Uplink

November • December 2003
Volume 15, Number 6

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

REMOTE SENSING

Images boost GIS work

By Holly Hacker
IRE and NICAR

Most GIS-savvy journalists are familiar with shapefiles, points and polygons. They can place dots on a map to show crimes in their neighborhood, or code Census tracts to identify neighborhoods that gained or lost the most residents.

But for journalists who use a geographic information system, the next spatial frontier to conquer is remote sensing, according to experts in the business.

The *St. Louis Post-Dispatch* recently discovered the dynamic stories that can be told by combining satellite images with other information in a GIS. The newspaper hired researchers in Minnesota to study development on the Upper Mississippi River Basin since the flood of 1993. The analysis, which blended satellite images, Census figures and flood area data, confirmed the lead reporter's suspicion: Missouri has developed more flood plain than any

continued on page 20

SPOTLIGHT: CITY GOVERNMENT

Data-based approach reaps great stories

By Michael Mansur, *The Kansas City Star*

Two years ago, I thought Excel was a meatpacking plant somewhere in central Kansas. I was spreadsheet ignorant.

This was particularly startling because I had worked on stories with other reporters who had analyzed large amounts of data using Microsoft FoxPro. And years ago, I had even sorted and analyzed data in dBASE. But for a variety of reasons, spreadsheets were foreign to me.

Last year I began a new assignment covering local government. And I almost immediately saw the need to

continued on page 12

SPOTLIGHT:

For more on covering city government see:

- Mapping analysis finds towing hot spots in St. Paul, p. 6.
- Database work finds strip-club money going to San Diego pols, p. 8.
- Spreadsheets help uncover mayor's unlisted contributions, p. 10.
- The November/December 2003 *IRE Journal*

INTRANETS

Newsroom data depots no longer a luxury

By Matt Carroll, *The Boston Globe*

In a few short years, searchable databases on newsroom intranets have gone from novelty to necessity.

Most newsrooms either already have databases – ranging from state payrolls to police bookings, from licenses for dogs to licenses for burglar alarms – or have someone thumbing through old "Dummies" computer books, learning how to use them.

Journalists who use the databases understand their value as a tremendous tool for finding people on deadline. With a few keystrokes they can find the address of the juror who voted to acquit in the murder trial, the unlisted phone of an indicted CEO, and all the campaign contributions from the real estate industry to the new mayor.

Other papers, such as *The New York Times*, have gone the next step and linked databases together so that journalists can intuitively follow the flow of

continued on page 22

Bits & Bytes

Updated files

The IRE and NICAR Database Library now has FDA Adverse Events Reporting System data spanning 1969 through December 2002 and NASA Air Safety Reporting data covering 1988 through September 2003.

The FDA database contains more than 2 million medical problems associated with pharmaceutical products reported by manufacturers, health-care professionals and consumers. The database holds information about each drug's probable role in a patient's death, disability or other such event.

In 2000, for example, the *Los Angeles Times* found information suggesting that the FDA ignored warnings from its own scientists and sided with drug companies chasing profits.

Journalists have used the NASA data to identify aviation incidents and gather information about them for news stories. The database is composed of voluntary and anonymous reports about aviation safety filed by passengers, pilots, and controllers – anyone is eligible to file. The data is rich with detailed narratives about events such as the 1998 near collision between two foreign jumbo jets at JFK International Airport in New York.

To learn more about these and other databases, visit the IRE and NICAR Database Library on the Web at *www.ire.org/datalibrary/databases* or call 573-884-7711.

continued on page 4

INSIDE NICAR

New faces, changes at database library

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

Fall is a season of change – something that holds true at the IRE and NICAR offices. Each new school year at the University of Missouri brings new staff members to our Database Library.

So I thought I'd introduce the new graduate assistants working in the database library. Chances are, if you call to order data or ask us a question, you'll speak to one of them.

Christina Caron is a first-year master's student who has spent the past three years managing and collecting outcomes data at the Dana-Farber Cancer Institute in Boston.

Megan Clarke is also a first-year master's student who has interned with KEVN-Rapid City, S.D., and Wisconsin Public Radio.

Hsuju Ho is a second-year master's student who started working in the Database Library last year as a volunteer. She has interned at major English and Chinese newspapers in Taipei, Taiwan.

Tad Vezner is a second-year master's student who switched to journalism after a career in social work and public relations. He was director of communications for the Epilepsy Foundation of Greater Chicago. Since making the switch to journalism, he has reported for *The* (Toledo) *Blade*.

We also have two new Database Library volunteers: incoming master's students Florence Kizza and Andrea Lorenz.

Second-year master's students Jaimi Dowdell and Holly Hacker have returned to the Database Library.

Two former Database Library staffers have departed. Megan Christensen got her master's degree and is reporting for the *Detroit Free Press*. Stephanie Kang is reporting for *The Wall Street Journal* in Los Angeles.

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

Hands-on CAR training

IRE and NICAR have a number of training opportunities coming up for journalists who want to learn more about computer-assisted reporting.

Journalists can learn how to use geographic information system software to map data and uncover compelling information for news stories at the Mapping Data for News Stories Workshop Jan. 9-11 in Columbia, Mo.

Journalists who want to expand their knowledge of data analysis can register for the Advanced Statistics Workshop Feb. 13-15 in Tempe, Ariz. Pulitzer Prize winners Steve Doig of Arizona State University and Sarah Cohen of *The Washington Post* will lead the session.

IRE and NICAR also will conduct intensive CAR Boot Camps for journalists who want to learn how to use spreadsheets and database managers to produce high-impact news stories in Columbia, Mo. The dates are March 21-26, May 16-21 and Aug. 1-6.

FIRST VENTURE

Party town arrest patterns uncovered

By Shelley Nelson, *Duluth* (Minn.) *News-Tribune*

Duluth's city council has a theory – keep the bars open an hour later and you'll keep the money from flowing over the bridges that connect Duluth, Minn., and Superior, Wis.

Superior police officers scoff at the idea.

Superior's reputation as a "party town" is nearly as old and well-documented as the city's history. The availability of cheap alcohol and a wide variety of bars and nightclubs in a 10-block area of a 45-square-mile city create a nightlife culture that is far different from Duluth's.

As I talked to Superior police officers trying to gauge the expected impact of a later bar closing time in Duluth on law enforcement in Superior, I discovered most police officers didn't think it would make a difference. The size of Superior's larger-than-average police department has long been attributed to late-night partying and violations committed when alcohol is consumed.

One night shift officer told me that if the police department could close off the bridges at night, about half their problems would go away. Several others on the night shift agreed with him.

I was skeptical.

I read police arrest logs routinely. Simple observation suggests a lot more Superior residents are arrested than Duluth residents.

That's where the computer-assisted reporting techniques that I learned at an IRE and NICAR Boot Camp earlier this year came in.

Addresses are key

Getting the data to answer those questions was the easy part. Department command staffers were willing to accommodate my request.

The database itself was a problem.

The lack of a system expert proved to be the biggest difficulty. The department doesn't have the know-how on staff to retrieve the full body of information. The system expert retired several months before and was never replaced because of budget constraints.

That meant I had to manually enter home addresses of arrestees from reports to find out where Superior's crime problems were coming from. While the data included arrest dates and times, arresting officer, arrest district, incident numbers, the suspect's name and birth date, and charges leveled at the time of arrest, staff had no way of retrieving the data to include the arrestees' home addresses.

