worth. I then re-created the maps in ArcView 8 and worked closely with graphic artists to convert from ArcView to Adobe Illustrator a set of maps that showed readers the dangerous consequences of the two scenarios.

That could not have been done in words alone — and certainly not without the computer analysis.

We ran a pair of population density maps side by side in the paper that we prepared with ArcView. One showed the real debris field. The other showed the scenario. The simple contrast of population density made a very clear point about NASA's good fortune. Detailed counts of people and homes were given too.

Our online staff turned the maps into a telling interactive presentation on the Web. See the story at *www.floridatoday.com/columbia/columbiastory2A52178A.htm* .

Meanwhile, Halvorson and I also were doing the research and reporting to frame the issue in the appropriate context.

We wanted to show how NASA fretted in the past about where debris from satellites the size of school buses might hit. Yet for weeks leading up to the publication of our report, the agency said it never deemed necessary a study of what might happen with a craft the size of a jetliner. Officials explained the space shuttles were designed to survive re-entry and the chance of a breakup was so remote that studying where debris might fall was neither necessary nor practical.

Our story acknowledged NASA's options are limited. There are only so many paths the shuttles can fly to get back from space. But there are options to reach landing strips in Florida and California while substantially reducing the number of population centers beneath the flight path.

Within a week of the story running, the Columbia Accident Investigation Board said it was looking for contractors to study the issue we raised. NASA Associate Administrator Bill Readdy told *The Washington Post* the agency commissioned a similar study. Neither will say if the studies were initiated before our report or because of it.

The official government investigation report will ask if public safety should have dictated that NASA try to land Columbia in California that morning instead of flying over so many major population centers — from San Francisco to Dallas. NASA is not finished studying its re-entry options for the future, but investigators agree changes to shuttle flight rules are inevitable.

Contact John Kelly by e-mail at jkelly@flatoday.net

***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: JOBS, PAY & PERKS

# Combing through data finds top-paid workers

By Hanah Cho, *Los Angeles Times*

The story was supposed to be simple. My task was to update a 1990 analysis by Capital News Service, a wire service operated by the University of Maryland, examining the highest-paid public employees in the state. The analysis then found that the highest-paid state employees were mostly white and male and came primarily from the academic world, particularly from the public medical schools.

After preliminary research and interviews, I determined that state employee salary data in Maryland would come from two sources – the state Office of Personnel Services and Benefits, and the University System of Maryland. (Looking back, after attending a NICAR Boot Camp this year, I also should have contacted the state comptroller's office for salary information.)

I didn't have much time, maybe about two months, to finish the project. So, after pinpointing the right contact person at each agency, I immediately filed public information act requests with both agencies. I was very specific in my request and asked for a number of items, including the top 100 highest-paid employees, names, salary, job title, department, gender and race, even though I was unsure I would get the latter because race may be considered private information. I also asked for the information in electronic form, such as Excel or Access. If I had had more time, I would have requested the agencies' complete payroll database so I could perform a more extensive analysis.

The university system complied with the request within weeks, providing the information in an Excel spreadsheet at no charge. The state Office of Personnel Services and Benefits wasn't as efficient or quick. The agency's lawyer argued that my request would essentially create new records, since it would require the agency to reprogram its database – something that's not required under the public information act. The lawyer also argued that my request imposed a burden on the agency because it would require extra work.

The lawyer was strict in her interpretation of the law, while I believed the law left room for negotiation. For example, Maryland's open records act ultimately leaves the choice of format up to the agency. However, the law also urges the agencies to voluntarily agree to the requester's choice of format if it does not require much work or significant cost. I knew my request would take a simple query or two, especially since the university system appeared to have no problems with it.

**I found that the database did not tell the complete story.**

So the lawyer and I wrangled back and forth on the law's intent. As the last straw, I put our position in writing to the lawyer as well as to the Maryland state attorney.

Meanwhile, I began analyzing the university employees' data, a table that included all the requested fields except race. The seven columns had the last name, first name, salary, title, department, university and gender. The first thing I did was make a copy of the database, a point that was drilled into my head by my computer-assisted reporting professor, who warned of the dangers of working off the original file. And, although the database was fairly clear, I ran an integrity check to see that there were in fact 100 rows representing the 100 highest-paid employees and whether there were any null fields.

I found discrepancies in how the names of departments were entered into the database, a quirk that I had to straighten out. I manually typed in the full name of departments by individually going through the department fields, because at that point my query skills were pretty basic.

After that, I imported the file into Access to run some queries, such as group by department and gender and count. I noticed something immediately. The University of Maryland men's football and basketball team coaches – Ralph Friedgen and Gary Williams, respectively – appeared quite low on the list, even though the media had reported that they make salaries in the millions. So I inquired about the coaches' contracts. But the university's athletic department refused to disclose contract details, arguing that the university was required only to release state-funded salaries.

Yet the state law makes it clear that bonuses and performance awards, which Friedgen and Williams most likely received, are considered public. So I filed a public information request even though I wasn't sure I would get the information before my story deadline. My editors also talked to the lawyer.

