Uplink

March • April 2004
Volume 16, Number 2

Published bimonthly by the National Institute for Computer-Assisted Reporting

www.nicar.org

LAW ENFORCEMENT

Finding 'missing' kids

By Marc Chase,
The Times (Munster, Ind.)

At first glance, the Indiana Missing Children Clearinghouse list showed that our circulation area led the state in the number of children reported missing by law enforcement agencies. The list also held the potential for a number of compelling stories about children who had been listed as missing for several years. But, as we tried to track down those compelling stories, we found the real story surpassed our initial impressions.

We found that at least 16 children, or 10 percent of those listed as missing from our region, were not missing at all. And the city of Hammond likely had more children on the state's list than Indianapolis, which is larger, because Hammond detectives had failed to follow up with parents in those cases. What began as an attempt to turn around a quick computer-assisted reporting project led to a three-month investigation that revealed serious flaws in a

continued on page 22

SPOTLIGHT: EMERGENCY RESPONSE

Survey helps explain heart attack survival

By Robert Davis, *USA Today*

In the early 1990s, a discrepancy between big city survival rates for cardiac arrest cases began to come into focus in medical journals – some cities saved 10 times more victims than others.

As a back-burner issue project that I pursued amid daily stories, I began looking for ways to detail those differences. Whenever I spoke with an emergency room physician, fire department official or cardiology researcher for a story, I asked them about the discrepancies.

I began to see that the officials explained away the differences with reasons they usually said were out of

continued on page 10

SPOTLIGHT:

For more on emergency response see:

- Mapping in Tucson, Ariz., finds potentially deadly lags in response times, p. 6
- Analysis shows EMS in Missouri fails to match suburban growth, p. 8
- Tips for using dispatch data to examine EMS response, p. 9
- A guide to evaluating EMS response, p. 10

PRESCRIPTION DRUGS

Making the connection between doctors, ODs

By Fred Schulte, *South Florida Sun-Sentinel*

In the recent newspaper series "Drugging the Poor," I tried to expose not only the human tragedy of prescription drug abuse, but also the huge financial cost.

The series revealed that a small group of Florida doctors have prescribed large quantities of narcotic painkillers and other addictive drugs to low-income people on Medicaid, costing taxpayers hundreds of millions of dollars and adding to a torrent of overdose deaths in the state.

The project linked some of the state's top prescribing Medicaid doctors to overdose deaths or histories of professional misconduct. It also uncovered millions of dollars in suspect billings for drugs, including claims generated using the license numbers of doctors who were dead or otherwise ineligible to write prescriptions.

Several databases, some created from paper records readily available to the public and others from data

continued on page 20

Bits & Bytes

Fresh consumer product data

How many people did space heaters or turkey fryers hurt in the past year? Journalists can use the U.S. Consumer Product Safety Commission Injury and Death database, recently updated by the IRE and NICAR Database Library, to investigate these and other products.

The database consists of four tables provided by the commission's National Injury Information Clearinghouse. The CPSC maintains this data so it can alert the public to potential hazards associated with consumer products. Fields include age, race and gender of the victim, state and city of the incident, product type involved and more. To order the database, call the Database Library at 573-884-7711 or go online to *www.ire.org/datalibrary/databases/cpsc.*

Tangled webs

IRE and NICAR are beginning to more deeply explore social network analysis for visualizing and diagramming relationships between individuals, businesses and institutions. It's a new approach for journalists, but business consultants, law enforcement agencies and public health investigators have used this approach for decades. In 2003, *The Washington Post* used social network analysis to report on a group of Los Angeles-area language schools that appeared to be "visa mills" for foreign students. For more information and a list of Web links and tipsheets on this emerging technique, go to *www.ire.org/sna.*

continued on page 4

INSIDE NICAR

Getting set for election coverage

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

This is a big year for election news, with the U.S. House, Senate and presidential races going full force. Opportunities abound for using computer-assisted reporting.

If you're new to CAR the elections are a great time to sink your teeth into campaign finance data and see where the candidates draw their financial support. If you've been doing CAR for a long time, there are plenty of compelling stories that look deeper into campaign finance, voter histories and election results.

Starting with this issue of *Uplink*, we'll offer articles that will help you cover elections and politics throughout the election season. Check out page 3 for an account of how Robert David Sullivan of *CommonWealth* magazine in Boston created a map of distinct U.S. political areas. We'll have even more in the May-June *Uplink*.