The data was provided in text format. I did an initial import into Microsoft Excel 2003 to conduct a visual review of the data to make sure it imported cleanly. From there, I saved it as an Excel file to import into Microsoft Access.

Deadlines limited the number of records that could be updated with the addresses of arrestees, but throughout the input process – every 100 records – I tested the database to determine how much difference there was in the number of arrestees from various communities and states. I saved the query to avoid having to rewrite it each time.

Throughout the gathering process, the difference – the percentage of people arrested coming from various local communities – was fairly small, varying from 0.3 to 1.2 percent at any point in the collection process.

Ready to write

After acquiring home addresses for about 1,800 of more than 2,300 arrest reports, covering a year ending in August, it was time to write. The consistent testing made me comfortable that the data fairly represented the arrestees.

With Access, I looked at the times of day and locations of arrest to determine where and when Superior police had their greatest problems. I looked at the types of charges leveled at the time of arrest. I looked at the home address of arrestees citywide, and then focused on the trouble areas and times, determining the home addresses of people arrested in the entertainment district and where those arrestees were coming from.

I copied and pasted Access query results into an Excel workbook for analysis. I could easily tab through the information as I wrote.

That allowed me to test my query logic without relying on recollection, and in one case, catch an erroneous result. It also allowed me to look at the results to see what was happening and determine what type of graphics would best tell the story.

It also proved handy when it came time to pass along information to the *Duluth News-Tribune*'s one-man graphics department.

Data findings, interviews

The findings substantiated what officers were saying about the draw of Superior's party-town culture and what they observed on patrol.

Superior residents accounted for about 60 percent of all arrests made by the department, but Duluth residents accounted for more arrests made in the entertainment district between 11 p.m. and 3 a.m. In that time frame and area, only about 33 percent of people arrested in the entertainment district and 27 percent arrested during those hours actually lived in Superior. Overall, Duluth resi-

continued on page 15

Bits & Bytes

continued from page 2

CAR Conference

Planning is under way for the 2004 Computer-Assisted Reporting Conference in Cincinnati, scheduled for March 12-14.

Panels and workshops will address campaign finance, census, military, federal contracts, international data, environment, local government, business, sports, health, education and much more.

For CAR beginners, special panels will be coordinated with hands-on classes. A special "Start to Finish" track will focus on broadcast needs and issues. Special pre-conference workshops will be offered March 11 for advanced CAR users.

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

For more information, registration and the latest schedule, visit *www.ire.org/cincy04*.

Breaking news

As major news breaks, IRE and NICAR update their Web pages to help journalists get a head start in reporting. When Hurricane Isabel struck North Carolina and Virginia, IRE and NICAR added hurricane story ideas, tipsheets and databases to the growing list of breaking news resources.

Although this year's hurricane season is over, many of the resources can be tapped during other disasters. See *www.ire.org/inthenews_archive*.

NICAR UPDATE

New service opens door to most recent data

By Hsuju Ho, *IRE and NICAR*

Deadline pressure can force journalists to set aside their data because of the time it takes to process and digest it. The IRE and NICAR Database Library now has a solution to that problem: a user-friendly online subscription service that provides cleaned and updated databases for instant use.

In the past, the Database Library staff took government data orders from journalists and made the data available on CD or via FTP download. Journalists can still purchase data this way and use the online subscription service to report on breaking news.

The online subscription service, introduced in spring 2003, allows journalists to obtain data as fast as a government agency updates it. IRE and NICAR set up a special Web page for each subscriber and links that page to the latest available data files.

Twenty news organizations have signed on so far.

The Database Library processes the data and puts it in a format that is easy to use with Microsoft Access, FoxPro or other database managers. The service also includes the record layouts and other documentation.

The service not only provides a shortcut for journalists who lack the time to deal with raw data, it also helps save time for journalists who could process the data themselves but would rather spend their time on other pursuits.

Another advantage of the online service is that it makes data available to subscribers any time. Subscribers can download the data when they like, even on weekends.

The databases that are offered are released weekly or monthly by government agencies and then immediately processed by the Database Library. News organizations can update their own database as frequently as they wish.

Weekly data updates offered on the service include the Federal Aviation Administration Accidents and Incident database, the FAA Service Difficulty Reports and Federal Election Commission campaign contributions.

...it helps save time for journalists...

Monthly data updates include the FAA Aircraft Registry, FAA Airmen Directory and the Internal Revenue Service tax-exempt organization database.

The Database Library is adding databases from its extensive collection slowly to ensure proper processing and the accuracy of the data. Keep an eye on the IRE and NICAR listservs for additions to the online service.

Subscribers also have the option of getting the data updates on CD every month.

For the latest database announcements and more information about the online subscription service, contact Jeff Porter, director of the Database Library by e-mail at jeff@ire.org, or see *www.ire.org/datalibrary/databases*.

Contact Hsuju Ho by e-mail at hsuju@ire.org.

IRE AWARDS 2003

call for entries

The annual contest of Investigative Reporters and Editors, Inc.

DEADLINE: Postmark by Jan. 12, 2004

The contest recognizes the best investigative reporting in print, broadcast and online media, and helps identify techniques and resources used by entrants.

Categories

Newspaper:
For outstanding investigative reporting at a daily or weekly newspaper. (Use highest one-day circulation of the week.) Categories are: Circulation less than 100,000; Circulation between 100,000 and 250,000; Circulation between 250,000 and 500,000; Circulation more than 500,000; and Local-Circulation Weekly.

Television:
For outstanding investigative reporting by a television outlet. Categories are: Network or syndicated program; Top 20 market; and Below Top 20 market.

Other Media:
For outstanding investigative reporting in other media such as Magazine, Newsletter, Specialty Publication, Book and Radio.

Online:
For outstanding investigative reporting.

Special Categories:
Tom Renner Award: Outstanding reporting in any medium covering and exposing organized crime. (Limited to 10 stories) You must submit TWO sets of copies if submitting an entry in the Tom Renner category and in another category.

IRE FOI Award: Honors individual or organization in any medium whose significant actions further open records or open government. You must submit TWO sets of copies if submitting an entry in the FOI category and in another category.

Student Award: Outstanding investigative reporting by a student in a college-affiliated newspaper, magazine or specialty publication, or broadcast work that has been publicly reviewed, screened or aired.

International Entries: International entries will be placed into appropriate categories by IRE staff. Contest judges can then move entries into other categories. IRE can award a special citation for deserving international work.

NOTE: Judges reserve the right to give more than one award in a category or to declare no winner in a category

For entry forms and additional information, visit our Web site at **www.ire.org/contest**

MAPPING IT OUT

The latest uses of mapping in news reporting.

Finding a snow job after storm

By Janet Roberts
St. Paul Pioneer Press

Our story began when a reporter's friend woke up one winter morning to find her car had been towed. She'd come face to face with St. Paul's confusing snow-emergency rules, which govern on-street parking after a significant snowfall.

Over lunch that day, the reporter and his colleagues wondered whether the rules are applied fairly. Are some neighborhoods towed more heavily than others?

With a few computer-assisted investigations under his belt, reporter Chuck Laszewski knew he could answer that question. ESRI ArcView 3.2 geographic information system (GIS) software — and a powerful extension called Spatial Analyst — would make it possible. All Laszewski needed was the right data.

Confusion aside, the snow-emergency rules make sense: Plows can't get the snow off the streets if parked cars are in the way. If streets aren't plowed clean, neighbors cope with a mess of frozen slop that can last all winter long.