About a month before my deadline, the state finally relented and gave me its top 100 highest-paid state employees in the form I requested. After the databases were combined (which I did manually by copying and pasting, since most of the highest-paid employees came from the university), it was easy to see that the highest-paid employees came mostly from the

*continued on page 16*

**SPOTLIGHT: JOBS, PAY, & PERKS**

# Analysis spots hiring during state crunch

By Spencer Hunt, *The Cincinnati Enquirer*

In the midst of a growing budget crisis, Ohio Gov. Bob Taft outlined his short-term plan to help keep the state's finances in order.

In addition to cutting back spending at state agencies and raiding the state's "rainy day fund" for $350 million, Taft told reporters he would halt hiring. "With the exception of people critical to health, safety and those involved in revenue-generating activity, we will put all hiring on hold effective immediately," the governor said.

That statement, an incorrect tip from a good source, and several e-mails from the state's payroll system helped lay the groundwork for an *Enquirer* analysis that showed state departments and agencies hired thousands of workers, including several highly paid administrators, while the budget continued to tank.

The tip was a complaint that Taft's own cabinet leaders ignored his freeze and hired new people "left and right." It turned out that the tip was wrong, but that rumor had a kernel of truth.

A few months after he announced the freeze, Taft lifted it. That was on July 1, 2002, the beginning of the state's new fiscal year.

There was still a story to write when a cursory glance at state payroll data in an Excel spreadsheet revealed more than 5,000 people had been hired or promoted since July.

These hires were made while dwindling income and sales tax revenue led Taft to cut spending again. This time he eliminated $140 million that would have gone to schools and universities. He'd also recently announced his plans to raise more than $3.1 billion in new taxes.

The story, published March 30, showed 3,859 people had been hired or promoted to new jobs since July 1. Nearly 1,000 of those people were paid for work not directly linked to health care, public safety or highway maintenance at a taxpayer cost of up to $39 million a year.

The analysis was done in Excel, using data obtained under the state public records act from the Ohio Department of Administrative Services. This agency handles the payroll for all non-university state employees.

Getting data was not a problem. Getting the best, most useful data was. The data arrived by e-mail in delimited text at no cost, usually within a day of a request. The spreadsheet had about 5,000 rows, one line per person.

The first spreadsheet showed the names, agencies, hiring dates and hourly pay rates. I eventually got a spreadsheet that included a more detailed job description, instead of a code, and columns showing if the person hired was still employed.

I sorted the data by job description, name and hiring date to identify people and take out those who weren't included in the analysis. For instance, there were hundreds of college interns and temporary state fair workers who had to be cut out because they worked for only a matter of weeks.

There were people who had been hired under a federal program that supplied all the money. These hires would have occurred regardless of the freeze. Since the payroll records didn't identify which jobs were paid by federal funds, I had to do some additional reporting by calling agencies that had hired a significant number of new employees.

In several cases in which more than 20 people had been hired during the same month with the same job description, it turned out that the hiring was done with funds from a new grant or federal program. Additional reporting with officials in other agencies turned up even more federally paid employees

At the end of this process, 3,859 employees remained. Because state agency officials argued without exception that every hire would have been made under the freeze, my editors and I had to come up with our own definition of an essential employee. We considered anyone directly involved in making sure others were healthy and safe to be essential. We included prison guards, highway patrol cadets, cooks, hospital workers, mental health workers, workers at state institutions for the mentally retarded, nurses, doctors and psychiatrists. Psychiatrists were a tough call because almost every one was paid more than $100,000 a year. We also included highway maintenance workers, because it's essential to keep the roads in good repair.

That eliminated another 2,870 employees. Finding out exactly how much money the non-essential state employees cost taxpayers was another problem because the data did not show whether they worked full time or part time. Other states' payroll records may show this, and anyone else thinking about doing a similar story should certainly ask for it.

In this case, the $39 million figure we arrived at had to be qualified. We wrote in the story that the figure is probably high, but we also decided not to include the cost of health insurance and other benefits, which payroll officials say tack on another 30 percent onto the base pay.

Once the data was in shape and I had created a separate spreadsheet showing only the non-essential hires, the remaining analysis consisted of simple sorts in Excel. Sorting by pay revealed the most expensive hires. Sorting by date showed who had been hired on or shortly after July 1.

