

EDUCATION

Charting test cheaters

By Holly Hacker
The Dallas Morning News

"We just worked real hard."

That's how the principal of Wilmer Elementary near Dallas explained her school's astonishing performance on the state third-grade reading test last year.

Wilmer finished No. 1 in Texas, ahead of more than 3,000 other schools. Amazing.

A little too amazing, in fact. Standardized test scores are a big deal in Texas, as elsewhere. States and the federal government rate schools based on test scores. In some schools, bonuses or promotions hinge on good results. In past years the state had rated Wilmer low-performing or acceptable, nothing higher. So how could scores climb sky-high in a year?

Our efforts to answer that question turned into a statewide analysis of test scores. We found nearly 400 schools with unusual swings from one test to

continued on page 20

SPOTLIGHT: WATER & WATERWAYS

Boat accident data sheds light on dangers

By Natalya Shulyakovskaya, *The Orange County Register*

Three local teenagers were killed and one was severely injured in a hit-and-run collision on a narrow stretch of the Colorado River called the Parker Strip at the end of the 2003 summer boating season.

The *Register* first reported the story by talking to the kids' friends, boaters, locals and law enforcement officials

who patrol the waters on the Arizona and California sides of the river and its

continued on page 10

SPOTLIGHT:

For more about water & waterways see:

- Tracking hurricanes' paths, p. 8
- Data and tipsheets for your stories, p. 10
- Synopses of recent stories using CAR, p. 12
- The May-June 2005 IRE Journal

OUTSOURCING

Finding overlap in public-private interests

By Catherine Rentz Pernot, *IRE and NICAR*

From security in Iraq to voting machines in the United States, the private sector's work with the government is taking an increasingly visible role in the lives and finances of Americans. Below are three ways journalists have recently reported on government contracts with private business using computer-assisted reporting.

To investigate the politics behind state government transportation contracts in Wisconsin, reporter Patrick Marley of the *Milwaukee Journal Sentinel* did a fairly simple analysis of two databases.

For the article, "Transportation contracts jumped 45% in 5 years," Marley compared five years of historical transportation contract data given to him by the Wisconsin Department of Transportation with a campaign contributions database maintained by a local nonprofit. Once he obtained the data, Marley said it took him less than an hour using Microsoft Excel and a Web-based database to make two large discoveries.

First, from the government's contract data, Marley found out that engineer-

continued on page 18

Bits & Bytes

Fresh OSHA data

The IRE and NICAR Database Library has updated the Occupational Safety and Health Administration inspections database to include records into February 2005. The data contains more than 3 million records of workplace safety inspections since 1972.

The OSHA data consists of 10 tables detailing workplace violations, penalties and accidents. Journalists have used previous releases of the data to investigate the safety of local businesses, and the effectiveness of OSHA inspections and enforcement.

Many of these stories and tip-sheets about investigating workplace safety are available through the IRE Resource Center.

Contact the Resource Center at 573-882-3364 or rescntr@ire.org.

Journalists can order the data for a state or the entire country by contacting the Database Library at 573-884-7711 or 573-884-7332.

For more information about the OSHA workplace safety database, or to download an order form, visit www.nicar.org/data/osh.

Campaign finance

Journalists can tap in to the resources inside IRE's Campaign Finance Information Center (www.campaignfinance.org) to

continued on page 4

Time to say thanks to our writers

By David Herzog, NICAR and Missouri School of Journalism

Uplink is a true grassroots publication.

Look at any issue and you'll see articles written by journalists from news organizations big and small. For this issue alone, journalists from major market television, a nonprofit reporting center and a diverse mix of newspapers have pitched in.

We don't pay our contributors a dime for their work. They generously share their stories so others can learn from their successes and their missteps. And they share their contact information so if you want to ask a contributor

some questions, you can simply shoot them an e-mail.

Our contributors provide a great service to journalism by giving their time. Starting this issue, we're going to tell you more about their backgrounds and accomplishments in our new "About our contributors ..." feature (see below). If you're looking for their e-mail addresses, you can find those where they've always been: at the end of the articles.

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

About our contributors ...

Jeff Claassen is the CAR specialist on the *Fort Worth Star-Telegram* investigative team. He has used data analysis to report on pollution and sports. Claassen is a 1994 alumnus of the IRE and NICAR CAR computer-assisted reporting boot camp and the 2004 advanced statistics mini-boot camp.

Dani Dodge recently joined *The San Diego Union-Tribune* as a general assignment reporter. Before that, she reported for the *Ventura County Star*. Dodge is a 2005 alumna of the boot camp.

Ron Nixon is the CAR editor for projects team at the *Star Tribune* in Minneapolis, Minn. He is the former training director for IRE and NICAR. Nixon is a 1998 alumnus of the boot camp.

Natalya Shulyakovskaya is an investigative reporter for *The Orange County Register* who specializes in data analysis. She is a graduate of the

Missouri School of Journalism and a former analyst in the IRE and NICAR Database Library. Shulyakovskaya is a 2000 alumna of the boot camp.

Holly Hacker is the higher education reporter for *The Dallas Morning News*. She is a former analyst for the IRE and NICAR Database Library and a graduate of the Missouri School of Journalism.

Matthew Waite is a reporter for *The St. Petersburg Times* who specializes in computer-assisted reporting and the use of geographic information systems (GIS). Waite is a 1999 alumnus of the boot camp.

Matt Williams is a government reporter for the *News & Record* in Greensboro, N.C. He is a graduate of the Missouri School of Journalism, where his studies included computer-assisted and investigative reporting. His Web log, *The Inside Scoop*, is available at <http://blog.news-record.com/scoopblog>.

FIRST VENTURE

Uncovering bad bridges at risk in mudslides

By Dani Dodge, *Ventura County Star*

I had just jumped onto the hotel treadmill in Columbia, Mo., for a quick morning workout before computer-assisted reporting boot camp started when a picture of my hometown fire chief flashed on the television news.

Record rains had triggered massive mudslides back home in Ventura County, Calif. Ten people were dead.

And I was missing the big story. My editors wanted me to finish the week-long IRE and NICAR Boot Camp, so I did what any journalist would have done and figured out an angle that I could begin to cover from more than 1,500 miles away.

My story: Money for bridge maintenance was going down in the state and county as the bridges were aging beyond easy repair. The consequence: An extra surge of water could wash them away.

It had taken nearly a year to persuade my editors to allow me to attend the January 2005 Boot Camp. We were into the third day of training when heavy rains flooded Ventura County, causing mudslides and mayhem. So while other boot campers researched federal Small Business Administration loan records or calculated AIDS rates for women, I used Microsoft Access to delve into the data for bridges and dams. If hillsides were coming down in the rain, bridges might be also. I bought the National Bridge Inventory database from the IRE and NICAR Database Library before I returned to Ventura County.

After I got back, I was quickly thrown into the daily coverage to relieve reporters who had been working 18-hour days. But I also continued to probe the data. The flooding had destroyed two bridges in Ventura County and damaged others beyond repair. Had those bridges been maintained well? What about the other bridges in the county? I looked in the data for

the county-maintained bridges.

When I called county officials, though, they said their bridges were completely safe despite the poor condition ratings and the rains: "Just a little damage." So I went after the paper inspection records. At first, the county public works officials balked: Someone would have to pull them for me, and they were too stretched, they said.

I threatened to file a public records request, but at the same interview I also asked to see the area where the records were stored. Once I had seen how easy they were to access, the county official realized he had no excuse for not letting me go through them myself. From there, the fun began.

In their reports, state inspectors wrote that several county bridges were "unstable" and needed immediate attention. The date on the inspections: 2001. But, according to the reports and interviews, no work had yet been done. County officials could no longer continue to deny problems existed. Looking at some of those bridges and several others in the county that inspectors had rated as structurally deficient, I found one bridge with rotted timbers that I could poke a stick through. I also

watched heavy trucks rumble over another bridge with a visibly undermined foundation — water had washed away the river bottom under the concrete pier that was supposed to hold it up.

I spent two and a half weeks on the story. Only a small portion of that was spent working with the data, but I went back to it often. Several times, I found something in the paper records, and then found a corresponding trend in the data. I realized only after looking at the reports the importance of words like scour, which means erosion.