But the rules cause a lot of consternation for newcomers — and even long-time residents who pay them too little mind.

Laszewski set out to get ticketing and towing data.

It would be the following winter before we sat down at our computer and began mapping. To make a long story short, it took more than six months — and the intervention of our lawyer — to wrest towing records from the St. Paul Police Department. The ticket records proved to be an even bigger hurdle, so Laszewski spent about three weeks' time, spread over five months, typing in ticket addresses for one snow emergency — about 1,200 in all.

Now we were ready to fire up ArcView, right? Not so fast.

Imagine it's just snowed more than 3 inches. A bunch of tow-truck drivers go to work. It's cold. It's wet. It might even be dark. The drivers get out of their trucks, hitch the car to their rigs and — oh, yeah — scribble down the address. Then it becomes the job of a clerk at the Public Works Department to decipher that scrawl and type the address into the database.

You can imagine how clean the address field was. We had gems such as "CAR X HUH?" Other records contained street names but no house numbers. The "good" addresses were listed this way: "Sherburne/1033." I used Microsoft SQL Server and VB Script to parse the address field and format it in a way ArcView would recognize.

My first pass at geocoding the records — asking ArcView to plot the tow locations on a street map — yielded a disappointing success rate, so it was back to address cleaning. Fewer than two-thirds of the addresses plotted, so I spent the better part of a week examining the records that failed to geocode and correcting misspellings. Ultimately, we had an 88 percent success rate — lower than I like, but the best we could do for this project.

Concentration of tows

Finally, we could begin our analysis. Immediately, we noticed the tows appeared to concentrate north of Grand Avenue, a commercial thoroughfare that roughly divides the city in half from east to west. Indeed, a quick count in ArcView and a few calculations in Microsoft Excel showed 78 percent of the tows were north of Grand. We had a story.

The Grand Avenue observation would become part of our nut graf: "It's almost as if the towing companies and city officials stood on Grand Avenue and drew a line in the snow. To the south, considerably fewer cars get hauled away."

Next, we overlaid a map of the neighborhood boundaries and did a spatial join to assign each towing record to its neighborhood. Exporting the data back to SQL Server, I summed the records by neighborhood. Clearly, towing was heavier in some neighborhoods. But we wondered if that was because more scofflaws parked in those neighborhoods. Enter our ticketing data.

In ArcView, I geocoded the ticketing data Laszewski had gathered. Here, there was no "line in the snow." Ticketing was dispersed fairly evenly across the city. Again, I summed by neighborhood. With those totals in hand, I was able to compute each neighborhood's tow rate: the number of tows as a percentage of tickets issued.

We found wild variation: The neighborhoods with the most tickets had some of the lowest tow rates. Neighborhoods with far fewer tickets had high tow rates.

Fine-tuning with extension

Our story was getting better. But we wanted to make more sense of our data. In particular, we wondered where the densest concentrations of tows were. Common sense told us they probably were in areas closest to the impound lot, where towed cars are taken. We knew — because we used ArcView to draw one-mile buffers around the impound lot and counted the tows in each circle — that 68 percent happened within four miles of the impound lot.

It was difficult, however, to see concentrations in the geocoded points. Because we mapped so many tows, they all blended together into a blob of largely undecipherable color.

Enter Spatial Analyst, an ArcView extension that can do density analysis. We didn't have the software, and at $2,500, it's not cheap. So we sought help from a college mapping instructor. We brought our data to her lab, and she showed us how to use Spatial Analyst to draw our hot spots. The result was eye opening.

We drew separate hot-spot maps for the ticketing and the towing data. The density patterns did not correspond. In particular, a definite concentration of ticketing on the city's West Side lacked any corresponding concentration in the towing map.

Where did the towing hot spots concentrate? Around the impound lot.

Now, not only did we have a story, we had a great publishable map and good reason to buy the Spatial Analyst software.

With time to play with our own software we tested several more theories: that towing patterns would correspond with density of population, rental housing, retail development and immigrant populations (language barriers make it hard for immigrants to understand the complicated parking rules).

All of these theories had some truth. But they didn't explain why the city wasn't towing on the densely populated West Side, an area with many Hispanic immigrants, dense clusters of rental housing and a high ticketing rate.

This is where Laszewski's shoe-leather reporting was invaluable. Tow-truck operators told him they are loath to cross the Mississippi River to tow on the West Side because it is a long round-trip to the impound lot. The city's towing contracts offer incentives to operators the more cars they tow. Towing operators said they focus their efforts north of Grand Avenue because that strategy yields the most cars.

The city's public works director, who coincidentally lives in the seldom-towed West Side neighborhood, said he would review city policies in light of our story. Across the river in Minneapolis, city officials were so enchanted with our analysis, they did a similar one of their own, finding the same disparate patterns.

Read the story online at *www.twincities.com/mld/pioneerpress/4658246.htm.*

Contact Janet Roberts by e-mail at jroberts@pioneerpress.com.

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

SPOTLIGHT: CITY GOVERNMENT

Showing contributions from strip-club workers

By David Washburn, *The San Diego Union-Tribune*

At first blush it sounds like the premise for an episode of a TV drama chronicling the adventures of a hard-nosed newspaper reporter:

The reporter hears rumors that a strip club owner from Las Vegas bribed a few city councilmen, hoping that they will push through legislation that would allow more touching between strippers and their customers. The FBI gets wind of it and plans an afternoon raid at city hall.

The reporter gets a tip about the raid from her sources at city hall. The raid goes off as planned and the reporter has a blockbuster story.

That's what happened in San Diego. FBI agents in May visited the offices of city council members Ralph Inzunza, Charles Lewis and Michael Zucchet and seized their files, telephone records and computer hard drives. Federal agents suspected the three had solicited bribes from Michael Galardi, the owner of two strip clubs in San Diego and several more in Las Vegas.

That was where the similarities between a TV script and reality ended.

In TV-land, the strip club kingpin hires a couple of goons to whack the city councilmen. The reporter walks in on the goons getting ready to do the dirty deed. Using her martial arts skills, she knocks out the goons and the grateful councilmen confess.

But in real life everybody hired lawyers, the FBI stayed quiet and all the journalists in town hounded their frazzled city hall sources. Following the money and establishing a link between the councilmen and the strip club owner through traditional reporting was far easier said than done.

This is the spot *Union-Tribune* reporter Caitlin Rother found herself in during the days after the raid. She heard lots of rumors about Galardi and the councilmen but little in the way of hard facts.

She later got some of the hard facts from a database of campaign contributions that we built on the fly using Microsoft Access.

An analysis of the database showed that people involved in the adult entertainment industry had given more than $20,000 to the three councilmen and two other council members whose offices were not searched by the FBI.

We were able to show that strippers and bouncers were giving to the campaigns of the three councilmen. Also, we found thousands of dollars in contributions coming from Nevada, especially from Las Vegas "homemakers."

Here is what else we found:

- Campaign finance reports varied in the way they listed people associated with the adult entertainment industry and their occupations, employers and home addresses. For example, someone listed as an entertainer or waitress on one report would be listed on another as a student, homemaker or some other less obvious affiliation. Some donors had unlisted occupations.

- Several people who were listed as giving the maximum $250 contribution to more than one candidate listed apartment addresses in less-than-affluent neighborhoods that usually are not home to political donors.