*continued on page 15*

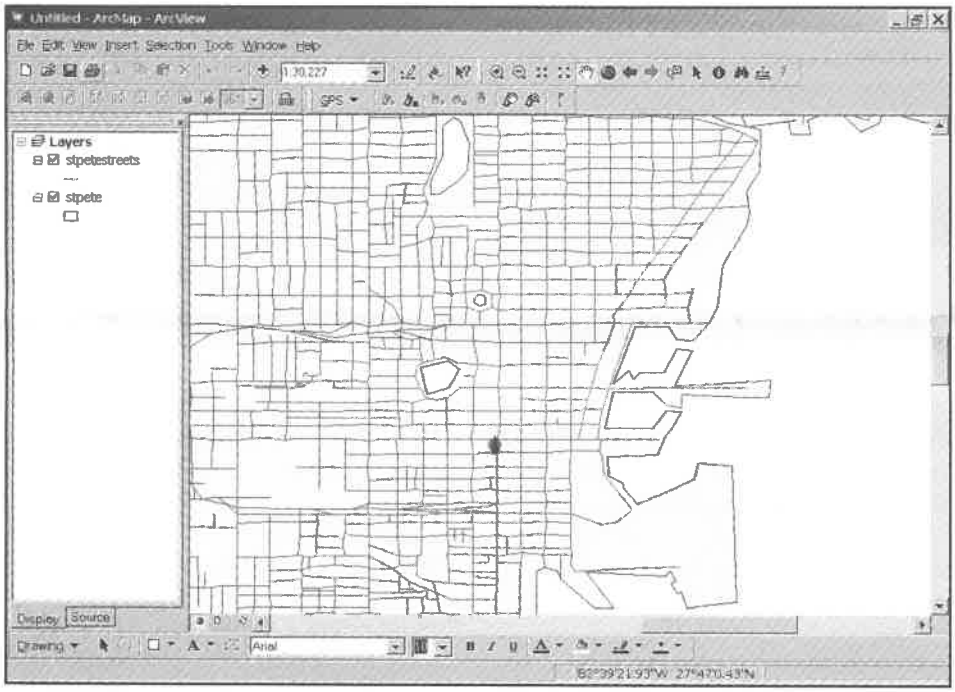
CAR TOOL

# GPS units add depth to reporting

By Matthew Waite, *St. Petersburg Times*

You won't hear me complaining about military spending anytime soon. The U.S. government – bless it – dropped $90 million to add another Global Positioning System satellite at the end of March, and I gleefully paid my taxes two weeks later.

Giddy about paying taxes? Sure. I'm happy to contribute my small sum to the $2.5 billion Global Positioning System constellation of 28 satellites that surrounds the planet and blankets it with low-power radio waves. Before you report me to the mental health authorities, let me explain why I'm so happy: It helps me do my job as a reporter.




I believe GPS is the best little-used tool in journalism. A GPS receiver, mapping software and a little know-how are all you need to do your own geographic analysis. Think of it as building your own databases, only with maps.

You can do this all very quickly and accurately. In the past, a reporter going after a plane that crashed in a remote area had to try to explain the accident location over the phone. Now, a reporter with a GPS unit can, depending on how close he gets to the crash, plot the crash location within a few feet.

GPS is a handy tool on deadline and off. Off deadline, I used GPS to create what the state of Florida lacked a few years ago: a map of traffic signal locations. The state had only engineering drawings and computer-aided drafting files showing the locations. So I decided to create a map. I got in the car with my GPS and my wife at the wheel. Within an hour I had a point map of every traffic signal on a stretch of U.S. 19, a deadly road I was writing about. That map became an important part of my analysis in ArcView, because I could layer deaths and traffic control lights. On one section of roadway with the most traffic control lights, there were no deaths in more than a decade. Of all deaths, more than 80 percent were people crossing away from crosswalks. (See the story at *www.sptimes.com/2002/01/20/Pasco/At_any_time__crossing.shtml)*

On deadline, GPS can help locate events or create maps of new roads.

Here's a step-by-step example of how I used GPS for a daily news story:

One night, after deadline, I heard an emergency search call on the police radio scanner for a missing elderly man. This happens all the time in Florida, so I wasn't very interested. Then I heard someone call for a U.S. Coast Guard helicopter to pull him out of a swamp in the middle of the night. That sounded like news.

So I went to the scene – a golf course covered in the darkness – and watched as a huge helicopter hovered over the man, lowered a basket down to a group of sopping wet sheriff's deputies and hoisted him away to a local hospital.

That night, I grabbed as many interviews with deputies that I could, and I grabbed a scorecard for the golf course.

Early the next morning I returned with my GPS unit and laptop. I plotted the points where the man went into the swamp looking for golf balls and where the sheriff's deputies went into the swamp to find him. I also plotted the points of the tee box and cup from the nearest hole to where the man was found. Why do this? You can get details like distances and reference points. And, using a satellite map of the area, you can also make great graphics. (See the story, without the graphic, at *www.sptimes.com/2002/04/09/TampaBay/Off_the_course_and_st.shtml)*

There are several ways to get GPS data from your unit into your geographic information system software. The one I used until recently is Maptitude 4.05 or later. Within the past few months, ArcView 8.3, through a free download, has added GPS support that rivals Maptitude's. ArcView 3.x

users can find shareware and free scripts for almost any GPS device, though they are not nearly as easy or intuitive as either Maptitude or ArcView 8.3.

Maptitude (*www.caliper.com/maptovu.htm*) and now ArcView 8.3 (*http://support.esri.com/index.cfm?fa=downloads.samplesUtilities.viewSample&PID=43&MetaID=537*) allow you to plug in your GPS unit via a serial cable (or USB, depending on the unit) that's often included with the unit. It plots your location live on a map. In one menu, you can configure and control your GPS unit, and control what data is displayed.

Using either software package you can record events (points) and save them in various GIS formats. You can use your GIS program's distance measuring tool to calculate the distances between points. That's how I measured how far the elderly man on the golf course was from the tee box.