As the campaign unfolds, go online to IRE's Campaign Finance Information Center at *www.campaignfinance.org* for resources, data and links to stories.

If you want to get a jump start on your elections and politics coverage, you'll want to get "Unstacking the Deck: A Reporter's Guide to Campaign Finance," an IRE Beat Book by veteran journalists Michael A. Weber, Aron Pilhofer and Derek Willis.

For more information about the book visit *www.ire.org/store/books/campaign.html.*

You can also sit in on campaign finance panels and get hands-on CAR training June 17-20 at the IRE Annual Conference in Atlanta. Keep an eye on *www.ire.org/training/atlanta04* for more information.

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

Upcoming hands-on CAR training

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

There are two weeklong CAR boot camps in Columbia, Mo., for journalists who want to learn how to acquire electronic information, use spreadsheets and databases to analyze the information and to translate that information into high-impact stories. IRE and NICAR provide follow-up help after participants return to their news organizations. The boot camps are May 16-21 and Aug. 1-6.

Editors can take advantage of an abbreviated version of the boot camp April 23-25 in Columbia, Mo. Participants will learn how to make CAR successful in their newsrooms and will get just enough hands-on experience to understand what their reporters are tackling and envision what is possible with CAR.

Hands-on classes will be offered at the IRE Annual Conference June 17-20 in Atlanta. Those registering for the classes will be able to learn how to use spreadsheets and database managers to analyze data.

ELECTIONS

Mapping a new look at presidential results

By Robert David Sullivan, *CommonWealth*

"Beyond Red & Blue," an exclusive report on *CommonWealth* magazine's Web site, takes a fresh look at presidential elections by identifying 10 distinct political regions that cross state boundaries. Drawing the new political map of the U.S. required elections and Census data, mapping and spreadsheet analysis and some judgment calls.

The project evolved from "The Lay of the Land," a story about the political regions of Massachusetts. That story, by comparison, was a no-frills affair. I typed election results for all of the 351 cities and towns in Massachusetts into Microsoft Excel, and then assigned each city and town to one of 10 regions with almost equal voting strength.

For "Beyond Red & Blue," I did the same for the 3,117 counties in the United States and used ESRI ArcView 3.3 geographic information system (GIS) to create detailed colored maps that later were turned into graphics for the report on our Web site. (See *www.massinc.org/commonwealth/new_map_exclusive/beyond_red_blue.html*).

Data into Excel

As I built an Excel master data file with the names of the U.S. counties I added elections data collected from Web sites and printed sources. In each state, the Secretary of State's office – or whichever agency oversees voting – posts this information on the Internet. In many cases you can download or cut and paste data directly into Excel. (See *www.politics1.com/states.htm* for links to all 50 states.)

Other great sources of data are Dave Leip's Atlas of U.S. Presidential Elections (*www.uselectionatlas.org*) and the "America at the Polls" books edited by Richard M. Scammon.

The first column of the spreadsheet listed the county and state like this: "Jefferson, AL," "Jefferson, AR," to distinguish between counties having the same name in different states.

In the adjacent columns, I entered how each county voted in every presidential election since 1976. (This was not as time-consuming as it sounds. I often used Excel's "fill handle" to mark all the counties in a state as Democratic or Republican, then went back and noted the exceptions, which were usually few.)

I then sorted the results and used formatting tools to color the counties according to which party carried them in the close presidential elections of 1976 and 2000. I ended up with four categories: Democratic both times (blue), Republican both times (red), Democratic then Republican (yellow), and Republican then Democratic (green).

That was a good starting point, and I used it to get a preliminary picture of the 10 regions. I added another column to the spreadsheet and assigned each county to one of the 10 regions, basing the decision not only on the voting history but also on the results of certain state elections, and geographic factors.

To come up with 10 regions that had an equal numbers of voters (as far as that was possible), I added another column to the spreadsheet that included the total number of votes cast in each county during 2000.

Using Excel, I was able to automatically add up all the votes cast in every county labeled as being part of, for example, the Appalachia region that I had defined. That way, whenever I reassigned a county from one region to another, I immediately saw how the change affected the totals for both regions. By keeping an eye on which regions were significantly larger or smaller than 10 percent of the electorate, I was able to keep redrawing the boundaries so that none of them got too large or small.

At this point, I got greedy for more data. I wanted to use demographic factors to further refine the 10 regions. So I created separate Excel spreadsheets for criteria such as age, family income, and population growth. I downloaded county-level data for all of these factors from the U.S. Census Web site (*www.census.gov*).