- Many contributions from people in the industry were given on the same day. This fact helped Rother confirm low-profile fund-raisers held by lobbyists for the industry.

I was thrilled with the final results, but wish that the database had been built prior to the raid. The newsroom had not realized before how important it is to have a database of local campaign contributions on hand.

Building the database

Rother needed to do a thorough analysis of years worth of campaign contributions, and she needed to do it fast. The problem was that San Diego, like the vast majority of municipalities, does not require candidates to file their contribution reports electronically. So she ended up with a box filled with dozens of reports containing thousands of contributions.

I uttered a swear word and began pounding my head on my desk. I had been pushing for a few years to have the newsroom create a database of local contributions. These databases are great for seeing who candidates may be indebted to politically.

The editors had agreed and made creating the database a newsroom priority. But, as is so often happens in newsrooms, the news of the day trumped preparing for the future. The database never got built, so we were caught unprepared when the city's biggest political scandals in a decade hit.

Nonetheless, we decided that a database of contributions to the three councilmen and two other council members who we knew had received money from Galardi was crucial for our reporting.

I created a table for the contributions and then built a data-entry form using Access' form wizard. Whenever possible, I used tools on the form to speed the data entry. For instance, someone could pick the name of a city council member from a list to enter that name into the recipient field of the table.

After that, we rounded up some news assistants, brought in a temp with data-entry skills and went to work entering

details such as the name, address, occupation and employer of the contributor. We also entered the amount given and to which council member.

After more than a week of grueling data entry we had three years of contributions for the five council members.

Not long after that I caught a mistake. Some contributions had been entered twice because of addendum reports filed by the candidates. When candidates need to make a correction or addition to a report, they file the entire report over again, which means that, in some cases, the same contribution is reported twice.

It is easy to spot an addendum if you are looking for it. But a few slipped through. So I spent another day typing the data from the addendums into a table and then joining that table to the main table to identify the duplicates and remove them using a delete records query. We ended up with 7,500 records.

While the news assistants and I built the database, Rother mined her sources for names of people associated with Galardi. She compiled more than two dozen names with suspected ties. We queried the database to see whether they had contributed and found that almost everyone on the list had given to the candidates. And we found that a vast majority of those people worked for Galardi. In addition, we noticed that their occupations had been listed differently.

A federal grand jury indicted the three councilmen, Galardi and two lobbyists in August. All of them were charged with wire fraud and five of the six with extortion.

The time and the effort that went into the database were worth it and bolstered the story. Our earlier stories about the federal probe would have been much stronger if we had built the database sooner and used it.

Contact David Washburn by e-mail at david.washburrn@uniontrib.com.

SPOTLIGHT: CITY GOVERNMENT

Spreadsheet, good eye help find missing money

By Paul Egan, *Freelance*

When political candidates fail to report receiving campaign contributions, the money has essentially disappeared.

Unless the contributor reports the candidate's failure to publicly report the money, which is unlikely, there is almost no way of knowing the candidate ever received it.

The exception is with contributions from political action committees. Unlike individuals, whose contributions are only reported by the candidates who receive them, PACs must file public reports that detail their political contributions.

Using a spreadsheet or database manager to compare the contributions that PACs report making to a candidate with those the candidate reported receiving from the same PACs is an excellent way of testing the accuracy and integrity of a candidate's campaign-finance reporting.

In July, *The Detroit News* used Microsoft Excel to conduct such an examination of Detroit Mayor Kwame Kilpatrick, a former Democratic minority leader in the Michigan House of Representatives. Over the two-year period studied, 1999-2000, The *News* found that PACs reported writing 30 checks totaling $16,650 to Kilpatrick's state House campaign fund that Kilpatrick never reported receiving. Journalists could use a database manager, such as Microsoft Access, to match the data, especially where there are many contributions.

Raising questions

Three factors prompted the *News* to take a look.

First, the PAC money Kilpatrick's state House fund reported receiving – a little over $120,000 over the two-year period – appeared low relative to other state legislators in leadership positions.

Second, Kilpatrick had a history of noncompliance with state campaign finance laws, accumulating more than $13,000 in fines over five years for late filing of reports to the state and county.

Third, the state had recently refused Kilpatrick's request to dissolve his state House committee, because of reporting discrepancies of between $1,400 and $2,900 – much smaller discrepancies than those uncovered by the *News*.

Without a computer the analysis would have been next to impossible. Even with one, it required considerable drudgework.

In Michigan, where some committees file electronically and others still file paper reports, data on candidate receipts and PAC contributions is supposed to be easily accessible in electronic format. Data filed electronically is added to the Michigan Secretary of State campaign finance database that can be queried on the Web (see *http://miboecfr.nicusa.com/cgi-bin/cfr/contrib_anls.cgi*.) State workers manually enter information filed on the paper reports into the same database.

Getting a file of Kilpatrick's reported contributions was simple: I punched in the candidate's name, set the date parameters for 1999 and 2000, and downloaded an Excel file with all the reported receipts. Then I deleted the records of contributions from individual contributors, leaving only 109 PAC contributions.

Getting a file of contributions that PACs reported making to Kilpatrick in 1999 and 2000 was much more difficult. Most PACs file on paper. The state uses the candidate reports as its source of PAC contribution data. So Kirkpatrick's unreported PAC contributions would be missing.

The only way to get the data was to examine scanned reports from all 1,000 or so Michigan PACs during 1999 and 2000 and manually create a spreadsheet of contributions they reported making to Kilpatrick. Using a similar format to the first spreadsheet, my new table included a column for the name of the contributor, one for the amount of the contribution, and one for the reported contribution date.

Contributions reported to Kilpatrick by the PACs had 145 records. Since the databases were relatively small, I sorted each spreadsheet alphabetically by contributor and used my eyes to do the matching, striking out pairs of contributions and receipts until I was left with a list of what appeared to be unreported contributions. If I were to try a similar analysis again, I'd clean up the files, load them into Access and try joining them by contributor name and amount.

Checking on the PACs

Looking at the unmatched contributions, I saw that I still had plenty of work ahead.

PACs are sometimes known by more than one name, and it's important to check for inconsistencies between the name used by the contributor and the one used by the recipient when recording the contribution.

We found instances with matching contributors but differing contribution amounts. Because there was no easy way to determine which table contained the correct amount, the *News* did not include these contributions in its results and focused solely on contributions that were not reported at all by Kilpatrick's committee.

It's also possible that a contributor reported making the contribution within

the time period covered by the analysis, but the candidate did not report receiving it until a later time period. For that reason, we searched all reports filed by the candidate after the time period covered by the analysis.

Contributions can go astray in other ways. For example, the *News* found instances in which a PAC reported contributing to Kilpatrick's campaign committee, but Kilpatrick reported depositing the money into his own political action committee, called Generations. Since he reported the contributions as received somewhere, these were not included as missing.

Interestingly, among the missing contributions the *News* found were two totaling $3,150 that Kilpatrick's PAC reported making to his campaign committee but his campaign committee never reported receiving.

Lost in the mail?

Another problem to be aware of is contribution checks that get lost in the mail or become outdated because they are not cashed quickly enough. In those cases, Michigan law requires the PAC to adjust the balance by reporting the amount of the lost or uncashed checks as "other income" in a subsequent report.

It makes sense, then, that the time period chosen for the analysis should be far enough in the past so such anomalies can be discovered and accounted for in the reports. That was one of the reasons the *News* examined the final two years when Kilpatrick was in the state House, rather than the period during his more recent job as mayor of Detroit.