Believe it or not, GPS units aren't just for recording a position. One day I covered a new toll highway that was opening to traffic. Many, including my editors, wondered what the road would be like, how fast you could go from point A to point B, and so on. I went out with my GPS unit and drove my county's leg from one county line to the other several times. The first time I did it mostly to make a GIS map of the road that I could include in other maps of the county. But my GPS – a Garmin GPS III+ — also records speed: how fast was I going when the plot was taken. A few more times up and down the road, with traffic, stopping for tolls, and I had a pretty good average speed for the roadway.

The analysis gave readers one sentence that no one else had: Average speed on the new road was 58 in a 65 mph zone. Not bad when you consider a driver has to stop twice unless his vehicle has an E-ZPass electronic tag that allows automated toll collection.

Some journalists like to call those statements computer-assisted paragraphs, and GPS is great for them if you can get used to operating your unit quickly. For instance, I covered an experimental-plane crash deep in a nature preserve. We had to get an escort from the sheriff's office to take us nearly two miles into the preserve where the plane went down. The experimental plane apparently lost power and tried to land on a nature trail. The cut in the trees wasn't wide enough for the wings, and you can figure out the rest.

With my GPS, I was able to plot my location near the downed plan and the takeoff location (the distance was less than a half mile). I told our graphic artist about the location and they created a graphic that ran in the paper. I guess you call that computer-assisted graphics.

The applications for GPS are as limitless as the uses of GIS. With GPS units starting under $100, this is an inexpensive tool for the newsroom, one that we as a profession are missing out on.

Contact Matthew Waite by e-mail at waite@sptimes.com.

## readme.txt

You can find many good GPS tutorials on the Internet. This link contains a collection of information for beginners: *www.edu-observatory.org/gps/tutorials.html*.

## Buying a GPS

In a few short years, the GPS market has taken off. Buyers will find plenty of choices and a range of prices to fit almost any newsroom budget.

If you have a few thousand dollars to burn, Trimble and ESRI offer a promotional deal on a Trimble GPS, a Windows Pocket PC and ArcPad, ESRI's GIS for Pocket PC. At the other end of the spectrum, you can find a Magellan GPS 310 for $99 on Amazon.com.

What you should buy depends on how you plan to use the unit. Are you, the GIS user, going to be the only one using it? Or are you going to circulate it in the newsroom? Do you need a map on screen, or can you live with just seeing coordinate points?

The basics: If you don't already have one, you'll need a GIS. You also will need a GPS unit with a PC connection of some kind. Most, but not all, have them. You should probably get one that's a little more rugged, since journalists seem to have a habit of dropping electronics.

The extras: If you want to save a considerable amount of money and are comfortable dealing with latitude and longitude coordinates, buy a GPS unit without a built-in map and screen. Low-end GPS units, which take position points just as well as more expensive models, don't include a scrolling map on their screen, just a compass. However, if you plan to hand the unit to novices, having a screen with a map on it is a good idea: reporters tend to be more comfortable with the idea of seeing a map, pushing a button, and having a dot appear where they are. Consumer model GPS units with maps are getting more reasonably priced. A Garmin eTrex Vista, a popular consumer model, costs about $300.

# Tech tip...

*continued from page 13*

the end of the rankings and move it to Column B where it belongs. To do that, you could do a LEFT function for Column A (example: =LEFT(A82,2) for two-digit rankings, =LEFT(A104,3) for three digits) and a CONCATENATE for Column B. (Insert a new column and type =CONCATENATE(RIGHT(A82,1)&B82).)

So, is there any way to import data into Excel and keep the exact formatting of columns? The good news – if you're willing to break the process into steps — is that yes, it's possible with Acrobat Reader 6.0. It has a new feature that lets you select a column of text. Just click the "Select Text" tool and move the cursor toward the column until it looks like a vertical bar with a box around it.

Then highlight the column (See figure 8), copy it, and paste into Excel. With steady mouse skills, you can even strip the column of rankings and keep them separate. Be creative – sometimes it's easier to copy a column starting somewhere besides the upper-left hand corner.

It works almost perfectly. The one big goof was that the column for newspaper ownership was blank in a couple of cases, but the data pasted into Excel without the spaces – so it stopped two rows short. That means you have to go back and reinsert the blank cells where they belong. You can also import the column headers with the data, but that requires tinkering, too. For instance, "Rank in top 200" filled up four vertical cells in Column A, just as it looks in the original.

The other big drawback is that the Column Select tool highlights only the data that appears on your screen. So for this table, which runs six pages, you'd have to select each column and paste for each page. This is okay for a small table – but if you've got hundreds of pages, you might want to stick with one of the free downloads.

The Challenge included a third document that several lists of the top-scoring schools in Missouri, motop10.pdf. The same methods work with that document that worked with the ASNE table. Willis, for one, copied and pasted the data into UltraEdit and replaced the ending parentheses in the school column with tabs, coded in the program as ")^t". He then replaced every number with spaces with tabs, coded in UltraEdit as ^1^t Willis imported that file into Excel, added the header row and saved it. (Note: UltraEdit costs $35.)

In summary, here are some tips to keep in mind:

- Make sure you've got the most recent version of Acrobat Reader. Check back for updates. Here's a link for the latest version: *www.adobe.com/products/acrobat/readermain.html*

- You can convert documents for free or close to nothing (less than $40), especially if you're willing to do a little cleaning.