Deficient Bridges by State and Highway System - Mozilla Firefox

http://www.fhwa.dot.gov/bridge/defbr04.htm

U.S. Department of Transportation
Federal Highway Administration

Bridge Technology

Search FHWA: [Keywords] Go

Deficient Bridges by State and Highway System - NHS Bridges
As of December, 2004

Area is in Sq M

State	# Bridges	# SD	# FO	# Def	Area	SD Area	FO Area	Def Area
ALABAMA	2,672	95	577	672	3,946,778	92,500	680,380	772,880
ALASKA	360	42	46	88	322,268	48,923	49,571	98,594
ARIZONA	2,525	16	160	176	1,812,283	29,681	190,231	219,912
ARKANSAS	1,943	55	283	338	2,504,620	104,135	320,311	424,445
CALIFORNIA	7,422	966	993	1,959	15,057,189	3,336,561	2,245,820	5,582,381
COLORADO	2,005	119	282	401	1,896,936	129,195	335,819	465,014
CONNECTICUT	1,585	59	324	383	2,113,884	270,134	551,320	821,454
DELAWARE	236	4	24	28	599,775	28,565	134,448	164,013
DIST. OF COL.	118	6	61	67	365,635	26,966	185,519	212,486
FLORIDA	3,933	28	478	506	7,766,274	165,853	877,511	1,043,364
GEORGIA	2,464	40	258	298	3,262,765	83,875	256,825	340,699
HAWAII	403	29	148	177	832,543	20,520	109,899	129,419
IDAHO	731	34	138	172	665,135	39,108	79,283	118,391
ILLINOIS	3,801	260	395	655	5,149,333	438,516	927,736	1,366,251
INDIANA	2,487	36	335	371	2,624,942	111,074	436,257	547,331
IOWA	1,792	111	177	288	1,954,949	184,806	260,199	444,806
KANSAS	2,432	89	355	444	2,341,802	81,506	552,274	633,780
KENTUCKY	2,151	50	312	362	2,533,714	144,273	353,563	497,836
LOUISIANA	2,516	94	569	663	9,788,872	585,332	1,588,279	2,173,612
MAINE	496	42	93	135	562,471	59,677	79,825	139,501

answers. I ran queries for the state, neighboring counties and Ventura County. In the queries, I asked about how many bridges were deficient and obsolete. I compared the rates of disrepair. I looked at the condition ratings of bridges.

The data showed that 99 of Ventura County's 485 bridges were deficient or obsolete. Four of the bad bridges were the ones that had been washed out or severely damaged. I discovered that within Ventura County, city- and state-maintained bridges had a much lower percentage of bad bridges than

continued on page 13

Bits & Bytes

continued from page 2

investigate national and state campaigns. The CFIC Web site features a searchable database of federal contracts and campaign finance data, and links to relevant tipsheets and news stories.

In addition, the Database Library has Federal Election Commission data for the 2006 election cycle.

The FEC data include records of all contributions and expenditures reported by federal campaign committees. News organizations can order a subscription to the FEC data that is updated weekly. One-time purchases and archives are also available.

Seleman joins IRE

IRE and NICAR welcomes Pricel Issa Seleman, a 2005 Alfred Friendly Press Fellow. Seleman is a senior reporter for *The (Tanzania) African*, where he also assists with editing and translation from Kiswahili to English.

Seleman has worked for the *African* since 1999, in addition to writing for the *Berliner*, and London-based *Africa Confidential*. In addition to English and Kiswahili, Seleman is fluent in Nyanja and Shona, the languages of Zambia and Zimbabwe.

Seleman is spending six months at IRE and NICAR studying the U.S. Freedom of Information Act, investigative reporting and computer-assisted reporting. He also will be writing for IRE publications and reporting for the *Columbia Missourian*, the community newspaper operated by the Missouri School of Journalism.

EMERGENCY RESPONSE

911 data shows system's shortcomings

By Scott A. Zamost, WTVJ -Miami

It started with an intriguing tip. And it ended with major changes in the works for the 911 emergency response system in Broward County, Fla.

A source suggested we examine a drowning in which a nanny called 911 three times. Instead of getting an operator, no one answered the phone, so she rushed the child to a nearby fire station. The toddler later died.

This was something entirely new. Reporter Patricia Andreu and I knew there had been periodic stories about problems in 911 systems around the country. Most of those stories had focused on delays in getting to locations depending on where someone lived. But none had shown a systemic failure to answer calls in a timely manner using police records.

We first obtained the police file of the drowning, which confirmed the nanny had dialed 911 three times. The police reports did not specify how long the phone rang each time. To get that, we asked the Broward Sheriff's Office, which handles most of the 911 calls for Broward County for the call records the day of the drowning.

Prior to our request, a source told us that every time someone calls 911, a record of the call is logged into the sheriff's computer system. This information, stored in a system called AccuTrak, developed by Gemini Software Technologies, includes the time the call came in, when an operator answered the phone, when the operator disconnected the call and if the caller hung up. That call data would prove to be a gold mine for our story.

At the same time, we learned that any time there is a question or complaint

about a 911 call, an official investigation is opened. We obtained the Broward Sheriff's Office investigative files about the drowning and found remarkable information: The nanny's three phone calls to 911 lasted 37 seconds, 23 seconds and 11 seconds before she hung up and took the child in her own car to the fire station. This was the kind of detail that would form the basis for a broader investigation. If this happened in a drowning where a child died, how long do other emergency phone calls ring before they are answered?

At first, the sheriff's office told us it wasn't sure such records existed. We then made a public records request for the AccuTrak records for the drowning.

We received the paper records for the drowning incident, but there was a problem. It was impossible to decipher them without a law enforcement expert from the agency we were investigating. Some of the lines were coded with information about which operator picked up. For instance, "C08" meant the operator in the eighth position. The public information officer gave us a basic explanation, but we needed more.

We decided to ask for a telephone interview with the head of the police communications division to explain the records and how the 911 system worked. Since they knew we were looking into answer times anyway, we figured knowing this information was critical to deciding whether we had a bigger story.

Over the phone, I spoke with Maj. Paul Lauria of the Broward Sheriff's Office. He explained what the fields meant — when a call came in, when it was

answered and when the operator disconnected it. The records also showed how long “abandoned calls” rang before the caller hung up.

Lauria confirmed the call answer times in the drowning incident. He told me the operators answered calls “as quickly as possible” and there was no standard they had to meet because it would be “self-defeating.”

However, we later learned there was a standard the sheriff’s office agreed to meet: 90 percent of 911 calls must be answered within 10 seconds. It’s a state standard we discovered in a phone interview with Florida’s 911 coordinator. It would turn out to be a key element in our story.

We looked at the AccuTrak records for the day the toddler drowned. The call answer times for that day were well below the 90 percent standard. It appeared we were onto something so we asked for six months of records. The records are kept electronically and the Florida public records law states that agencies should be able to redact non-public information. Nevertheless, the sheriff’s office insisted it could not electronically redact the confidential phone numbers of the callers. We consulted someone in our own information technology department who told us this was possible. The sheriff’s spokesman insisted it couldn’t be done and said the office could provide the records on paper with the numbers crossed out, a process that would take weeks or even months.

In the interest of moving the story forward, we asked for six months of AccuTrak records on paper. The sheriff’s office, as it had done throughout our investigation, responded in a way that kept delaying the story. It told us it would take several months to cross out the phone numbers and estimated the cost could be higher than \$8,000. That was unacceptable to us. We continued to press the office about getting the records in electronic form but kept getting denials.

We ultimately decided to look at two randomly selected weeks, which would amount to nearly 29,000 emergency calls. The state 911 coordinator said that would be an excellent sampling and more than anyone had ever analyzed. The paper records, which still took nearly a month to get, cost about \$600.

Keying the call information into an electronic format was a monumental task. A group of dedicated interns helped input the time each call came in and the time it ended into a Microsoft Excel

We then wrote formulas that calculated the answer time.

spreadsheet. We then wrote formulas that calculated the answer time for every call, and the daily totals. We broke down the calls into 10-second increments. We included a column for the abandoned calls and the times it took before the callers hung up.

The office’s performance was much lower than the state standard. Only 65 percent of the calls were answered within 10 seconds. Hundreds of other calls took 20, 30, 40 seconds – even more than a minute. When we finally interviewed Lauria on camera, he not only acknowledged his department could not meet the standard, he said it doesn’t do what we did – track call answer times.

While we were gathering the numbers, Andreu concentrated on persuading the parents of the boy who drowned to go on camera. Still in shock, they finally agreed with the hope that the attention would ensure this wouldn’t happen again. We also tracked down the nanny who made the call as well as an elderly woman who hung up after her call to 911 rang for more than one minute. We ended up with two compelling

examples on camera as well as all the strong data we had assembled.

Reaction was swift. Shortly after the story aired in November, two cities agreed to study whether they would be better off handling 911 calls on their own.

In late January, Broward Sheriff Ken Jenne credited our investigation with major changes in how 911 calls are handled. The sheriff’s office installed software that allows it to track call answer times on a daily basis. Keeping track of calls and better training have already improved answer times. In addition, the sheriff told us he would ask for additional staff to handle non-emergency calls. He hopes to set up a system in which those calls are immediately routed to special operators to free up emergency lines. Finally, the sheriff insisted his agency must work toward reaching the state standard of 90 percent.