Smart linking

I also linked all of these new spreadsheets to my original spreadsheet through the VLOOKUP function. Thus, each row of data in my Census spreadsheets included a county name and state, and I told Excel to find the equivalent county name and state in my master data file, then plug in the right region names in the new spreadsheet.

As in the master data file, I used the SUMIF function – and simple calculation formulas – to get raw Census numbers and percentages for each region. Because all the spreadsheets were linked to the master data file I only had to reassign a county from one region to another on my original spreadsheet to change all the spreadsheets.

If I looked at the spreadsheet for educational attainment and discovered that the demographics in an Appalachian county more closely matched the Farm Belt region, I would change the county's designation in the master data file, and it would automatically change in all the other Excel files, producing new regional totals in each of them. Because of this system I didn't have to make any final decisions as to which counties were in which regions until the

continued on page 12

Bits & Bytes

continued from page 2

Baby dentistry

The IRE and NICAR Database Library worked with WCNC-Charlotte, N.C. to analyze Medicaid data for a story that aired in December. The Database Library transferred tapes of Medicaid billing records to Microsoft Access and FoxPro and identified pediatric dentists spending millions of dollars of taxpayer money on unneeded root canals in babies. The analysis found that the main metropolitan Medicaid centers in North Carolina treated fewer than 10 percent of the kids on Medicaid in the state, but performed almost half of the baby root canals. To view the story, go to *www.wcnc.com/news/topstories/stories/wcnc-121803-al-medicaid.b4d99017.html*. The series won a Scripps Howard Foundation national journalism award for excellence in electronic media, large market television/cable. For more information on how IRE and NICAR can help your news organization with data analysis, contact the Database Library at 573-884-7711.

News resources

When news breaks, journalists can find help at IRE's "In the News" Web pages. The special sections list useful databases, publications and Internet sites that can add depth to deadline reporting. When a tanker truck fell from an overpass above I-95 near Baltimore, IRE and NICAR suggested several databases from the U.S. Department of Transportation containing truck accident and hazardous material release reports. To learn more about this and other resources available, visit *www.ire.org/inthenews_archive*.

FIRST VENTURE

Principals cash in on sick pay

By Art Golab, *Chicago Sun-Times*

Elementary school principal Anthony Forst Jr. was rarely, if ever, sick in his 35-year career with the Chicago Public Schools.

But he collected $129,000 in sick pay his last year, bringing his total salary to $247,000 – significantly more than the mayor or governor makes. And it was all legal.

That $247,000 figure appeared on the top of one of the first sorts of teacher salary data *Sun-Times* education reporter Rosalind Rossi and I received from the state.

It led us to a story detailing how, under a little-known provision in their contract, Chicago teachers and principals can collect up to 315 days of unused sick pay in their last year of service, a practice unheard of in private industry.

We also found that 244 of those sick days could be added to their service time for pension purposes, a double-dip that was unheard of even in government. As a result, principals could get bigger pensions.

Rossi and I were relatively new to this sort of thing, as our paper had just lost both of our computer-assisted reporting specialists and decided to hurl us into the breach.

Fortunately, the paper sent us to an IRE and NICAR Boot Camp where we learned the tricks of the CAR trade. We were much better prepared to tackle the story than if we had tried to teach ourselves.

As a warm-up, we did the annual school report card, which involved mostly importing school test score data into Microsoft Access and formatting it for a special section. It was mind-numbing work, importing and boiling down 11,000 columns of data for 3,900 schools. But it helped hone our CAR skills.

We had to learn to select and import large chunks of delimited text files and make sure we assigned the correct heading to each column of data. Then we found ways to check its accuracy. We also learned to link the tables we had generated to tables that had been created by another reporter the year before.

We ran queries that calculated the difference in school scores between the years for several grade levels in several subjects.

In the course of all this work and while we were going back and forth with the State Board of Education on discrepancies in the data (the boot camp had prepared us well to expect this), Rossi thought to ask them if the latest teacher salary data was available.

It was, and it arrived on a CD by courier the next day, free of charge. When we finished the report card, we felt ready for anything, and decided to tackle the teacher salary data, which was technically a lot simpler.

Now the challenge was to get something interesting out of this data containing 159,000 records of the salaries of every teacher and school administrator in the state.

The teacher pay story had already been done to death.

The previous year our paper had written about the best-paid high school teachers and what they did to earn their salaries.