- Yes, cleaning. No matter how you import something, chances are it won't be perfect. Spot-check your imported version with the original. Also be prepared to do a little work to get the column headers in the right place.

- Experiment. You can import the same file using more than one method.

With some practice and patience, you might find that importing PDFs into Excel is almost as simple as A-B-C.

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

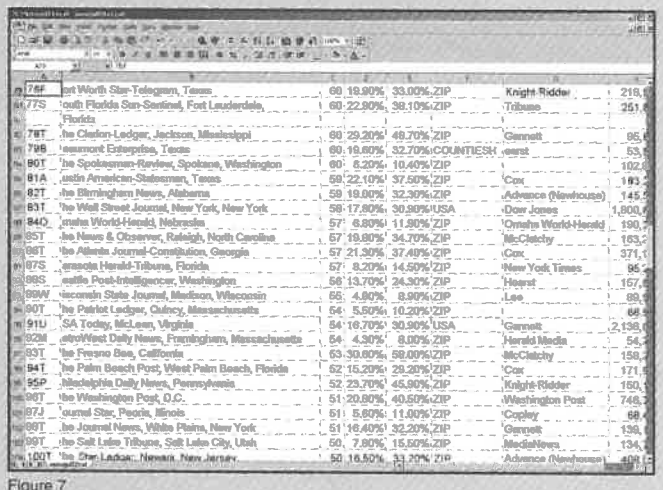


Figure 7

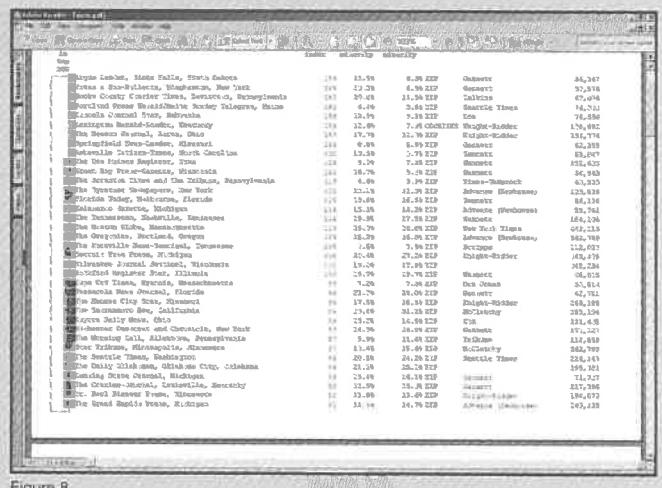


Figure 8

SPOTLIGHT: JOBS, PAY & PERKS

# Data for covering workers, compensation

By David Herzog, *University of Missouri and NICAR*

Journalists can use several key databases to cover public employees, their salaries and benefits. Here are a few of the most valuable and what you can expect to find inside them:

## Payrolls

Payrolls list the names, departments, jobs and pay of government workers. The workers' hiring dates, pay rates, work status (part- or full-time) also should be included. Check the pay field carefully because agencies sometimes mix the hourly pay rates of wage earners in the same field with the monthly or weekly salaries of professional workers. Also, make sure you request the most detailed pay records you can get so you can look at overtime pay.

Depending on your state's open records law, you might be able to get even more detailed information about workers, including home addresses and dates of birth. The more detailed information you can get, the better.

Request this data from agency human resources or personnel offices. Sometimes finance offices maintain the data. Update this data every year so you can build an archive that shows salaries over the years.

## Pensions

These databases list the names of pensioners, last date of employment and amounts of annual pensions. Many of these databases include the addresses of the pensioners and the types of pensions (disability or retirement).

Depending on the complexity of the retirement system, there may be several types of pensions with distinct rules and cost-of-living adjustments. If there is more than one pension system, make sure your data includes a field that indicates the type of pension.

Journalists have used pension data to show how some pensioners get more money off the job than when they worked. They've also used it to project how much pensioners will collect in the future.

You can request this data from agency pension boards. If they do not keep the data, try the human resources, personnel or finance offices.

## Check register

The check register can be a great place to find information about public agency worker travel. A check register is a database that contains a record of every check issued by a government agency.

Registers often include the check number, name of person or entity that the check was issued to, date, amount, department and purpose of payment (often a code). Sometimes the data includes an employee identification number.

Journalists can use this data to scour for employee travel, which may or may not be legitimate, and reimbursed business expenses. In addition, journalists can use the data to look for payments to consultants and contract workers who are not paid through the usual payroll channels.

Try requesting this data from the treasurer, comptroller or another government fiscal officer.

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

# Hiring

*continued from page 9*

In a few cases, expensive employees were hired right after the freeze was lifted. The prime example was a Department of Education employee who was hired July 8 for $120,000 to help make sure state education reforms dovetailed with the federal "No Child Left Behind" initiative. Sorting by job description showed how many of each type of employee was hired. Sorting by state agency showed which sector of government did the most hiring. Ohio's lead welfare agency, the Department of Job and Family Services, hired 140 people.

I also used Excel to calculate the average cost of a subset of new hires. For example, I determined that the Department of Education hired 28 education consultants, people who act as liaisons with local school districts, at an average $53,500 a year.