Contact Scott Zamost by e-mail at Scott.Zamost@nbcuni.com

readme.txt

The IRE Resource Center offers reprints of stories for journalists who are interested in learning more about covering emergency response using computer-assisted reporting. Some examples:

Robert Davis of *USA Today* analyzed data collected from a survey he designed for a series about EMS response and heart attack survival. (Story No. 21418)

Brad Branan of the *Tucson Citizen* used databases to find shortcomings with the city’s fire department response. (Story No. 21752)

Contact the Resource Center at 573-882-3364 or rescntr@ire.org to order any of these reprints.

MAPPING IT OUT

The latest uses of mapping in news reporting.

The dividing line

By Matt Williams
The (Greensboro, N.C.) News & Record

Not long after moving to Greensboro more than a year ago, I realized that to buy groceries or clothes or see a movie, I always drove in the same direction: west. This, even though I lived right in the middle of downtown.

Decades after the end of legal segregation, thousands of residents still did most of their shopping on the "white" side of town.

It's the same problem I'd hear often from minority residents of the east side of town: There just aren't as many places to shop in our part of town as there are in yours. The places we used to shop steadily closed and have been abandoned, while all the new stores open up on your side of town.

But community leaders were never really able to quantify the sense shared by many that there was an inequity. Faced with this, I started to piece together the News & Record's attempt to do that.

Putting together the numbers was the easy part. Finding out the "why" was the challenge.

I began by creating the base for my comparison, in this case block-level

data from the 2000 Census. Using the smallest geographic level of Census data allowed us to see with ESRI ArcView 8.3 geographic information system (GIS) the patterns of racial distribution. We got the demographic and map data through the Census Bureau's Maps and Cartographic Resources Web site at <http://tiger.census.gov>.

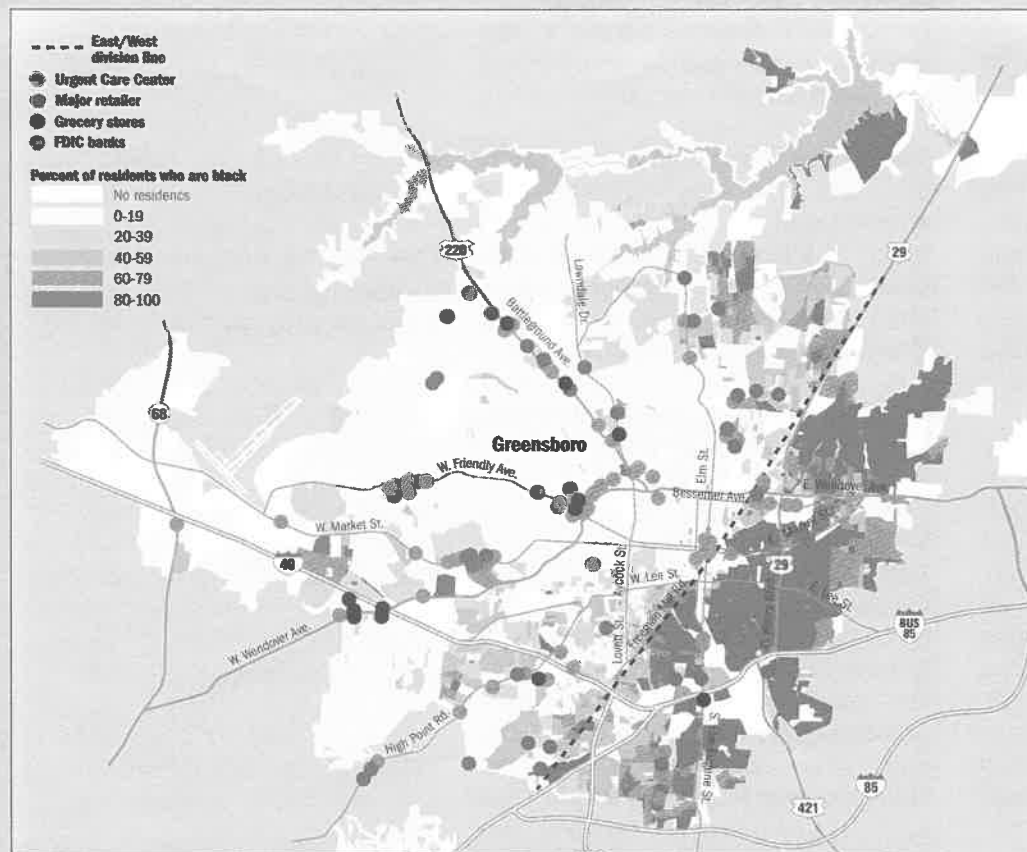
In ArcView, I joined the demographic data file, which includes population by race, to the Census blocks map. I then set the symbology properties to show the percentage of whites living in each block.

Right away, the racial segregation was clear to the eye, with two major streets in the Census TIGER street file running diagonally through downtown delineating the two sides of Greensboro. We selected those roads as our "dividing line."

The western half of the city was home to a little less than two-thirds

of the city's population and was 80 percent white. The eastern half was home to a little more than one-third of the total population and was 77 percent black – a virtual mirror image. Within those two there were large areas where the dominant ethnic group made up more than 95 percent of the residents.

Before we even started, we had numbers that would surprise most of our readers. Like me, they might have assumed that the city was somewhat segregated but had no way of quantifying how much it is segregated. From there, I gathered names and addresses to place businesses of all types throughout the city. Some of the data was publicly available, such as the location of bank branches



insured by the Federal Deposit Insurance Corporation. Those files can be found at <http://www2.fdic.gov/idasp/main.asp>. The rest were simply copied into Microsoft Excel spreadsheets one-by-one out of on-line phone directories. Those were then exported to comma-delimited text fields that were imported to ArcView.

ArcView's geocoding function takes a table containing street address data and turns it into a point map using a street file, such as a Census TIGER street file, as a reference. We displayed each type of business with a different color, and we could display only banks or just grocery stores.

After those businesses were geocoded, we created a simple tally of the number of locations on either side of the dividing line. Once we took population into consideration, it was easy to see the disparities. We obtained the population of each side of the city by selecting the Census blocks for each and using ArcView's summarize function to calculate a total.

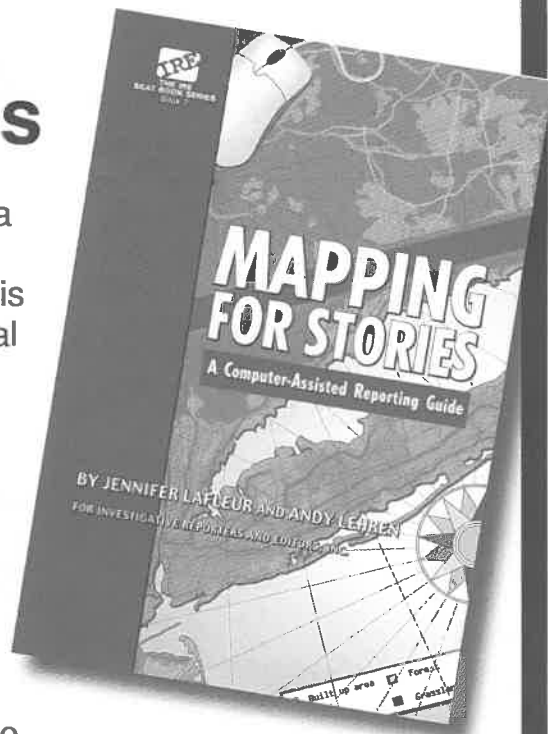
Per capita, there were half as many banks on the "black" side of town and a third fewer grocery stores. In other categories there was no contest: all of the 60 movie screens, eight urgent care centers and all but one major retailer were on the "white" side of town. If the movie screens were distributed evenly based on the population, there should be 40 on the west and 20 on the east.

Fast-food restaurants and dry cleaners, which were also included in our analysis, showed roughly equal distributions. Pawn shops, check cashers and payday loans were skewed the opposite way. They were twice as common in minority neighborhoods as white ones.

continued on page 22

Mapping for Stories

Learn how to map data for daily news stories and larger projects. This book gives professional journalists, as well as journalism educators and students, the foundations for using mapping software to dig deeper into stories. It includes story examples in every section to help show how these skills translate into daily journalism. Downloadable maps and data allow reporters to work along with each chapter and follow lessons at their own pace. It is structured to help journalists complete better stories.



IRE members: \$15 each
Nonmembers: \$25 each

**Plus media-rate postage: \$4 for the first book;
\$2 for each additional book.**

E-mail questions to rescntr@ire.org

ORDER NOW!

By MAIL

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

By WEB

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

By PHONE

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

SPOTLIGHT: WATER AND WATERWAYS

Tracking hurricanes' path to the coast

By Matthew Waite, *St. Petersburg Times*

When a hurricane barrels over water toward the coast, everyone wants to know one thing: Where is it heading? Everyone will talk about it, your family, your neighbors, your editors, coworkers, shoppers in line at the grocery store. Where will that storm hit?

Every six hours, the National Hurricane Center uses its computer models to predict a hurricane's path and publishes the latitude and longitude coordinates for the projected track. That's good news for journalists who use geographic information system (GIS) programs to map data.