A *Sun-Times* investigative reporter no longer with the paper took it up a notch, doing the definitive story on school

continued on page 13

JUNE 17-20, 2004

The best in the business will gather for more than 100 panels, workshops and special presentations about covering civil rights, public safety, courts, national security, the military, business, education, local government and much more. Visit *www.ire.org/training/atlanta04* for more information and updates.

Conference Location and Hotel:
Atlanta Marriott Marquis
265 Peachtree Center Ave.
Atlanta, GA 30303

To make a hotel room reservation, call 404-521-0000 or 800-228-9290 and ask for the "IRE 2004" room block. If you'd like to make your reservation online, go to *www.marriott.com/DPP/propertyPage.asp?marshacode=ATLMQ* and use online code IREIREA. By using one of these methods, you should receive the following discounted rate:
• $119 single/double occupancy plus tax

The room block expires May 21.

Local Host: *The Atlanta Journal-Constitution*
Primary Sponsor: CNN
Sponsors: *The Wall Street Journal* • Atlanta Press Club
American University • Poynter Institute • Chicago Tribune Foundation
The Carter Center • Dart Center for Journalism & Trauma

REGISTRATION FORM.

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

Name: _____

Employer/Affiliation/School: _____

Address: _____

_____

City, State: _____ Zip: _____ – _____

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

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

To register, mail this form and a check to IRE, 138 Neff Annex, Missouri School of Journalism, Columbia, MO, 65211. To register by credit card, you must have a VISA or MasterCard. We cannot accept American Express. You may fax your credit card registration to 573-882-5431 or register online at *www.ire.org/training/atlanta04.*

Cancellations need to be sent via e-mail to jgreen@ire.org. There is a $50 processing fee for all cancellations until June 16, 2004. Refunds will not be given for cancellations after June 16.

Card Number: _____ Expiration Date: _____

Card Holder Name: _____

Card Holder Signature: _____

To attend this conference, you must be an IRE member through July 1, 2004. Memberships are nonrefundable.

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

REGISTRATION
____**$150** I would like to register for the main conference days as a ❑ Professional ❑ Academic ❑ Associate or ❑ Retiree member.
____**$100** I would like to register for the main conference days as a student.
____**$25** Late fee for registrations postmarked or faxed after May 15.

CAR DAY – optional
Thursday, June 17 *(requires additional fee)*
____**$50** Professional/Academic/Associate members
____**$35** Student members

BLUES BASH
Thursday, June 17 at 7 p.m.
____**$15** Advance tickets *(specify how many tickets)*
Tickets at the conference will be $20.
Additional information is available at *www.ire.org/training/atlanta04*

TOTAL $ _____

MAPPING IT OUT

The latest uses of mapping in news reporting.

Response times fall short

By Enric Volante,
(Tucson) Arizona Daily Star

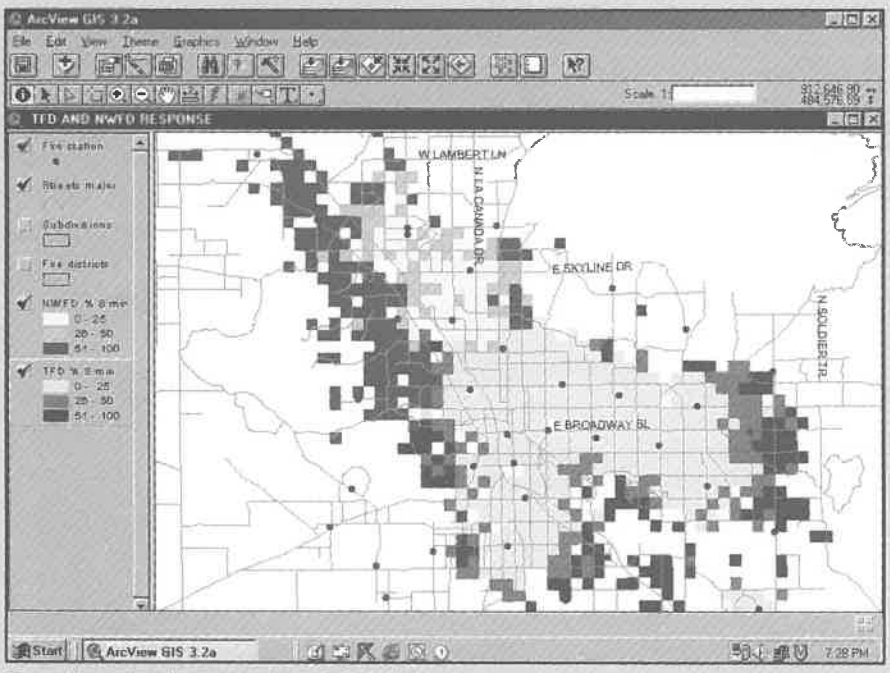
Manager Shawn Bell called for medical help when a man collapsed in a diabetic seizure one night at Lil Abner's Steak House. Then Bell waited 10 minutes for help – even though an ambulance station is one minute down the road.