As a further check, the *News* contacted a sample of the PACs that reported making contributions to Kilpatrick that Kilpatrick never reported receiving. Contributors confirmed that at least eight of the unreported checks, totaling $5,150, had cleared the bank.

After the *News* published its report, a spokesman for Kilpatrick acknowledged that there appeared to be a reporting problem and said Kilpatrick's committee had hired an attorney who specialized in financial matters to try to resolve the issues.

Paul Egan is a former staff writer for *The Detroit News* and can be contacted by e-mail at paulegan@sympatico.ca.

readme.txt

IRE offers tipsheets for journalists interested in using databases to cover local campaign finance.

Michael A. Weber, co-author of the IRE Beat Book "Unstacking the Deck: A Reporter's Guide to Campaign Finance," offers a list of Web addresses where you can find local, federal and state campaign finance data in Tipsheet No. 1919.

Tipsheet No. 1222 provides guidance to journalists for learning about campaign contributors and their employers. This tipsheet also explains how to use and sort a donations database to look for story ideas.

In Tipsheet No. 1221 Bob Warner of the *Philadelphia Daily News* offers advice on obtaining campaign finance databases from local and state officials who may be reluctant to hand them over. Warner also provides tips for cleaning the data before you analyze it.

IRE members can download the tipsheets from *www.ire.org/resourcecenter*. You can also order tipsheets from the IRE Resource Center at 573-882-3364 or rescntr@nicar.org.

City

continued from page 1

quickly add, subtract, order from highest to lowest, find the average or median and calculate percentage increases. I had to learn Microsoft Excel.

Now, I realize I risk ridicule here. Many readers of this publication are interested in how to write the shortest program to sort through millions of government records in Microsoft Access or FoxPro.

But I also know many local government reporters who find themselves too busy to regularly build spreadsheets for their stories. I've come to find those spreadsheets are essential in almost every story. In fact, they can help a reporter take more control of the beat and its stories by generating information – and stories – that no one else has.

What I will attempt to detail here is some of the key data sets on the local government beat, how I used them to build spreadsheets and the stories that resulted.

The questions that prompted my requests for the data or documents I used to build the spreadsheets are basic to most any beat.

How much does that government official make?

How much does that perk or program cost taxpayers?

Start with your local government's annual budget. It should break down expenses in a general way. Look for any interesting notes and be sure to compare that budget to previous years, to search for major changes.

Also be sure to obtain a copy of the annual management letter, normally done by outside auditors. It should detail problems in financial controls, reconciliation of accounts and other problems related to revenue collections.

The city budget also should total personnel costs. But you'll need to dig a bit deeper to find out which government employees have the best deals in the way of pay.

Personnel and payroll

You need basic salary information about everyone. That should include name, department, salary, overtime and any other pay. A quick and easy story could focus on how many officials earn more than $100,000. You might be surprised by what you find. In Kansas City, we found 60 city officials in the "$100,000 club."

Then focus on the overtime. Examine by departments. What are the trends over the years? Examine also by individuals. Did any employees double or greatly boost their annual pay with overtime?

Start with your local government's annual budget.

I obtained some of this information electronically. It was in a rather clunky form, however. It was organized by payroll date – so each employee had 26 entries per year. I imported the data to Excel, then created pivot tables on the names to group all payroll information. Then, I sorted in descending order.

Another angle to pursue on overtime is to look at how often city officials, including police and fire, are paid for overtime they don't actually work. In Kansas City, police are paid a minimum of four hours overtime for a callback, even if they only work one hour. *The Star* obtained a few weeks of the police department's callback sheets, which detailed the actual time worked versus the amount paid.

Using a spreadsheet, I logged each callback, the amount paid in overtime and the amount actually worked.

Perks

A number of basic, but interesting, city or county records will be a good starting point in determining the perks that your local government employees enjoy.

Early retirement programs are popular today as local governments attempt to trim budgets. Obtain a list of every employee, position, salary, years of service and monthly retirement payment. I used such a list, building most of it by entering the data myself, when about 350 employees opted for an early retirement program in Kansas City.

With the spreadsheet, I was able to detail who was receiving the largest monthly retirement payments, the average payment, the years of service lost to the city and losses by department.

And then there are cell phones. When I set up shop in city hall, I quickly found it difficult to find any official by leaving a message on their office phone. The best way to reach them, I found, was on their city cell phone. That, of course, prompted the question: How much does that cost taxpayers?

Using an open-records request, I obtained electronic data about each cell phone, its city user, department and total monthly bill for three months. We totaled the three months of bills by user, then ranked them.

We filed another open-records request, asking for detailed billing information about the top 25 users. That revealed that many, including two city council members, were abusing their phones or racking up exorbitant bills. Others were doing work on their cell phones, but they were on the wrong plans, regularly exceeding their allotted "free" minutes.

Take-home vehicles traditionally are another city or county government

perk. We requested the number and assignments of vehicles by city department. The city auditor had studied this a couple years earlier, so we compared our update figures to those in the city audit, creating a spreadsheet by department to track the increases or decreases.

To flesh out the story a bit more, we also obtained logs of city employees assigned to vehicles, which showed how often they use the vehicles for something other than commuting to and from work.

Property records

Be sure to obtain copies of county assessment records. These could be a foundation for a number of stories. Who has had the highest or smallest increases in assessment?

Also ask for delinquent property records, which may tip you off about local developers and businesses in trouble. Who owes the most in back taxes? Do many property owners allow their property tax bills to go delinquent for two years or more before paying the bill?

In addition, cities must create inventories of their city property. Check the city's real estate lists, matching city-owned addresses against lists of code violators. I recently found that the city owned land on which illegal junkyards were being operated.

Other fun records

Inspection records of restaurants, swimming pools, day-care centers and even salvage yards can produce great stories. Plug into a spreadsheet the inspection dates, findings, investigator's names and other document information. Patterns will emerge that you might not have caught otherwise. Also, spreadsheets are a great tool for keeping the information close at hand.

I recently had fun by obtaining city water department listings of water use at city-owned sites, including the city zoo and numerous fountains. With that information in a spreadsheet, we could calculate the amount of "free" water going to city sites.

The city kept spreadsheets on its free water users, including their addresses and the amount of water per month. With that information, we calculated usage per hour for some fountains and compared it to the average monthly usage of a family of four. We also toured several of these sites with water officials to refine our calculations.

Then there was water lost through leaks. In some areas the department lost as much as half of its water and ended up collecting fees for only two-thirds of the water it had treated for distribution. Big leaks happen not only in Kansas City but also in almost every city east of the Mississippi River because of the age of the city water distribution systems.

Finally, *The Star* also recently obtained the parking records for municipal judges, who had long been accused of working short hours in the city courthouse.

For many courthouses, this idea of using parking records to mark their workweeks would probably not work. But for judges whose cases are all heard from the bench in a docket-style fashion and where little legal research is required, it seemed like a good way to determine how much time the judges were spending on city business in the courthouse.

For this story, though, I had to advance beyond spreadsheets, enlisting the support of Greg Reeves, the newspaper's FoxPro guru, to sort through the electronic swipes of judges, who used cards to come in and out of the courthouse parking garage.

With Reeves' help, I had moved beyond Excel. But it would never have been possible without those initial steps 18 months ago into the world of spreadsheets.