This kind of analysis takes some time – especially if you're learning just what payroll records can and cannot show you along the way. The result, I believe, is a story that offers a perspective on state finances and spending that only an independent observer can provide.

Of course, you will also have to deal with state officials who probably will argue that every single person they've hired was out of an absolute necessity.

The best response came from the Ohio Department of Taxation, which hired 55 auditors in December and January to beef up collections. While each auditor was paid an average $34,500 salary, they claimed each collected about $457,000 in unpaid taxes.

That prompted a spokesperson to say, "They are more than carrying their weight."

Contact Spencer Hunt by e-mail at shunt@cincinna.gannett.com.

# Salaries

*continued from page 8*

university's medical schools, a finding that was true back in 1990.

The breakdown was almost too good and too easy to be true. So I did more reporting, calling officials at the medical schools and experts on medical school salaries. I found that the database did not tell the complete story. At face value the highest-paid state employee, a renowned surgeon, made about $600,000 a year from the state. But, as my reporting later revealed, most of the medical school deans and professors also practice medicine, thereby earning most of their so-called state salaries from private sources.

After culling through names of individuals who fit this category, I found that the highest-paid employee, whose $359,570 salary was completely state-funded, was the chancellor of the University System of Maryland. That revelation changed the focus of the story, especially since the salary and deferred compensation package of the incoming chancellor had recently raised questions at a time when the state was trying to close a $1 billion deficit.

Instead of analyzing the database in the aggregate, I decided to focus my story on the chancellor's salary and look at why medical school personnel received such high six-figure salaries, outranking the governor, judges and heads of state agencies.

I also decided to write a separate story on the two coaches whose state salaries did not accurately reflect their overall compensation package. The database showed that the two coaches, Friedgen and Williams, made $183,820 and $207,584, respectively, when in fact, they were making nearly triple the pay of the chancellor. Because I was still waiting for my request to be answered at the time of deadline, I relied on published reports of the coaches' salaries to write my story. The story also touched upon the controversial issue of paying college coaches million-dollar salaries.

The stories I found went further than the original intent of my reporting: to update the 1990 analysis of the 100 highest-paid state employees. If I had taken the database at face value, I would have incorrectly written that medical school personnel received the top salaries, a trend unchanged from 1990. Instead, I found that the university chancellor was the highest-paid employee whose salary was paid completely from state funds.

The salary database did not require complicated queries or have many quirks. But the project was worthwhile because I learned a valuable lesson: No database is a substitute for good, thorough reporting. I also learned it's crucial to know the state's public information law so you can be armed when going into battle with lawyers or agency administrators who don't want to give up information easily.

As a final note, a few weeks after the story ran, the University of Maryland athletic department responded to my request and provided more details of the coaches' salaries, such as the amount of bonuses tied to performance. It pays to be persistent.

Contact Hanah Cho by e-mail at hanah.cho@latimes.com.

## readme.txt

The "State Salaries" articles by Hanah Cho, published while she was a student at the University of Maryland, were finalists in the student work category of the 2002 IRE Awards. To order a reprint, contact the IRE Resource Center at 573-882-3364 or rescntr@ire.org and ask for Story No. 19441.

# Overtime

*continued from page 1*

ficers making at least $90,000. Abundant overtime pay and a contract that allows police to work double shifts every day helped three dozen rank-and-file cops earn more than $100,000. The top-earning cop was a patrolman: Joseph Smith grossed $143,405, with overtime making up $86,215 of his total pay.

The findings, published on May 25 in an article by Matt Burgard, our city police reporter, and me surprised not only Hartford residents, but also police administrators and union officials. The Hartford Police Department did not monitor overtime earnings by individual officers, only by units.

"Is this right?" the police chief asked, when shown a printout of the top overtime earners. Then he turned to an aide and said, "Who the hell is Joseph Smith?"

We uncovered the high salaries by analyzing data that all municipal beat reporters should have on their computer desktops: a city worker payroll list, broken down by department, base pay and overtime.

I requested the data earlier this year with no particular story in mind. At a minimum, I figured the payroll would be helpful in checking names and salaries on deadline.

Obtaining the city payroll data was easy, and analyzing it was simple even for a reporter using an Excel spreadsheet for the first time.

## Laying groundwork

Before requesting the data under the Connecticut Freedom of Information Act, I consulted with the city's information technology director and asked questions. How did the city maintain the records? What would it take to copy the entire payroll file to a CD-ROM? I had made a point of getting to know the director months earlier, asking him to give me a tour of the recently merged city-school board information technology system. By the time I filed my FOI request by e-mail, I knew that producing a copy of the payroll on a CD-ROM was a matter of a few keystrokes for one of his staffers.

Similar requests by *The Courant* have not always gone so smoothly.

A few years ago, the Connecticut State Police tried to charge Jack Dolan, our resident computer-assisted reporting guru, a heart-stopping sum for a copy of its criminal convictions database. The state police figured copies of individual records cost $25 and there were 815,000 records in the database. So they multiplied $25 by 815,000.

The bill came out to more than $20.3 million. Try burying that in an expense report.

The newspaper instead opted for a lawsuit, resolved last year by the Connecticut Supreme Court. As a result, it is well-established law in Connecticut that public agencies can charge only reasonable fees for copying electronic records.