No wonder Bell's interest was keen when I sat down at the steak house bar the next day and showed him a map that compared how long ambulances took to respond to emergencies in his neighborhood and others.

Thousands of people peered at that map a few days later. Last April the *Arizona Daily Star* published "Rescue Roulette," a three-day series that revealed that the Tucson Fire Department ambulances failed more than half the time to reach emergencies within the eight-minute goal recommended by a national industry group.

The stories also showed how other local ambulance operators often missed the target. And many – including the company that responded to the diabetic man in the steak house – operated under licensing requirements less stringent than Tucson's.

We analyzed emergency response data using ESRI's ArcView 3.2 geographic information system (GIS) software to find out where emergency medical workers were going and how long they took to respond.



To view this map in color see *www.nicar.org/techtips.html*

Getting dispatch data

If you want to analyze EMS response, you'll need to get the CAD, or computer-aided dispatch, database. CAD data typically includes all the information found on traditional ambulance logs, such as time dispatched and time arrived, as well as information on the incident, patient, property and vehicles that responded.

Our city serves as dispatcher for the area's largest ambulance operator, the Tucson Fire Department, and for several fire districts and private ambulance companies. I requested the city's data for all agencies under the Arizona Public Records Law.

Tucson officials balked. In addition to patient names, they refused to release addresses to which ambulances responded, saying that might identify patients. They cited medical privacy concerns and the new restrictions posed by the Health Insurance Portability and Accountability Act of 1996 (HIPAA). They also withheld all medical data, such as whether the call involved cardiac arrest or an asthma attack.

My colleague Rhonda Bodfield and I met with city officials to negotiate which fields of data should be released.

We would have liked to follow the lead of journalists at *The Hartford* (Conn.) *Courant*, who joined a database of EMS responses with a public database of deaths to identify patients who died after long waits for medical aid. (For more information see "Cardiac crapshoot" in the July/August 1999 *Uplink*.) But death certificates

are not public in Arizona, despite a fairly broad public-records law that hasn't been tinkered with much since territorial days.

So we gave up the address fields on condition that the city substitute plat numbers from a grid of quarter-mile land sections. Unable to argue that patients would be identified, the city then also gave up the medical data for each ambulance call.

Lacking names and addresses, we found patients to interview the old-fashioned way. We hit the courthouse to review civil lawsuits and we asked sources in the EMS business to name people who had waited a long time for help.

The city provided the database reluctantly and slowly, but at no cost. We ended up with a text file of 270,968 records of dispatch calls over 18 months that included fires, routine medical transports and medical emergencies.

By querying in Microsoft Access we found 92,400 were emergency ambulance calls. We deleted calls in which an ambulance was diverted to another scene. Sometimes in the heat of an emergency, drivers simply forgot to record when they arrived, so we also threw out calls with imprecise arrival times. That left 89,818 records.

Then we tabulated how long it took from the time an ambulance was dispatched to the time its crew reported arriving at the scene. We flipped the results into Excel to chart a comparison of the different agencies.

We mapped the calls in ArcView using a quarter-mile grid layer downloaded from the county transportation department's Web site, along with shapefiles of streets, fire stations and subdivisions that helped us see routes to the most populated areas. We used ArcView's legend editor to create a color-graduated grid of response times within each cell of the grid. Slower response times on the edges of the metropolitan area immediately jumped out, particularly in the northwest, where a once-rural fire district was coping with a growth boom.

But, like many parts of the country where fire departments and private ambulance companies have fought for turf, the Tucson region is a patchwork of competing providers. So we also asked the county's largest ambulance operator outside the city for its CAD database. Rural/Metro Corp. denied the request. That lack of accountability to the public became part of the story.

We knew the Arizona Department of Health Services had paper logs of ambulance calls submitted by Rural/Metro as part of its three-year license renewal. We ordered copies. Typing the data into Excel and double-checking it took three days, but it was worth it.