Contact Michael Mansur by e-mail at mmansur@kcstar.com.

CAR TOOL

Free tool reads, converts SAS data

By Jeff Porter, *IRE and NICAR*

SAS, the expensive statistical software used widely in government and academic circles, might be out of reach of many newsroom budgets. But there's still a way – for free, no less – to read SAS data files and translate them into a format your spreadsheet or database manager can use.

From the company's Web site, you can download the SAS System Viewer at no cost. To find it, go to *http://support.sas.com* and click on the link to the left called "Software Downloads." That takes you to a page of downloads, categorized by product type. Click on the link called "Base SAS Software." You should see a short list of items available for download, including the SAS System Viewer. Click away, and follow the screen instructions. You'll have to create a login, including your e-mail address.

Installation should be a snap. Download the executable file and just double-click on it to start the installation.

After the program installs, it's time for some practice. Go to the Web page *www.usda.gov/cnpp/healthyeating.html*. There you can download data on the USDA Healthy Eating Index, a summary measure of overall diet quality. It provides a picture of the type and quantity of food people eat and the degree to which diets comply with specific recommendations in the Dietary Guidelines and the Food Guide Pyramid.

Choose to download the 1999-2000 Healthy Eating Index data in compressed, self-extracting SAS 6.12 format. Also download the 1999-2000 Food Guide Pyramid Servings database. The files are relational, creating a summary of food intake by respondent.

This will not be the most exciting database you've ever worked with. It does, however, give you some practice in using SAS System Viewer.

You'll need to extract two files: hei9900.sd2 and pyraserv9900.sd2. Start the SAS System Viewer, then go to File > Open. Keep the default setting under Files of Type as All SAS Datasets, then navigate to the appropriate folder and open the hei9900.sd2 file. It should look something like Figure 1.

hei9900.sd2

	Respondent sequence number (SEQN)	Total Fat Score (FAT_SC)	Saturated Fat Score (SFAT_SC)	Sodium Score (SODI_SC)	Cholesterol Score (CHOL_SC)	Grain Score (GRAIN_SC)	Fruit Score (FRUIT_SC)	Vegetable Score (VEG_SC)	Meat or Meat Alternates Score (MEAT_
1	1	10	10	10	10	10	10	6.395704	2.9540
2	2	10	10	0	9.0689333	7.7505194	10	10	
3	3	10	10	10	10	4.0819607	6.1875	4.1446270	4.9571
4	5	4.15034	5.6220	4.3485	10	7.9170272	0	10	5.6

Figure 2

Now, let's check out a few features of SAS System Viewer. First, notice that the column headers are somewhat cryptic. So go to the View menu item and choose Headers and Names. Now, you'll see something a little nicer-looking in Figure 2. Prefer just the shorter version? Toggle back to headers.

You can also view the variables, too, under the same View menu, and you'll see a list of all the variables (also known as columns) in the SAS dataset. To go back to viewing the data, just go back to the View menu and choose Formatted.

Now, some final few steps to make the data useful in different formats. Choose how you prefer to view the data – with headers or the lengthier names. Typically, if I'm going to use a file in Microsoft Excel, I like longer names. If I'm going to throw something into a database manager, shorter field names are my preference. So consider how

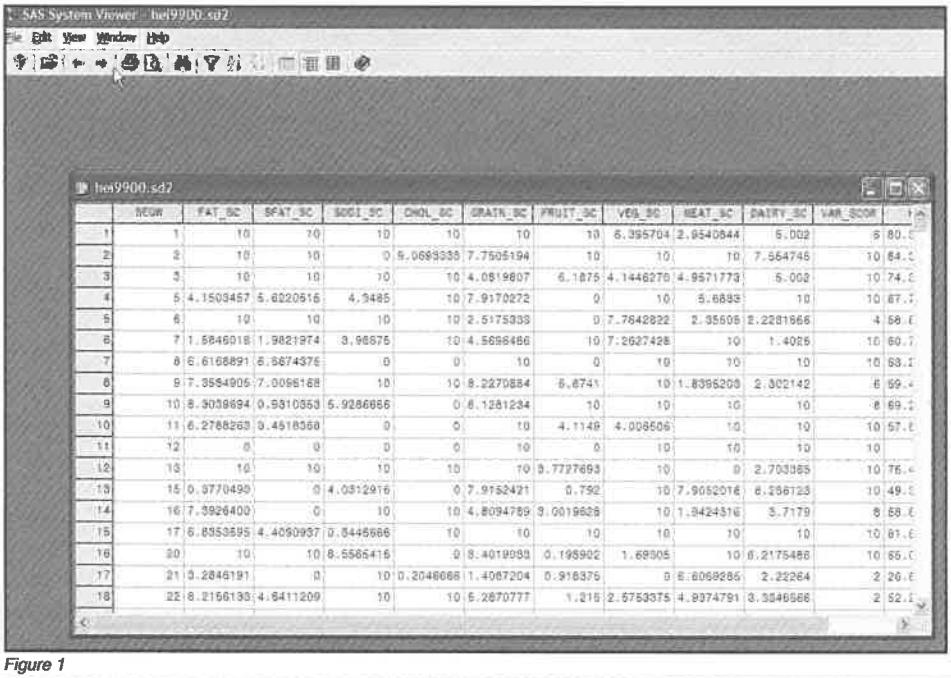


Figure 1

readme.txt

Have a favorite CAR tool that Uplink readers should know about? Drop us a note at uplink@nicar.org and tell us about it.

you're going to use your SAS file, then simply go to File > Save As and choose your File of Type.

Now, you've saved the file as a text file of your choice, ready to be opened in a spreadsheet or database manager.

Of course, the viewer doesn't offer many of the features of the full-blown version SAS. You can download, review, even edit SAS program files with the extension .sas, but you can't run those programs with the viewer alone. There's a long list of SAS-type files you can just open in read-only mode, including the data files such as .sd2.

Of course, you can then save those as text, reopen them with different software then edit data tables. Text files opened with the SAS viewer – such as .csv, txt, prn, .dat – can be edited, although there are better applications for that.

SAS isn't the most common data type that government agencies post on their Web sites, so you probably won't use the SAS System Viewer every day. Even government agencies that process data with SAS have the ability to provide the data in almost any way you request – text, Excel, Access, dBASE – so theoretically, all you have to do is call or e-mail with a request and they'll send it right out to you.

But if you're on deadline, sometimes you don't have that luxury of time, so you can at least view and convert SAS data files with a free addition to your CAR toolbox.

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

Arrests

continued from page 3

dents only accounted for about 18 percent of all arrests, but 37 percent of the arrests made in the entertainment district and 40 percent of the arrests made between 11 p.m. and 3 a.m.

I was surprised by the results, but when I went back to talk to officers who work the night shift about the problems they deal with, they weren't surprised at all. That was what they expected based on their patrol experience and what they were talking about during earlier discussions.

Since the story ran on Sept. 8, the Superior city council has gotten behind the chief's efforts to create a satellite police office in the city's entertainment district. It's also prompted the chief to look into ways to make people who live outside the community responsible for some of the law enforcement costs they generate and city residents bear.

It has also prompted the chief to seek better training for his records staff so they can look at the information in different ways. It's training that will be useful a year after the new Duluth bar closing time goes into effect and I look back to see if the extended drinking time makes a difference in Superior's law enforcement efforts.