The city of Hartford gave me a copy of its payroll data at no charge.

With brief tutoring by Dolan, I started playing with the city data.

The city provided the data on an Excel spreadsheet at my request. I quickly realized, however, the data was incomplete. The city had provided me every employee's salary, not the actual pay. Missing were overtime and longevity payments, small bonuses paid for years of service.

I amended my FOI request to include all payments to employees, including longevity payments and overtime. This time the I.T. director gave me the bad news: Marrying overtime records with the payroll would require staff time for which he would have to charge me. We agreed on a price not to exceed $100 – a bargain for 6,700 payroll records.

The payroll records listed each employee in alphabetical order with six other columns of data: department, regular pay, overtime, longevity pay, other pay and gross pay.

Sorting by salary, I created a list of the city and school employees in descending order of gross pay. The list showed rank-and-file cops making more money than the city manager, police chief and other department heads. To find all the top overtime earners, I sorted again by overtime.

## More reporting

We put the analysis into a broader context by reporting more. Opportunities for overtime had been increasing every year as the department's force had been allowed to shrink. The police contract encouraged officers to take overtime when offered, especially as they neared retirement. Police pensions in Hartford are based on an officer's "final average income," taken from the best three of the last five years before retirement.

Patrolman Smith declined to talk to us, though we were able to answer the chief's question: "Who the hell is Joseph Smith?" It turned out that Smith had achieved a measure of notoriety under a previous administration. He was accused of loafing on the job and stealing city time. He was reinstated after a jury acquitted him in 1999.

Officer Plummer Carroll, who finished just behind Smith with $80,606 in overtime, was happy to talk about his industrious ways. He took all the overtime offered to run up his pension calculation. He estimated that he averaged about 80 hours a week, with a high of 110 hours, over three years.

"I missed out on a lot of beautiful things: holidays, weekends," Carroll said.

His reward: an annual pension of $74,093 – nearly $20,000 more than his base pay of $56,095.

Contact Mark Pazniokas by e-mail at mpazniokas@courant.com.

# Back tax

*continued from page 1*

Since Charlotte County stores delinquent taxpayer information on a mainframe system, I asked for the data in ASCII. They provided the entire database for free. I imported the text into Microsoft Access 97.

Not long after beginning a search through all the records, I discovered my first roadblock. The tax collector kept detailed information on each delinquent, but surprisingly, no one had tallied up how much people owe in back taxes over the years. In fact, I was told no one had ever done a countywide calculation showing how much each delinquent owed in penalties over multiple years. Also, because interest kept adding up, the dollar figure changed every day. But the tax collector's database showed only how much delinquents originally failed to pay. To get an accurate picture of the top tax delinquents, I needed to calculate the back tax information for more than 40,000 records.

## Prep work

After talking with the information technology office for the tax collector, I began writing a program in Access to calculate back taxes. To accomplish our task, we had to first calculate how many months had passed since each delinquency began. The more months without a payment, the more in back taxes owed.

I created two new fields in Access and ran the calculations to show the number of delinquent months and the monthly interest rate for each delinquent. We used this information to calculate how much each delinquent owed in back taxes, but I still had to add a $4.25 penalty to each record to cover the cost of advertising the delinquency in a local newspaper. After completing the work, I finally thought we had all the information we needed.

It didn't take long for us to discover I needed to merge the file with another county database to get an accurate picture of local delinquents. The database of tax certificates showed only the name of the person who owned the property the year its taxes became delinquent. We needed a second database to see the name of the current owner. Florida law states the current owner of property is responsible for all back taxes.

I called the property appraiser's office and arranged to receive the electronic list of properties in Charlotte County. I merged the tax collector's information with the property appraiser's list by joining the files using the "strap number" field that serves as a property identifier. Finally, with information about current property owners in hand, we were able to generate the true list of top tax delinquents.

## Beyond the data

Our queries found that one company had piled up more than $950,000 in back taxes on more than 500 properties. With a quick visit to the county clerk's office, we discovered the county's top tax delinquent was going to have a very difficult time paying its bill. Court records showed the company had gone bankrupt and hadn't paid its taxes for seven years in a row. Despite this fact, the county continued to spend tax dollars sending out delinquent notices and advertising its delinquency in the newspaper year after year.

This struck us as odd, since Florida law allows county governments to auction property if the owner remains delinquent for more than two years. For some reason the county was bypassing the option to auction the properties off, a move that could recoup money and return the real estate to the tax rolls.

By law, the tax collector must spend tax dollars advertising delinquencies three times a year for each piece of property that qualifies. We began more detailed queries, searching for all the delinquents that hadn't paid for years but were still draining the county's coffers. We calculated the tax collector had spent $251,712 since the mid-1990s trying to collect from deadbeats it could have cracked down on long ago.

We eventually uncovered a total of $8.3 million in uncollected taxes across the county since the mid-1990s. It's money that wasn't getting to essential services such as schools, fire departments and road construction crews. More than $1 million of that would have gone to the Charlotte County Fire Department. Fire Chief Dennis DiDio told us it could have added new fire trucks, firefighters and equipment to fight bio-terrorism.