Unlike the city, the state had no problem releasing the patient addresses in those logs. We used ArcView's geocoding function to display about 7,500 addresses as points on a street map from the county transportation department. As we queried the data in ArcView, we saw long waits of 15 minutes or more had occurred in many places, not just distant areas. Dennis Joyce, assistant managing editor for news, had us run side-by-side maps showing different wait times.

While Arizona doesn't release death certificates, it was one of the first states to make campaign finance records available by FTP download. We found Arizona-based Rural/Metro kicks off the legislative session each year by spending close to $20,000 on a bash for lawmakers at places such as Phoenix's Ritz-Carlton hotel. As one ex-legislator told Bodfield, "I always graded the party by the shrimp – and they would have shrimp."

After we mapped the data, I visited some neighborhoods with long response times. I spoke to a few people on the street and then walked into Lil Abner's Steak House. When a cook mentioned that a man had been convulsing and kicking tables the night before while the manager phoned for an ambulance, I knew I'd brought the map to the right place.

You can read the stories online at *www.azstarnet.com/specialreports/emsresponse.html*

Contact Enric Volante by e-mail at volante@azstarnet.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: EMERGENCY RESPONSE

Sprawl outpaces private ambulances

By Jeff Sonderman, *Missouri School of Journalism*

Investigating emergency response times takes a lot of upfront reporting and data chasing before you know if you have a story. So how do you know if you should jump into doing this story in your community? There are a couple of ways.

If you notice specific cases where people died waiting for an ambulance, that's an obvious red flag. But you don't need to wait for such a clear signal.

In any community with sprawl and growth, journalists may wonder whether key public services, including emergency medical services (EMS), are keeping up. That's what sparked my interest and led to "A Matter of Time," a November 2003 story in the *Columbia Missourian*.

The story was based on the analysis of Boone County computer-aided dispatch data using ESRI's ArcView 8.2, as well as Microsoft Excel and Access.

I obtained three years of computer-aided dispatch (CAD) records of emergency medical calls, 39,744 total, from Public Safety Joint Communications, a city agency that operates the 911 service.

The agency's database manager was cooperative and told me what data was available. Because of privacy concerns under the Health Insurance Portability and Accountability Act of 1996 (HIPAA), the agency would only release the data to the hospitals, which then handed it to me on CD. The data came in a comma-delimited text format at no cost because it required no special programming.

HIPAA also challenged our desire to map response performance. The hospitals refused to provide addresses that were logged for each call because the information could easily identify patients. I talked to the database manager about alternatives, and he said the dispatch agency logs the general location of each call using a grid with cells of about one-half square mile each. He e-mailed me an ArcView shapefile of the geographic grid system that day.

The response data was relatively clean; 911 system computers, not people, had entered much of the data automatically. The biggest problem was that response times were not properly logged on some calls. The response time is recorded when the dispatcher presses a button at the time of dispatch and again at the arrival on scene. In some cases, the dispatcher missed one of those events, and the response time was recorded as zero. These records were excluded because the errors seemed to affect calls in a random pattern that did not skew the data. This anomaly, as well as all of my analysis, was explained to the readers in a sidebar.

We imported and analyzed the data into Access and exported the results of some of the queries in Excel.

The analysis included looking at overall response time, as well as grouping in Access to see response times by ambulance, specific location, city or town, and time of day. I looked at the division of call volume among various ambulances and locations and times of day.

To analyze performance relative to the eight- and six-minute response time standards of the hospitals, I created two fields to indicate whether the unit had reached the scene within each of those two time frames. I also created new fields to display the response time for each call in seconds, based upon the start and end times included in the data.

In ArcView, I used the grid to map responses by grid section, which produced a color-coded mosaic showing the percentage of calls in each sector that were reached within the eight-minute goal. To do this, I first used an Access query to group by the coordinates for each grid sector. I then summed the number of calls reached within the goal and counted the total number of calls in the area. I imported the result of that query into Excel to figure out the percentage of successful response times for each grid sector. Then I imported that simple table – just grid coordinates and the percentage of success – into ArcView and applied it to the shapefile that used the same grid coordinates.

Talking to the people who kept the data also helped us to extract useful data from two of the fields in the database.

The grid coordinate field of each record corresponded to a single cell on the grid map, but it also contained a letter at the end of the four or five numbers (the "B" in section 14041B, for example) that indicated the general location of the call. Each of the smaller cities and towns in the county had its own assigned letter. I parsed the grid coordinate field in Access and placed its letter into a new field so I could easily examine all the calls in those towns using grouping and counting queries.