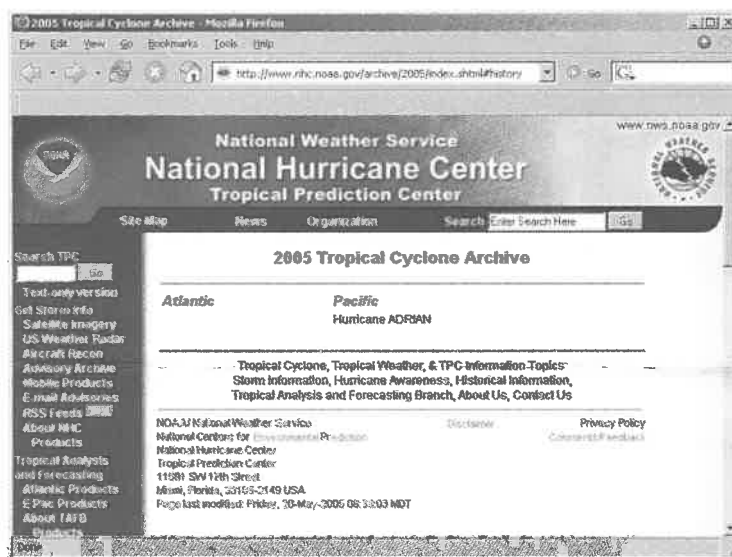
If you've used a GIS before, you've probably plotted x/y coordinates, usually latitude and longitude. It's not difficult to do, but doing it repetitively, every six hours, would be tedious. Fortunately, with ESRI's ArcView 8 to 9, you can automate the task with little effort and create updated maps for reporters and editors.

The place to start is the National Hurricane Center's Web site at www.nhc.noaa.gov. The storms are listed by name, with links to the Public Advisory, Forecast/Advisory, Discussion, Probabilities, Maps and Charts and Archives. If you've never been to the site, each link is worth reading. For mapping purposes, go to Discussion, where the forecast points are kept, usually toward the end.

Building the foundation

Copy the forecast points and paste them into a new Microsoft Excel work-

sheet. You'll have to use to Data/Text to Columns . . . function in the menu to place the data into individual columns. Name the columns, using lat for latitude and long for longitude so ArcView can use these later.



Now we will get the latitude and longitude in a format that ArcView can recognize. Use the search and replace tool in Excel to delete all the Ns in the lat column and then do the same to delete the Ws in the lon column. The lon needs to be represented as a negative number, so create a new column and multiply the cells in the lon column by negative 1. (Longitudes in the Western hemisphere are negative numbers). Copy the contents of the new column and choose Paste special/Values to put the values in the original lon column.

With properly formatted longitude and latitude points, you could go ahead and map these points. But let's add some more data.

Insert a column at the beginning called forecast_age. Here we're going to add data that will tell our GIS which points to display. For your newest forecast, fill in the cells in forecast_age with something that says this is what I want – I use "current" and then call the older points "old." The use for this will come later.

Next to that column, add another and call it forecast. Here, manually fill in which forecast it comes from (i.e. Friday 5 a.m. or Saturday 11 p.m.). That will help you if you want to look at the evolution of the forecast track.

Then add a column called forecast_step. Fill that in, starting with 1 and ending with 8. If you want to connect the dots later using automated means, this step will save you time in having to edit the file.

Last, and this is optional, convert the forecast wind speeds from knots in the last column to miles per hour. To do that, use search and replace to delete the KT's and multiply knots by 1.15077945 in the next column.

With each new forecast, just open the spreadsheet, paste the new information in, copy the formulas down and save it. I usually use storm names for my spreadsheet names.

Tracking maps

Now it gets much easier.

Open Microsoft Access. Create a new Access database and import your spreadsheet. Close the database.

Open ArcMap. For hurricane tracking base maps, I've found that Census TIGER maps work just as well as anything. I have statewide county boundary maps, statewide place maps, statewide major roads (interstates and freeways) and statewide water polygons in a map.

With those layers up, go to Tools/Add XY Data in the menu. Under "Choose a table from the map or browse for another table," browse to your Access database file and find your table inside it. (The newest generations of ArcView, from versions 8.x on, directly read Access data). Make sure that ArcMap recognized your latitude and longitude fields in the box that opens. Click OK and hurricane track points should appear on the map.

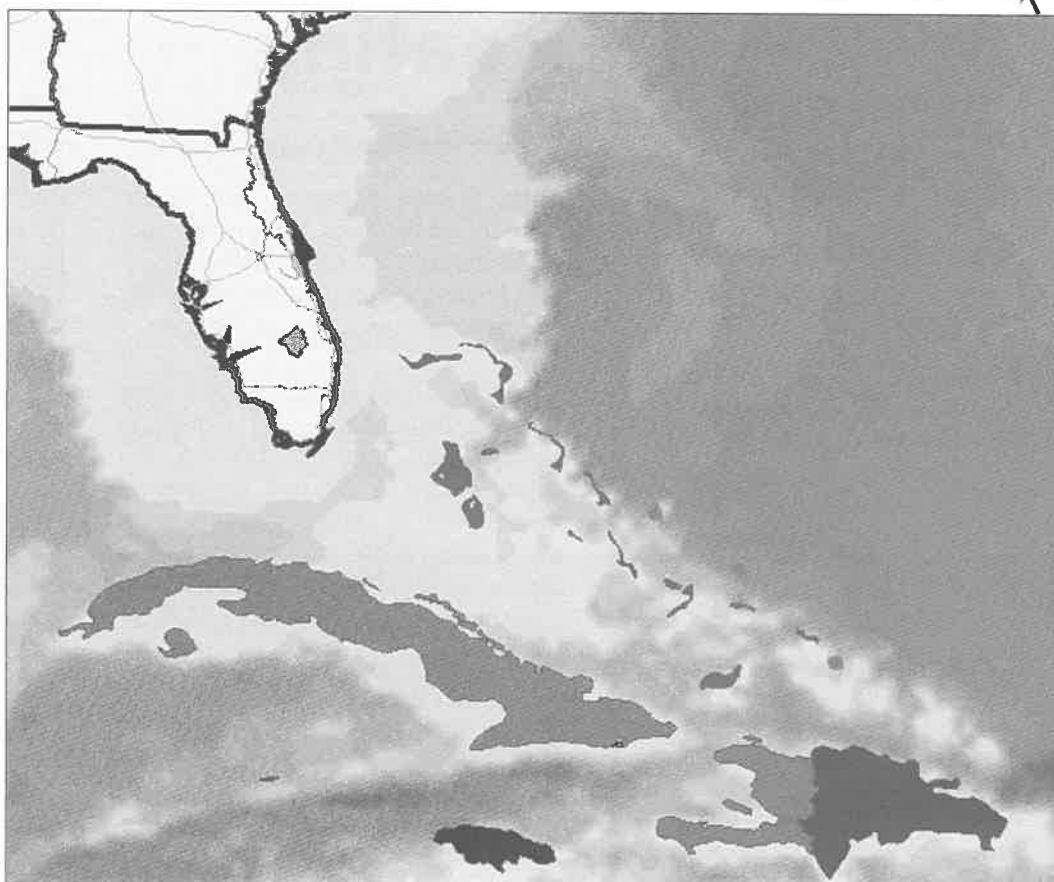
If this is the first time you're mapping points, you'll only have one track on the map. If this is the fourth or fifth forecast you're mapping, you'll see more sets of points on the map. We can show just the current forecast track using the forecast_age field.

Double-click on your hurricane layer name in the ArcMap table of contents. Then select the Definition Query table and enter forecast_age = "Current" to limit the displayed data to the current points.

Repeat as needed

When you get a new forecast, all you have to do is update your database. Copy the points into Excel, repeat the steps and be sure to change the forecast_age column to reflect the new forecast as the current one, then import it into Access and ArcMap will update the points without you having to do a thing.

If you do this consistently – putting each forecast into your database – by the end of the storm you can query all of the forecast positions that start with Initial. Those are the exact point of the storm every six hours. With those points, you have the general path of the storm. One weakness is that hurricanes don't move in straight lines, which is what your map will show. They wobble, angle and turn slowly. So remember that the path you can create from initial forecast points is relative.



If you'd like to turn your points into a line, there are few options available. The best is an extension called XTools Pro (See www.xtoolspro.com). XTools costs \$129 for a single-use license, but is worth the money because of the number of tools in it. One of the tools converts points to a polyline file. With three clicks, you can turn your hurricane points into a hurricane path.

Since Hurricane Jeanne, I mapped the predicted track of the storms that followed, and created maps that helped editors decide where to send reporters.

Mapping forecast points won't result in a story itself, but it will help guide dozens of them. And it only takes a few minutes to pull off, after you get used to the steps.

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

readme.txt

Additional resources for journalists looking for hurricane and tropical storm information are available on the Web.