The one finding that really took me by surprise – and the records staff was able to verify – is the number of underage drinking arrests police make. The vast majority of those types of arrests in the entertainment district occur around one nightclub during its "dry" nights, a weekly alcohol-free event.

That's a story for another day.

Contact Shelley Nelson by e-mail at snelson@duluthnews.com.

Upcoming CAR Workshops

Computer-Assisted Reporting Boot Camps

These unique seminars, taught by IRE and NICAR's experts, train journalists to acquire electronic information, use spreadsheets and databases to analyze the information and to translate it into high-impact stories. In addition, IRE and NICAR provide follow-up help when participants return to their news organizations. Four CAR Boot Camps are scheduled for 2004 in Columbia, Mo.

- **Jan. 4-9**
- **March 21-26**
- **May 16-26**
- **Aug. 1-6**

Mapping Data for News

Jan. 9-11 • Columbia, Mo.

Intensive hands-on training using mapping software for news stories will be offered by David Herzog, of NICAR and the Missouri School of Journalism, and Jennifer LaFleur, computer-assisted reporting editor at *The Dallas Morning News*. Participants are asked to have a basic knowledge in using relational database programs such as Access or FoxPro.

Advanced Statistics Workshop

Feb. 13-15 • Arizona State University

Pulitzer Prize-winning journalists Steve Doig, interim director of the Walter Cronkite School of Journalism at Arizona State University, and Sarah Cohen, database editor of *The Washington Post*, team for this workshop. This session is aimed at strengthening the skills of reporters who want to move beyond basic CAR and use statistical analysis in their work. Reporters should know spreadsheet and database manager applications and have experience in CAR.

Fellowships

A limited number of fellowships are available. Visit www.ire.org/training/fellowships.html for more information and an application.

More information is available at www.ire.org/training/bootcamps.html

Tech tip...

Geometry, scripting find nearby points

By Brad Heath, *The Detroit News*

Your state is pondering a new law: No convicted sex offenders should live within a mile of a day-care center. But what's the context? To find out, you need to answer two questions: How many sex offenders actually live near each day-care center in your neighborhood and, for each one, exactly how far away is the nearest offender?

You know how to start answering these questions using your geographic information system (GIS) program – I used ESRI ArcView 3.2. You've already used geocoding to turn a data table of day-care center street addresses in your community into a point map, and did the same for a table of convicted sex offenders. So now all you have to do is count up which ones are near each other.

How?

That's where it gets interesting.

ArcView will need to go through and check each day-care center for the closest sex offender. And it will have to count how many offender points are within a mile of each center. That's simple to do by hand if you're only looking at one or two facilities. But what if you're looking at dozens, or thousands?

ArcView doesn't come with that ability. You have to build it yourself using ArcView's scripting language, called Avenue. Once you learn Avenue you can make ArcView do pretty much anything.

In this case, it's using coordinate geometry. To answer the questions, you'll need to fill out both themes with accurate x and y coordinates. Then you can use scripts or queries to search through the data to figure out which sex offenders are closest to the day-care centers.

Step one is creating the coordinates. If you have a point theme in ArcView, odds are you're dealing in latitude and longitude – decimal degrees. While these are flexible, they aren't the best for measuring distance, so you need to put some coordinates in your data.

Start by changing the projection of your view. Switch to the NAD 1983 projection for your state plane. (These projections are built in to ArcView 3.x.) Also, change the distance and map units to feet. Then open the attributes table for the theme you want to work on first. Start editing it and add two fields: "ProjX" and "ProjY". Set them up as numeric and be sure to allow for several decimal places. Close the table but don't hit the "Stop Editing" button yet.

Now you're ready to add some coordinates. To do it, you'll need an Avenue script that cycles through each record and retrieves the x and y coordinates that correspond to the map projection you've chosen. The code looks like this:

```
curView =
av.GetProject.FindDoc("mymap")
thmToUpdate =
curView.FindTheme(".shp")
tabToUPdate =
thmToUpdate.GetFTab
prjView = curView.GetProjection

fldShape =
tabToUPdate.FindField("Shape")
fldProjX =
tabToUpdate.FindField("projx")
fldProjY =
tabToUpdate.FindField("projy")

for each rec in tabToUpdate

curRec =
tabToUpdate.ReturnValue(fldShape,
rec)

ProjX =
curRec.ReturnProjected(prjView).GetX
ProjY =
curRec.ReturnProjected(prjView).GetY

tabToUpdate.SetValue(fldProjX,
rec, ProjX)
tabToUpdate.SetValue(fldProjY,
rec, ProjY)

end
```

After you enter the code, make sure you update it with the name of your view and theme. Then compile it with the checkmark icon and run it.

Now back to your table. Look at the ProjX and ProjY fields and they should be filled in with numbers – the eastings and northings of your point. Instead of using latitude and longitude, these coordinates identify a point based on how many feet north and east it is from a predefined point on the map. Click "Stop Editing" and save the results. Then export the table you just finished and repeat the process for your second table. Import both into Access.

This leaves you with a database, two tables that lack a common field for joining. But Microsoft Access and other database managers will let you

use more sophisticated relationships.

Instead of using a join, you can link the points in the two tables using the Pythagorean distance formula, which can calculate the distance between any two points on a plane using their coordinates.

Here's how you'd use Access SQL to link the tables based upon points that are at most a mile (5,280 feet) apart:

```
WHERE sqr((daccare.ProjX -
offenders.ProjX) ^2) +
((daccare.ProjY-
offenders.ProjY) ^2)) <= 5280
```

Now you can insert that into a SQL statement to answer the two questions.

First: How many sex offenders live within a mile of each day-care center? You'll want a query that will check the data and count every point where the distance is less than 5,280 feet. Here's the SQL:

```
SELECT daccare.Name,
Count(offenders.Name) AS
offenders_within_mile
FROM daccare, offenders
WHERE
(((Sqr((((offenders.projx-
daccare.projx)^2)+((offenders.projy-
daccare.projy)^2))))<5280) AND
((daccare.projX) Is Not Null)
AND ((offenders.projX) Is Not
Null))
GROUP BY daccare.Name;
```

It's important that your query filter out null values, which will appear in any records that you were unable to geocode. Even if only one record wasn't coded to a spot on the map, the null value will cause Access to choke.

Second: How far from each day-care center is the nearest sex offender? You'll want a query that, for each day-care center, chooses the minimum result from the distance calculation. Here's the SQL:

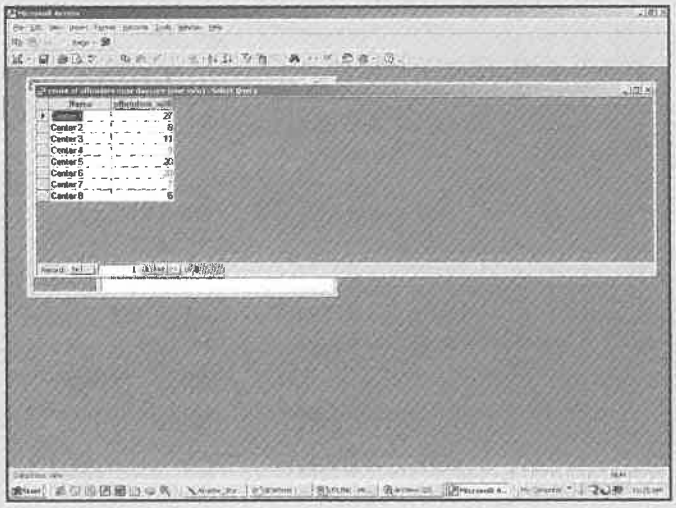
```
SELECT daccare.Name,
Min(Sqr((((offenders.projx-
daccare.projx)^2)+((offenders.projy-
daccare.projy)^2)))) AS dis-
tance
FROM daccare, offenders

WHERE (((daccare.projX) Is Not
Null) AND ((offenders.projX) Is
Not Null))
GROUP BY daccare.Name;
```

It's possible to handle these calculations with Avenue (and you don't even have to get Pythagorean to do it, because the language has a built-in distance function). But bringing it back into a database is easier, not to mention much faster. And once you've figured out how to do spatial math in Access, it's easy to change your analysis, ask different questions of your data or bring in other data, maybe the location of sex crimes or neighborhood police stations.