## Putting it together

Further research revealed Florida law requires county governments to auction off all land worth more than $5,000, if the land has been delinquent for more than two years. If it's worth less than $5,000, counties can decide for themselves if they want to go after the property. The $8.3 million in uncollected revenue we discovered was largely made up of properties worth less than $5,000.

We found the county failed to put together all the pieces on its own. It looked only at how much each delinquent owed on individual pieces of land. No one ever thought to keep track of just how much delinquent land existed under one person's name. We eventually discovered systematic proof that the county wouldn't crack down on a majority of delinquents who owned land assessed at $5,000 or less.

We took our findings to the tax collector and the real property manager for Charlotte County. The tax collector told us she wanted to immediately respond by auctioning off all the delinquent land we discovered. However, that wasn't her decision to make. The real property manager decides which land to sell at tax auction. When we showed him our findings, he pledged his support for the tax collector's idea. The real property manager told us he would begin the process of auctioning off all the delinquent properties that qualified for the program.

The county clerk has agreed to postpone some of its normal fees to speed the sale of the large number of properties that will be auctioned off on the courthouse steps. The Charlotte County tax collector and the real property manager credit WBBH for discovering a potential source of millions of dollars in uncollected revenue. We're told Charlotte County will complete its auctions over the course of the next year.

Contact Mark Greenblatt by e-mail at mark.greenblatt@nbc-2.com

## readme.txt

The IRE Resource Center has copies of stories and tipsheets offering ideas for using CAR to analyze municipal and county tax records. Here are two:

Story No. 20151: In 2002, *The* (Newark, N.J.)*Star-Ledger* analyzed local tax data and found that "the average residential property tax bill in New Jersey jumped almost 7 percent this year, the largest one-year increase since 1990." The story also discusses the overall rise in taxes from suburbs to rural areas.

Story No. 19005: A 2001 *Philadelphia Inquirer* data analysis of millions of property tax records found disparities between the amount of taxes paid relative to residents' wealth. They discovered that in some cases, people living in lower-income areas were paying more in taxes than those living in higher-income areas were.

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

# Incumbent

*continued from page 1*

Why is steak sauce important to a political campaign? That was the key element of my first computer-assisted reporting project: A question.

Finding New York's campaign finance database online was easy. After that, I had to fall back on my Boot Camp training. Two pieces of advice became immediately useful – even critical – in doing the story:

First, find a buddy to review your work. In my case, I tapped Brad Heath. He's a reporter in Detroit now, but he was my predecessor as Binghamton's CAR person and understands my experience as a reporter and my virtually nonexistent database skills. In fact, he knew the data I was working with, too, because he had briefly played with it.

Second, give the database to the source. The campaign treasurer had done all the paperwork in filing the information, and it was in his interest to make sure I got my facts straight. He found a couple of errors in my queries. And with more than 3,000 potential sources, I still kept total control of what the story said.

I learned a few other lessons.

Do everything two or three times. It wasn't until I exported my query results to Excel and checked the math that I realized the numbers didn't add up, and it was because of a problem in my query. At Boot Camp, they called these data integrity checks, and they are important.

Review the data. The quick and easy task is to find the extremes in your data, and that frequently can be useful. But go through the rest of it, too. You'll come up with more questions. Why, for example, couldn't I find the local Republican leaders contributing to the area's most prominent Republican, but I could find their spouses? You're not going to notice that unless you know both the data and your subject.

I'm primarily a beat reporter so that kind of analysis was easy, even fun, once my co-workers and I made a game of finding how many names we recognized. I don't know how a reporter could do that coming into the data cold.

I also had to learn what the data wasn't telling me, which meant I had to learn what it was supposed to say. Candidates for New York's state legislature generally contribute to a party "housekeeping" fund to pay for legal advice, polls and whatnot. Other politicians had contributed to it, according to my analysis, but the state senator hadn't. Either he was the state's biggest political cheapskate (unlikely) or there was another fund out there the data didn't have a record of.

I had to learn to keep focus on my story – how the senator's campaign spent its money. There were plenty of other tempting angles to pursue. Why was the ambassador to Grenada my boy's biggest individual contributor? And was the fact that he was indicted about 15 years ago (and later exonerated) on charges alleging he blew up trailers of a union protesting one of his facilities relevant? Fascinating stuff, but it wasn't part of the story I had in mind. Maybe next time.

That's the lure of a good database. It gives you a garbage load of ideas. You can't do them all. Even if you could, you probably shouldn't. Overkill can be a nasty habit, and with all the time I sank into this project, it was a temptation.

**The story described how campaigning works and what it means to a community.**

And that was the toughest lesson of all. Making the database work took a solid 20 hours – as much as I normally devote to a basic, no-frills enterprise package. And because a database generates about three numbers and a nut graf – and maybe some charts or lists or facts for a box – that meant that once I found the good stuff, I still had 15 or 20 hours of work remaining.

CAR stories are inexpensive compared with other investigative stories of similar depth, but they're inefficient compared with traditional methods. That can be an issue at a community newspaper, where cutting somebody free from the daily need to feed the beast can be a burden on everybody else. Fortunately, Heath taught my editors that so I didn't have to.

Through all of this, Heath double-checked my work and cheered me up when I felt dumb. With his help, I learned to rewrite queries to make them as simple as possible. The easy