The University of Central Florida offers an interactive mapping application at <http://hurricane.methaz.org> that allows users to plot the results of different hurricane forecast models on one map. This allows journalists to get an idea about the uncertainty of the storm track.

The tropical weather tracking page from the U.S. Navy (www.nlmoc.navy.mil/cgi-bin/main.pl?tropical) contains downloadable shapefiles of forecast tracks that can be used in a GIS program and basic wind models.

Boating

continued from page 11

to go with the national data.

Using the Coast Guard data, I ran queries summarizing data by year, looking at how boating accident trends in California, Arizona and Nevada were different from the national trends, comparing various lakes and rivers, looking at sizes of boats involved in accident.

While the number of boating accidents, fatalities and injuries nationally had declined steadily, I found a sharp two-year spike in the hit-and-run accidents and a growing number of alcohol-related accidents and fatalities.

The increase in hit-and-run accidents turned out to be a product of dirty data. After talking to boating law administrators in about a dozen states with the sharp increases, I learned there had been data entry errors when the U.S. Coast Guard switched to a Web-based reporting form.

So booze and boating became the focus of my reporting.

I spent the July 4, 2004, holiday weekend on Lake Havasu and Parker Strip. The lake and the strip, although connected by water, seemed like two separate worlds. The lake was open and calm, with a few large boats and partiers congregating in coves. The strip was more crowded, its waves choppy.

As I wrote my first draft, the story became more about the river and the people who gathered there for its party atmosphere. Soon, a phone call from a relative led me to the accident that became the center of the story.

The story turned into a tale of two Orange County men who went to the river for a July 4 weekend, where their paths crossed. Both had alcohol in their blood. One of them died in an accident. Another one was arrested and

later convicted of negligent homicide. The numbers and trends I got from the data became the frame that held the story together.

I returned to the river with my Garmin eTrex Legend Global Positioning System unit and logged the locations of accidents, the sheriff station, and the hospital. I drove to key locations checking times and distances. I also started noting and logging drinking establishments along the river.

Later, in the office, using the records obtained in my reporting, I created a timeline of the accident. I assembled a spreadsheet tracking witnesses mentioned in police reports and what they said they saw.

I used California and Arizona online liquor license databases to make sure I found drinking establishments along the Parker Strip. Both databases are accessible on the agencies' Web sites. I searched data for cities located on both sides of the Strip and pasted results into Excel.

Then, I tried to map the locations using ESRI's MapShop, a Web-based mapping system used to create map graphics. But the geocoding failed on most of the rural addresses.

So, I gave the spreadsheet of liquor li-

censes and my GPS unit to our graphics reporter, Chantal Lamers. She then spent a day driving along the Parker Strip looking for licensed bars, liquor stores and alcohol-selling boat launches and logging their locations with the GPS.

I converted the coordinates we had logged with the GPS unit into the format that MapShop accepted and uploaded the file into MapShop. MapShop uses degrees and decimal minutes and seconds. The GPS was originally set to record coordinates in degrees, minutes and decimal seconds. I re-calculated the coordinates in Excel.

All 57 drinking spots showed up on the map. Another 20 points showed other critical spots — locations of major accidents, hospital and rescue station. Our graphics department used the file to create maps for publication.

I used the aerial photography available through MapShop's Premium Imagery layer to double-check locations and make my descriptions more precise.

You can read the story, "A Dangerous Mixture" online at www.ocregister.com/ocr/2004/10/03/sections/news/news/article_260375.php.

Contact Natalya Shulyakovskaya by e-mail at natalysas@ocregister.com.

EXTRA! EXTRA!

YOUR GUIDE TO THE LATEST INVESTIGATIVE WORK

If your newsroom is parched for story ideas, Extra! Extra! on the IRE Web site (www.ire.org/extraextra), carries links to and summaries of articles about water and waterways that used computer-assisted reporting. Here are some examples:

The Salt Lake Tribune analyzed public utilities data to identify which homes used the most water in the area. One estate consumed 7.1 mil-

lion gallons in one year — more than 21 times the use of an average homeowner. Former NBA star Karl Malone's home used more than 3.3 million gallons, the newspaper found.

The *St. Petersburg Times* conducted an analysis showing that, since 1998, the median price of waterfront homes in Pinellas County, Fla., was increasing more than twice as fast as real estate off the water. The newspaper reported that the typical waterfront property cost more than \$500,000 in 2004 — more than double what it cost six years earlier.

Bridges

continued from page 3

"Scour critical" became an important benchmark in my story. I brushed up on the language of engineers by reading Web sites such as the American Association of State Highway and Transportation Officials at www.transportation.org/aashto/home.nsf/FrontPage.

Then I hit a snag: a discrepancy in my numbers. I was down to my final cross-checking when I decided to match the numbers I had from my analysis with the numbers on the Federal Highway Administration's Web site. Mine showed nearly one-third more bridges in California than the Web site. I felt that same cold rush of panic I had experienced seeing the Ventura County fire chief on the news.

First, I called a Federal Highway Administration public relations officer. He said the numbers on the Web site were correct. He didn't know how I could have come up with more bridges. I called Jeff Porter, the Database Library director for IRE and NICAR. He said it hadn't been a problem before, and suggested looking at several things, including whether I had double-counted bridges. He said that some bridges are counted twice because the administration considers the bridge underneath, or underbridge, as a separate bridge. I didn't think that was the problem because each of the bridges I counted had a unique ID number. It didn't make sense to me that any of them would be underbridges.

So, I searched further. I e-mailed a computer analyst at the state transportation department. He didn't respond. I called the number on the fact sheet for the Federal Highway Administration. A woman told me she couldn't talk to reporters.

The world around me started to get black. So I begged. I got teary. I explained I wouldn't quote her, but please don't let me put bad informa-

tion into the newspaper. She quickly told me several items to check. One was underbridges — just as Porter had said. When I explained I had already tried to check that by making sure I included only bridges with unique numbers in my count, she explained that the underbridges also had their own ID numbers. Five minutes later, the problem was solved. (If I had researched Porter's suggestion further, I would have saved myself an hour of panic.)

The story, "County's aging bridges at the breaking point," ran a month after I graduated from boot camp. Sidebars, graphics and a Web video presentation I had shot accompanied it. It can be seen at www.VenturaCountyStar.com.

The county official who hadn't wanted to give me the records was impressed with the report. He asked me if I expected to win a Pulitzer Prize with the story. Yeah, right after I hit the lottery. Nevertheless, I did learn some lessons:

1. Practice is key. To do computer-assisted reporting, I found I have to rewire my brain. The key to that is practice. I try to do a little bit each week.

2. Read first. Analyze later. On the bridge data from the Database Library, there is a readme file and a guide to the fields in the data. Though I thought I knew what I was looking for, I found there was much more to know and that the data had much more to give. I realize I would have saved a lot of time if I had read all the materials that came with the data before diving into the numbers. In my hurry, I did it the other way around and wasted time.

3. Log queries. Keep records. When I started queries on the data, it seemed so simple to do and then so tedious to write. I realized later I should have heeded the advice of the boot camp instructors and written down everything. Then, when I needed to re-create the queries with different modifications, I could have done it easily. Logging

queries also seems to give editors a much better sense of security. In addition, I learned it can be very difficult to re-create queries two weeks later under deadline pressure with a stressed-out editor looking over your shoulder. From now on, I'm keeping a notebook by the computer to log the queries.

4. Check your results against other sources. If the numbers are good, they will stand up. If I hadn't checked the California Department of Transportation Web site, and then the Federal Highway Administration Web site, I never would have known how far off I was. I would have risked a massive correction. I would have had fewer chances to do computer-assisted reporting in the future.

5. Give yourself time. The exercises learned in boot camp can be much easier to do than grabbing data in real time. I found it took me much longer to do CAR on my own. And getting the numbers is just the starting point.

Contact Dani Dodge by e-mail at Dani@DaniDodge.com.

readme.txt

The IRE Resource Center offers tipsheets and stories for journalists who are interested in using computer-assisted reporting to examine bridge safety.

Mark Greenblatt of KHOU-Houston provides guidance on using the National Bridge Inventory database. (Tipsheet No. 2399)

A *Boston Globe* investigation using the bridge data found that one of every seven bridges in Massachusetts was structurally deficient and in need of an overhaul. (Story No. 17011)

Contact the Resource Center at 573-882-3364 or rescntr@ire.org to order these.

Tech tip...

Grappling with data on EPA mainframe

By Jeff Claassen, Fort Worth Star-Telegram

We were struggling to verify a tip: the U.S. Environmental Protection Agency had dramatically slowed its enforcement air pollution laws at oil refineries.

The clues would come from the agency's administrative records – the ones listing the dates and results of inspections, enforcement actions, court settlements and administrative fines.

We asked the Dallas regional EPA office for the records, but the requests were taking months to fulfill, and the information they eventually provided was woefully inadequate.

Those EPA record-keepers were trying to pull information from the agency's enforcement-tracking databases. That piece of information gave us our first break – we needed to get copies of those databases to work with ourselves.