What else can you do?

Why not find out how many sewer pipes dump out near your community's beaches? Or look at whether fires are happening close to the fire departments. You could add the distance calculations to an analysis of ambulance response times. And you could incorporate it into your Web site by creating searches that let users find the day-care centers closest to their homes.

Contact Brad Heath by e-mail at bheath@detnews.com.



Remote

continued from page 1

other state affected by the disaster.

Those who study or work in the field of remote sensing expect to see similar projects from news organizations down the road.

In a nutshell, remote sensing is the science of gathering information about the Earth from a distance. The process typically involves sensing energy (such as heat, visible light or infrared radiation) that is emitted or reflected by the Earth's surface or atmosphere. This can be done in several ways, from attaching a camera to a hot-air balloon to sending a satellite hundreds of miles into space.

Remote sensing really isn't new to the news business. Think of TV weather maps showing satellite images of clouds or rainfall, or aerial photographs in newspapers.

Blending data, images

News organizations have embraced remote sensing more in the past year, largely because of the war in Iraq. They have shown satellite images of Baghdad, debris from the Columbia space shuttle crash over Texas, and areas in Washington, D.C., where the snipers attacked.

The *Post-Dispatch* series took remote sensing another step by blending other data with satellite images to find and tell a story. Environment reporter Sara Shipley grew up near St. Louis and remembers big floods in the 1980s. When a job at the *Post-Dispatch* brought her back to the area last year, she was struck by all of the businesses and roads that had sprouted on the flood plain.

Shipley wanted to measure how much development had occurred since the 1993 flood, and to compare Missouri to other Midwestern states. Existing documents and data, such as permits issued for flood plain development, could not answer her questions. (The permits are required only for federally backed flood insurance, which not everyone obtains. Plus, the permits don't cover all areas of a flood plain.)

So the newspaper hired GeoSpatial Services (*www.gis.smumn.edu*), a research group at Saint Mary's University in Minnesota, to analyze development in areas of seven states affected by the flood. To mark the flood plain and areas that flooded in 1993, the researchers used data from the Scientific Assessment and Strategy Team, a government group formed to study the disaster.

The researchers also obtained current and pre-flood satellite images from the U.S. Geological Survey. Their third major source of data, showing changes in population and housing, came from the U.S. Census for 1990 and 2000.

Remote sensing can be costly.

After some cleaning and processing, Saint Mary's researcher Jim Hipple imported the satellite images (as geotiff files), flooding and flood plain boundaries (as shapefiles) and Census block data (as database files) into ArcEditor, ESRI's mid-level GIS.

The Census figures helped Hipple zoom in on areas with big jumps in population. He also scrutinized the pre- and post-flood satellite images to scout out new highways, parking lots, businesses and homes.

The satellite images had 30-meter resolution – meaning each pixel of the image represents a 30-by-30 meter surface. To get clearer pictures with higher resolution, Hipple used aerial photographs to study areas with intense development.

Hipple traced the outlines of post-flood development using his mouse to create polygons and place them in one of four categories: industrial/commercial, residential, highway/interchange and recreational. Then he tallied the acres of land developed in flood plain and flooded areas since 1993.

Reporting results

The analysis revealed that Missouri had developed more than 4,200 acres of flood plain, more than any other state affected by the flood. And nearly all of Missouri's development had occurred in the St. Louis area, in areas that sat under water 10 years ago.

The series (*www.stltoday.com/stltoday/news/special/flood93.nsf/front?openview&count=2000*) then explained how lax state regulation and the political climate in the St. Louis region allowed so much development in areas that could someday end up under water again.

"We were able to quantify it. That was the biggest thing," Shipley said. And the satellite images taken before and after the flood drove home just how much development had taken place. "When you're sitting here on the ground, you don't really get that perspective."

Remote sensing can be costly. The Post-Dispatch paid $10,000 for the analysis. Nor is buying remote pictures from a commercial company cheap, at least by most newsroom standards. Images from Space Imaging (*www.spaceimaging.com*) start at $350, while those from DigitalGlobe (*www.digitalglobe.com*) start at $562. However, in an effort to tap into the media market, both companies have given images to some news organizations for free.

But journalists can do other projects and use other data that cost less.

Images on the cheap

"There are many very good shoestring-type projects that can be done using imagery, sometimes as a thing unto itself, other times as a layer in a GIS," said Chris Simpson, who teaches jour-

links, the reporter can also pull up the names of everyone else who donated from that company or from that profession.

But even with the wealth of data that has been put up in newsrooms, a common complaint was that journalists fail to use what's there.

"We're all such luddites in this industry," said James Wilkerson, CAR editor at *The Morning Call* in Allentown, Pa. The paper has an intranet with databases ranging from voter registration files to crime incidents and arrests going back to the early 1990s. "There's a handful of people who use it constantly and some people who never use it."

Matt Carroll can be contacted by e-mail at mcarroll@globe.com.

readme.txt

The IRE Resource Center has tipsheets offering ideas on how to create, develop and improve newsroom intranets. Here are just a few:

Robert Gebeloff of *The Star-Ledger* in Newark, N.J., offers a primer on what newsroom intranets should include so journalists will use them. (Tipsheet No. 1770)

Liz Donovan of *The Miami Herald* created a list of suggestions for directories, public records, news sources, people finders, searches and general reference data for newsroom intranets. (Tipsheet No. 1327)

George Landau of NewsEngin, Inc., provides tips and advice on setting up and running a newsroom intranet. (Tipsheet No. 895)

IRE members can download copies of the tipsheets at *www.ire.org/resourcecenter*. You can order copies from the Resource Center at rescntr@ire.org or 573-882-3364.

IRE and NICAR Services

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

Programs and Services

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

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

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

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

Publications

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

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

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

For information on:

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

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

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

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

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

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

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

Uplink Info

A newsletter of the National Institute for Computer-Assisted Reporting

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

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

Editorial
573-884-7711

Subscriptions
573-882-2772

Postmaster: Please send address changes to IRE-NICAR.

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

IRE members $40, nonmembers $60
Subscriptions

John Green
jgreen@ire.org
Subscription Administrator

Pia Christensen
pia@ire.org
Advertising Coordinator

Len Bruzzese
len@ire.org
Director of Publications

Brant Houston
brant@ire.org
Editor

David Herzog
dherzog@nicar.org
Managing Editor

Jeff Porter
jeff@nicar.org
Asst. Managing Editor

Lisa Triefenbach
Art Director

Pia Christensen
Copy Editor

Christina Caron
Megan Clarke
Jaimi Dowdell
Holly Hacker
Hsuju Ho
Tad Vezner
Contributing Editors