The index field that assigns a unique identifier to each dispatch also included characters that indicate which ambulance responded, for example "M31" for Medic 31. We parsed those references into a new field, which allowed us to look at response times and emergency call volume for each ambulance.

The two ambulance services that cover Boone County are owned and operated by the two major local hospitals. The good news for residents is that they don't pay a tax for a public ambulance service. But this also means the ambulance services are departments of large institutions that have to watch their financial bottom lines.

Emergency medical services tend to lose money, and that puts the opera-

tors in a sticky position. How much money are they willing to lose on an ambulance service to provide better patient care?

This is an especially important issue in small but rapidly growing regions, such as Columbia and Boone County, which has a population of about 135,000 people, with about two-thirds of them in the city of Columbia. But Columbia is sprawling outward, away from the ambulance stations, and populations are growing in other cities around the county.

Once I had the story, it was time to make it come alive with field reporting. I went on several ambulance ride-alongs, arranged by one of the hospitals, and shadowed the paramedics when they responded to calls.

This firsthand experience helped me understand the system and the numbers I was writing about. It also gave me extended access to paramedics in their work environment, where they were talkative and comfortable. The time I spent on ride-alongs was incredibly useful. The paramedics are on the front lines of EMS, so they often provided the most insightful comments on the data I showed them.

Paramedics can also help answer the big question: How much is better patient care worth? Paramedics almost always want more ambulances and more resources. They would take an ambulance on every street corner if they could get it, because their biggest concern is patient care. Here you might find, if not a whistleblower, at least an advocate for change in the system.

Contrast that with the concerns of managers and hospital administrators. Many of them have never worked as paramedics, and their focus is more on the financial costs of adding ambulances or other resources.

So in most EMS systems, you will probably find some amount of tension between what the EMS workers want and what the administrators think is necessary or practical. Each side will be eager to convince you, and you can use information from both sides to develop a balanced story.

Contact Jeff Sonderman by e-mail at jeffsonderman@yahoo.com.

readme.txt

Here are some tips for using computer-aided dispatch data for analyzing EMS response from Enric Volante of the (Tucson) *Arizona Daily Star*.

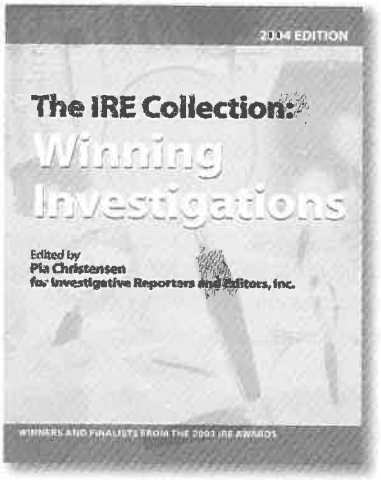
• Don't rely on average or median response times. They tend to mislead. Compare the percentage of ambulance runs that meet a set time goal, such as 90 percent of emergencies within eight minutes. Use a shorter time for first responders. Check the guidelines of national firefighting groups for different types of EMS calls.

• Pick your target. An increasing number of fire departments send fire trucks with paramedics or emergency medical technicians to stabilize a patient until an ambulance arrives to take the patient to a hospital. Are you going to map runs by ambulances or by first responders? Your choice may depend on the level of aid first responders are equipped to give and how many fire trucks in your city have them.

• Request only the data you need. CAD databases can be big and complex. Talk with the person who knows the data and review the record layout so you understand each field before you submit a request.

• Borrow a dispatcher's manual. It explains the process for prioritizing each emergency based upon the medical symptoms reported to the dispatcher.

COMING IN MAY!



Get the scoop on the year's best investigations

The latest compilation book from IRE not only salutes the winners of the prestigious IRE Awards, it offers testimony from the winners.

Information is shared on how the work was completed, which techniquest proved most valuable and what resources were used to win the print, broadcast, online and other categories. A great guide for launching projects in your own back yard.

Watch the IRE Web site for information on ordering.

www.ire.org

Results

continued from page 3

closing days of the project. I could keep tweaking the regions as I discovered new data.

Mapping regions

I turned to ArcView late in the project, and I did not use the mapping software to make any new calculations. I began by using a blank map that showed boundaries for all of all the counties in the United States, a map that came with ArcView. That map includes a data (or "attributes") file that includes standardized codes (known as FIPS codes) assigned to each county.