One database, the AFS (Air Facility Subsystem), covered enforcement of the Clean Air Act. Another, the PCS (Permit Compliance System), stored enforcement records for the Clean Water Act. A third, ICIS (Integrated Compliance Information System), held all the court and administrative settlement records.

The agency has good descriptions of these databases and others at www.epa.gov/compliance/planning/data.

The EPA's enforcement managers in Washington, D.C., were willing to help. But they thought it would be impossible to use the massive databases. Those databases contain terabytes, or thousands of gigabytes, of data.

Included are the enforcement history and all of the pollution-output records for every emissions point in every factory, refinery and other facility in the country that has a pollution permit.

Next, we went to the people who know that material the best – the agency's programmers, who run search services on the EPA's Web site. Those search engines include ECHO, an enforcement history search, which environmental reporters use often. (See www.epa.gov/echo/compliance_report.html)

The programmers told us the agency has much smaller versions of the databases we wanted, updated every month for public use. Those databases included administrative actions, such as inspections, but not the massive list of individual emissions reports.

Graduate students, environmental activists, and the people who run the EPA Web site searches used that trimmed-down data the most. The agency's Web-site searches pull from these smaller databases.

The only hitch: public access was through an old, awkward Web page connection, at <https://trex.rtpnc.epa.gov>. The core of this Web site is a terminal emulator that wires you into a supercomputer at the agency's computer center in North Carolina. Our company firewall blocked it, so all the programming and downloading had to happen from home.

It took weeks to learn the system and download the data. The system, called IDEA (Integrated Data

for Enforcement Analysis), uses a programming language created entirely by EPA computer technicians. Everything you want to do requires terminology you won't see anywhere else.

The EPA gave us several technical guidance documents, including one that is more than 200 pages long. Programmers also gave us some sample programs, and worked diligently with us to get the syntax right so it would create the data sets that we wanted.

To show how idiosyncratic the syntax is, here is one program we used. This script pulls specific fields from the AFS dataset, ("FS.AFSID," etc.), copies the unique identifier field to each record ("DEC 100") and designates enough server processing time and hard drive space for our new file:

```
INPUT
NONLINKED
SELECT AFS.AFSID NE .
OUTPUT
JOB TIME 40
REPORT AFS1
DATASET AFSNX
/*SHR
SPACE
LENGTH 190
PRIMARY 2350
SECONDARY 1175
RELEASE
/* DIR
TITLE "AFS PULL 1"
DETAIL
AFS.AFSID DEC 100
AFS.CNTY
AFS.STAB
AFS.PLNT
AFS.APC1
AFS.REGN
AFS.AQCA
```

```

AFS.PNME
AFS.STRT
AFS.CYNM
AFS.SIC1
AFS.SIC2
AFS.INSPDAY
AFS.FEDREP
AFS.ZIPC
AFS.DTA1
AFS.ACSEQ
AFS.ANT1
AFS.PAM1
AFS.INFE
AFS.YIFE
AFS.INFO
AFS.YIFO
AFS.RDTA1
AFS.RANT1
AFS.RATP1
AFS.RPAM1
AFS.DDTA1
AFS.DANT1
AFS.DATP1
AFS.DPAM1
PARM BACKYEARS_16
PARM MINHEAD
PARM DELIMIT_^

```

There's one note of caution hidden in the mostly unreadable program. If you download data from IDEA, use the shift-6 (^) as your delimiter, as specified in the last line of code. Fields contain commas, making those unusable and IDEA is unable to use tabs or other field delimiters other than alphabetic characters.

Downloading also took a while because we needed to sign onto IDEA in the evenings and weekends, because of costs. During weekday hours, the EPA charges \$2,100 per CPU hour (hour of computer processing time. For example, time to run queries and save data sets). Night hours cost half that, and weekend hours are one-quarter of the regular cost.

Once IDEA created the data sets for us, downloading a single file took hours, even though I was using a DSL connection at home. The EPA's FTP server in North Carolina is slow. Downloading was complicated, but worth it. We finally had three very

comprehensive tables containing:

- more than 2 million records for 175,000 air-permit facilities
- more than 1.3 million records for 330,000 water-permit facilities
- about 230,000 entries for 59,000 court and administrative settlement cases

The EPA supercomputers store the data in old-fashioned hierarchical format. That means that different types of information for a facility (address and location, inspections, violation notices etc.) are contained on different lines of data, one after the other, with no common ID number to link them together.

Fortunately, the EPA programmers had come up with a way to generate repeat ID numbers in the output data, so they were attached to each record. We were able to thread this information back together using the common ID. I used FoxPro, more out of habit than for any other reason. Access would have handled these files as well.

Another note for anyone who uses this data: to count inspections, enforcement actions, and other agency activities, you must use what EPA record-keepers call "rolled-up" data. For reasons the agency could not clearly explain, records are often duplicated in its "detail" data. To check that, we downloaded each type of record – "rolled up" and "detail" – and did see the repetition in the detail records.

The data went back to the 1970s and created a broad picture of enforcement patterns over the decades. The number of inspections, violation notices and other actions spiked upward under the first President Bush, after flattening

under Reagan, then crested during the Clinton years, and finally dropped sharply after 2001. The data showed that our tip was right.

Since 2001, inspections have gone down 52 percent for oil refineries and 4 percent for all industries, including refineries. Violation notices decreased 68 percent for refineries, and 24 percent for all industries. The number of court and administrative settlements dropped by 77 percent for refineries and 7 percent for all industries. Stack tests, examinations done by private contractors and turned over to the EPA, were down 9 percent for refineries and up 43 percent for all industries.

We also used other data – the EPA's Toxics Release Inventory and Texas' Point Source Air Emissions Inventory, among others – to flesh out our description of the pollutants that refineries have released to the air, water and land.

Our reporting found that the EPA had in recent years turned increasingly to legal agreements called consent decrees, and was not inspecting refineries during the negotiations or even after them. Experts said that the lack of comprehensive inspections had a cascading effect, driving down the number of violation notices, administrative penalties and other actions.

The two-day series got responses from around the country, including neighboring Louisiana, which also has a large number of oil refineries.

Contact Jeff Claassen by e-mail at claassen@star-telegram.com.

CAR TOOL Cleaning your files With TextPipe

By Ron Nixon, *Star Tribune (Minneapolis)*

Here are three challenges familiar to journalists who have worked with dirty data:

1. Your data exists in a text file, but when you try to import the file into your program, there's a problem: the data lines up incorrectly. There's a tab or comma delimiter that's out of place or a space that you can't see.
2. You need to search and replace hundreds of times in your data, sometimes using exact searches and others with pattern searches.
3. You need to extract information out of a text file that's unstructured or pull data out of a HTML or XML file.

For any of these challenges you could use a programming language such as Perl, Python, PHP, or even Visual Basic to craft a solution. There are several advantages of using any of these. Among them is the power and, with the exception of Visual Basic, the availability of free downloads on the Web. But there's a downside: each comes with a steep learning curve.

Even Python, which can be easy to use, has a learning curve. An alternative tool is a utility called TextPipe Pro. TextPipe costs \$299 and you can download a free 30-day trial from www.crystalsoftware.com.au/index.html.

Made by DataMystic, an Australian company formerly known as Crystal Software, the program has an easy-to-use interface. Users can have it up and running in a few minutes and it is especially well-suited for the non-programmer.

Don't let the ease of use fool you — it's

still a powerful text manipulation tool. TextPipe comes with 100 filters that can be applied to everything from a simple search and replace to extracting telephone numbers from a text file. What really makes it shine, though, is its ability to do hundreds, even thousands of search-and-replace routines at the same time. And that's just the beginning.

Along with the searching and replacing, you can add filters to remove white space, add a new line or add a line feed, all in one session. You can even do targeted searching and replacing by column or even by row. You can even use a Microsoft Excel spreadsheet or a text file as a lookup table for TextPipe's searching and replacing. Simply type the strings to find in one column and the replacement in another and import the data.

What's more, TextPipe is fast and can handle files of several hundred megabytes. I've used it on files 2 gigabytes or more. TextPipe can be set up to run as a scheduled program, so you can have it running and working while you do other things. You can also run it from the command line.

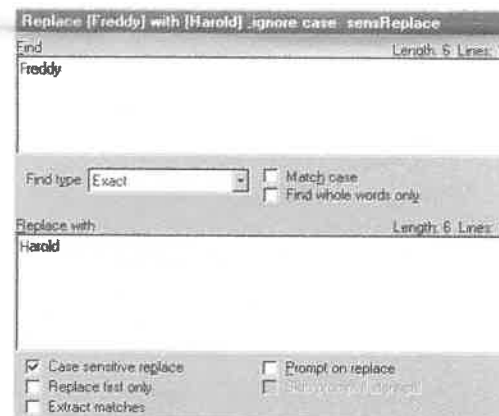
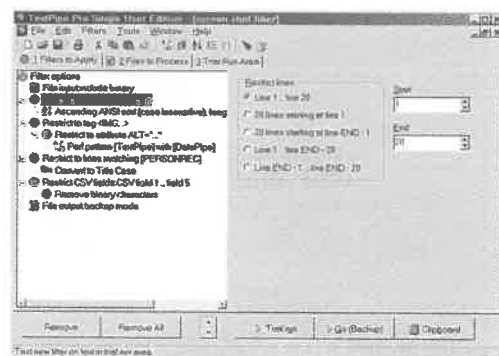
Other TextPipe text file capabilities include merging dozens of files, sorting information and converting text files from mainframe, PC and Unicode data formats.

TextPipe keeps an audit log of all your changes so you can document your work. The software has a feature called the Scratch Pad where you can try your filter on practice text before using it on your real text file.

Those who need to do even more complex text processing can use Perl-style regular expressions. For those

who are intimidated by regular expressions, TextPipe offers EasyPatterns, a text processing tool that has all the power of regular expressions but is easy to understand.

For example, to describe a phone number in regular expressions you'd write: `"\d\d\d\d-\d\d\d\d-\d\d\d\d"`. In EasyPatterns it's: `[PhoneNumber]`.



DataMystic provides an introduction to Easy Patterns at www.crystalsoftware.com.au/easypatterns.html.

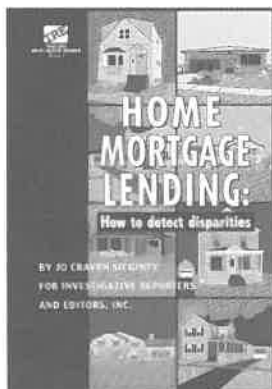
Even though there are programming languages available for free download, TextPipe's ease of use and flexibility make it well worth the price.

Contact Ron Nixon by e-mail at rnixon@startribune.com.

REQUIRED READING FOR YOUR NEWSROOM

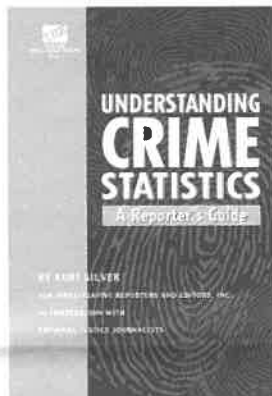
HOME MORTGAGE LENDING: *How to Detect Disparities*

Pulitzer Prize-winning journalist Jo Craven McGinty guides reporters through understanding and using Home Mortgage Disclosure Act data. Included are specific story ideas and lists of tipsheets and stories available through IRE.



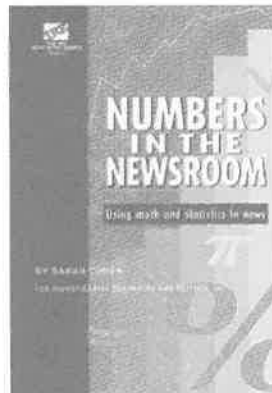
UNDERSTANDING CRIME STATISTICS: *A Reporter's Guide*

Covers using Uniform Crime Reports, National Crime Victimization Survey, National Incident-Based Reporting System, other major statistical sources, writing the crime statistics story and database analysis of crime statistics. Includes law enforcement contact information and stories and tipsheets available from IRE.



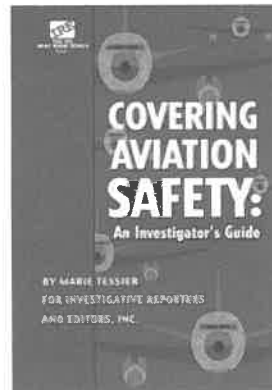
NUMBERS IN THE NEWSROOM: *Using Math and Statistics in News*

Pulitzer Prize-winning reporter Sarah Cohen guides journalists through working with numbers, including fractions, rates, percents, per capita, measuring change, making inflation adjustments, understanding averages, working with graphics, doing budget stories, questioning surveys and polls, and much more.

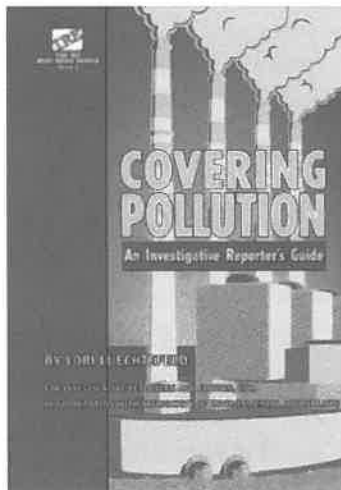


COVERING AVIATION SAFETY: *An Investigator's Guide*

Learn to develop a crash plan for your newsroom, report from the scene of a crash, start an aviation beat, interpret aviation records, negotiate Web data and investigate planes and airlines on deadline. Includes related stories and tipsheets available from IRE, as well as FAA regional contact information and useful Web sites.



COVERING POLLUTION:

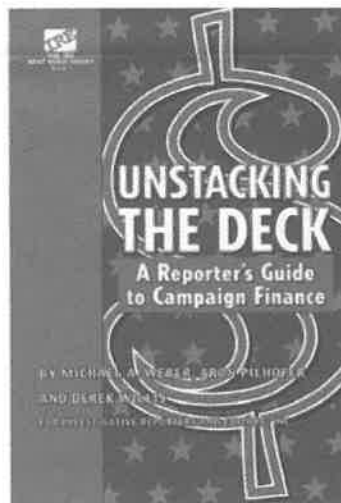


An Investigative Reporter's Guide

by IRE in cooperation with SEJ

This primer gives an overview of useful resources reporters can use for local investigations into environmental pollution. Its main focus is to show how to get to the heart of an investigation quickly and without waiting months for FOI requests to be fulfilled. Filled with examples and references to stories, tipsheets and other resources available from IRE and SEJ.

UNSTACKING THE DECK:



A Reporter's Guide To Campaign Finance

Invaluable for pursuing stories about the impact of money on elections, political parties and candidates at the federal, state and local levels.

- Understand the loopholes in soft money restrictions.
- Learn about the use of nonprofits to funnel money to candidates.
- Find out how to track where candidates spend the money they raise.
- Learn how to obtain and use pertinent documents and electronic data.
- Find story examples from the IRE Resource Center.
- ... and much more

ORDER NOW!

IRE MEMBERS: \$15 • NONMEMBERS: \$25 EACH • PLUS POSTAGE

POSTAGE RATES: Media Rate - \$4 for first book, \$2 for each additional book

BY MAIL

Send your check to IRE
138 Neff Annex
Missouri School of Journalism
Columbia, Mo. 65211

BY PHONE

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

BY WEB

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

Outsourcing

continued from page 1

ing consulting contracts for transportation work leaped 45 percent, to \$119.1 million from \$82.1 million, over the past five years – several times the rate for all state contracts.

Next, he found that employees of four of the top five recipients of engineering contracts were also some of the largest campaign contributors to state politicians. To compare contributors and contractors, Marley cross-checked state records compiled by a Wisconsin nonpartisan group advocating campaign finance reform, Wisconsin Democracy Campaign, against the contracts data.

To find change in government contracts, Marley did a simple percent-change calculation in Excel from the contract totals given to him by the state.

The relationship between the top contributors and contractors was a bit more difficult to crack. Marley began by summing amounts for the contractors. He then plugged in the top contractor names into the nonprofit's Web site, which has a searchable list of donors, to get a feel for how much the top contractors were giving to state officials.

Overall, Marley said the data and the analysis were fairly straightforward. He said obtaining the data from the state also was fairly easy. It took him just two days to receive the Excel file containing the transportation contracts from the DOT with several hundred rows of contract data.

View the article online at www.jsonline.com/news/state/oct04/269824.asp.

War profits

Reporters Joseph Neff and Jay Price of *The (Raleigh, N.C.) News & Observer* brought the Iraq war close to home by investigating a North Carolina company doing contract work in Iraq. At the

same time, their article, "Contractors in Iraq make costs balloon," exposed the private and complicated world of government contracting and subcontracting in Iraq.

The mutilation of several American contractors in Iraq propelled the investigation. Several of the contractors worked for Blackwater USA, a North Carolina company that guards the U.S. ambassador and provides other security details in Iraq.

The authors tracked a string of con-

The relationship between the top contributors and contractors was a bit more difficult to crack.

tracts starting with Jerry Zovko – one of Blackwater's contractors who was killed in Iraq while guarding convoys. They followed the contracts up from Zovko through three other subcontractors and eventually to Halliburton, the prime contractor.

The business of contracting in Iraq is a closed world, Neff said. Many people declined to talk until the articles started running. "If we were reporting from Kuwait or Iraq, it would have been easier."

Such an environment made the database reporting challenging, Neff said. Although U.S. taxpayers ultimately paid Blackwater's bills, all of the subcontracts were hidden from the public eye either because of the confidentiality of the work or because the company was a subcontractor.

The major part of Neff's analysis used a federal contracts database. That database, which Neff obtained from the

IRE and NICAR Database Library, gave him a picture of how Blackwater's business with the government grew since 1999 – before the recent wars in Afghanistan and Iraq. He analyzed the data using FoxPro relational database manager.

Neff used the contractor name field to select Blackwater contracts and tally how much its business had grown with the government to \$19 million in 2003 from \$205,000 in 2000. He also used the database to obtain contract numbers and make Freedom of Information Act requests for related documents.

However, the database provided only a partial picture of Blackwater's relationship with the government because it did not include any subcontracts the company had received.

The Blackwater contract that Neff and Price were looking into was a subcontract with Halliburton, so they had to seek out other sources. Getting information from Halliburton wasn't feasible. Halliburton had failed to document almost half of a \$4 billion invoice it submitted to the Pentagon, according to a recent audit by the Defense Contract Audit Agency. The missing half included the Blackwater contract.

So the reporters relied on other sources and unraveled four layers of contracts above the original Zovko contract. At each layer, the reporters found tacked-on profit margins hidden from the public. They found Blackwater marked up the Zovko contract by 36 percent. However, the other markups went undiscovered. The reporters said "whether the cost for Zovko doubled, tripled or quadrupled" remains unclear as Halliburton has refused to disclose the details of the spending.

Read the article online at www.newsobserver.com/nation_world/blackwater/story/1762376p-8044834c.html.

For more information about the Federal Procurement Data System and ordering options, visit the IRE Web site at www.ire.org/datalibrary/

databases/fedcontacts.

Defense dollars

In "Outsourcing the Pentagon," Larry Makinson and a team of reporters and analysts at The Center for Public Integrity completed what is perhaps one of the most thorough CAR analyses of the Department of Defense's private contracts.

The investigative team of about 20 journalists used database reporting to discover the Department of Defense has outsourced about half its budget since 1998. In addition, they found about 40 percent of the department's contracts were awarded without competitive bidding during the same time period.

The basis of their analysis was the Pentagon's procurement database that they described as "public information that had been posted for years on an obscure Defense Department Web site."

Aron Pilhofer, database editor for the Center (and former Campaign Finance Information Center director for IRE), said the Center used the procurement, Federal Election Commission campaign contribution and a lobbying databases for comparison.

The Center downloaded five years of contracting data in text format from the Defense Department Web site (www.dior.whs.mil/PEIDHOME/guide/pro-coper.htm), which houses data on contracts issues since 1966.

The Center then imported the files from the department's Web site into a Microsoft SQL Server database manager. The Center focused on the contractors with \$100 million or more in Defense of Department contracts.

However, the information in the database was far from perfect and required substantial cleaning, according to information posed on the Center's Web site.

We discovered very early that the Department of Defense database

was rife with errors – that literally thousands of records had incorrect or outdated IDs for the contractor's ultimate parent," the Center reported.

To mitigate the errors, the Center examined contracts for companies below the \$100 million threshold to make sure that did not miss any company due to possible miscoding.

The Center eventually found more than 700 businesses that had \$100 million contracts or more with the Pentagon.

After identifying the major contractors, the reporters researched campaign contributions to Congress and the president. For that, they analyzed the FEC campaign contribution database and lobbying reports filed with the U.S. Senate's Office of Public Records.

They said that the lobbying data online is searchable, but hard to use, so they scraped the data from the Web site using Perl into their own database. (Web scraping is using programming languages to gather data from Web sites and assemble it into database form).

In the end, they found that many of the top contractors were also generous political contributors, especially to President Bush, who was the top recipient of the contractors' contributions, receiving \$5.4 million from the top contributors since 1998.

Read the Center's stories at www.publicintegrity.org.

Contact Catherine Rentz Pernot by e-mail at Catherine@ire.org.

readme.txt

See "Finding backyard military contractors" in the May-June 2003 *Uplink* for more information about digging into federal procurement data to unearth information about local contractors.

MOVING?



Please send us your new address so you don't miss one issue of *Uplink*.

Drop us a note at
jgreen@ire.org

 Investigative Reporters and Editors, Inc.

REQUIRED READING For Your Newsroom

Unstacking the Deck:

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

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



ORDER TODAY!

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

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

Cheaters

continued from page 1

another, with several cases that pointed to cheating — not by students, but by educators.

In Wilmer Elementary's case, a state investigation confirmed what our reporting found: Teachers didn't work "real hard." They gave students answers to the test.

Wilmer Elementary belongs to the Wilmer-Hutchins school district, where finances, leadership and academics have been in chaos for decades. The district made headlines last summer when it had to delay the opening of a high school because the building had roaches, mold, mildew and a leaking roof.

A host of financial and legal troubles followed. The FBI and Texas Rangers (the law enforcers, not the baseball team) got involved. And then came some good news: high test scores brought an "exemplary" state rating to one elementary school and "recognized" status to the other two, including Wilmer.

My colleague covering Wilmer, Josh Benton, didn't believe it. He had heard rumors of cheating before in the district, but nothing concrete. Pointing to Wilmer Elementary's reading scores, he wondered how we could put the dramatic rise into statistical context. Just how likely was it?

Enter SPSS, the statistical software program. I had grown to love and respect SPSS (and statistics, for that matter) at my previous newspaper, the *St. Louis Post-Dispatch*. We used SPSS to analyze school test scores relative to family income, as several other papers have done.

SPSS and similar programs, such as SAS, also have a wonderful ability to draw pictures of data. You can see stuff with charts and graphs that you

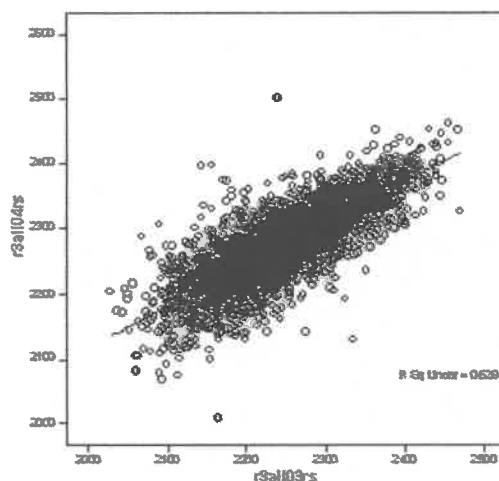


Figure 1

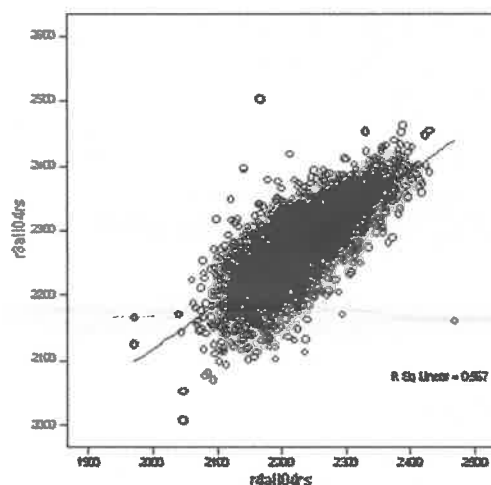


Figure 2

wouldn't in rows and columns. So to see how Wilmer Elementary's third-grade reading scores compared to the rest of the state, I plotted each school's scores for 2003 versus the scores for 2004. (Figure 1)

Each circle represents a school in Texas that tested at least 30 students. (Smaller numbers can skew results.) The scatterplot shows a trend: Schools do about as well in 2004 as they did in 2003. But see the circle hovering high above everyone else? That's Wilmer Elementary. It went from an average score of 2,240 one year to 2,501 the next. No other school had such a jump.

Then we looked at the scores another way: comparing third-grade reading in 2004 to fourth-grade reading in 2004. (Figure 2)

Again, Wilmer stands out at the top, with the stellar third-grade score but just a so-so fourth grade score.

But that wasn't the only school that stood out. Check out the school on the far right, with the near-perfect score for fourth grade but a lousy one for third grade. That's Harrell Budd Elementary in Dallas.

We ended up looking at reading and math tests, which Texas students take in grades three through 11. (Other subjects are tested only in certain grades.) We examined the scores four ways:

- From one grade and year to the next, such as third-grade reading 2003 vs. fourth-grade reading 2004.
- From one grade to the next for the same year, e.g. third-grade reading 2004 vs. fourth-grade reading 2004.
- The same grade over two years, e.g. third-grade reading 2003 vs. third-grade reading 2004.
- The same grade over two subjects, e.g. third-grade reading 2004 vs. third-grade math 2004.

The analysis went beyond scatterplots and used a statistical tool called regression, which looks for the relationship between two variables. Based on that relationship, regression makes educated predictions. So given a school's score last year, regression determines how well it should perform this year.

In Wilmer's case, it scored 231 points higher than expected in 2004, based on its 2003 score. To put that in context, most schools scored within 37 points of their predicted score.