The 10 Regions of US Politics

Our regions are based on voting returns from both national and state elections, demographic data from the US Census, and certain geographic features such as mountain ranges and coastlines. **Each region represents about one-tenth of the national electorate, casting between 10.4 million and 10.8 million votes in the 2000 presidential election.**

I then added the FIPS codes to the master data file, which I saved as a dBASE file in order to import it into ArcView. Using a method in ArcView similar to Excel's VLOOKUP function, I then linked the dBASE file to the attributes file in such a way that all of the counties were automatically assigned to the proper regions. Then I created a legend that assigned a different color to each region, and I used the "unique value" function to produce, with just a couple of keystrokes, a map that resembles the final "Beyond Red & Blue" map.

I wasn't quite done, however, because the ArcView map exposed certain aesthetic flaws, such as regions with fingers and tentacles made up of sparsely populated counties. I then went back and made some final adjustments to the master data file, making the regions look a bit more coherent. I decided that those odd shapes and extensions are justified only when significant numbers of voters are involved. (For example, I pushed the El Norte region up to include San Antonio's Bexar County because it's a major urban area with some 400,000 voters. I did not expand the Big River region to include several Democratic-leaning counties on the Upper Peninsula of Michigan because they include too few voters to be much of a factor in statewide elections.) After I made these final corrections to the master data file, I saved it as a dBASE file and imported it into ArcView.

To create separate ArcView maps for each region, I simply copied the national map 10 times. Each time, I changed the legend so that only one region appeared up in color, while the other nine were left invisible. I also imported other ArcView data files to add state lines and major cities.

The project took a couple of months while I worked on regular duties. The finished product, posted on our Web site in mid-December, includes close-up maps for each region and a set of tables showing the political and demographic characteristics for each.

Next comes what I consider to be the fun part. When the results come in for the 2004 election, I hope to redraw the "Beyond Red & Blue" map to reflect any changes in voting patterns. And all of my old data will be automatically updated to reflect these changes, with very little effort on my part – leaving more time for writing.

Contact Robert David Sullivan by e-mail at rsullivan@massinc.org.

© 2003 Massachusetts Institute for a New Commonwealth

Principals

continued from page 4

superintendent pay. He revealed that top-paid superintendents were getting retirement incentives that had the potential to cripple the taxpayer-backed state pension fund.

Our competition, the *Chicago Tribune*, borrowed that angle and applied it to elementary teachers.

So Rossi suggested we look at principals.

Our data had the name, salary, position and district information for all school employees in the state, but the state board had omitted the names of their schools.

We asked them to provide another disk with the school name information, which they did, and forged on, running simple queries for salaries of high school, junior high and elementary principals. To do that we selected the codes used for each position, which were provided in a separate text file that also contained the record layout.

The top ten ranged from $179,000 to $247,000, which was definitely very interesting. But the question arose, how do they make the money? Most were at or close to retirement age. Was it just another case of retirement incentives inflating pay?

The raw numbers only gave us total salaries.

We had to get the breakdowns the hard way, by calling each district. We made our jobs easier by focusing only on the top ten salaries.

But the district superintendents and business managers at first took advantage of the fact that we didn't know which questions to ask. They gladly broke out retirement incentives because they could make a case that this saved them money by enabling them to hire less-experienced, lower-paid replacements.

But accumulated sick pay, vacation pay and other benefits that we had no idea existed all were lumped into one category.

Gradually, by comparing notes between districts, we learned to ask for more specific breakdowns from everyone and kept calling until we got them.

Getting the numbers is sometimes the easiest part of the story.

As a result, the Microsoft Excel spreadsheet we used to keep track of the salary breakdowns kept growing wider.

Three facts quickly emerged. All of these principals had retired, and in many cases retirement incentives had pushed them into the top pay brackets.

But only suburban principals got the retirement incentives.

Chicago principals got a different going away present. They were allowed to cash out nearly a year of sick days and apply a large portion of those days to their service time, thereby boosting their pensions.

We still weren't sure this was a story.

To put it in perspective, we called school districts in New York and Los Angeles. After wading through their complex retirement packages (nothing is simple in educational bureaucracies) we learned they were far less generous with accumulated sick pay than Chicago.

And a compensation expert at the Illinois Chamber of Commerce told us such benefits were unheard of in private industry.

So we finally had a story.

School and union officials defended the system, saying it provides an incentive for teachers (who eventually become principals) to show up every day and thus saves money spent on substitute teachers. It also rewards Chicago teachers who stick it out until retirement and don't make as much money as suburban teachers.

And it serves as a *de facto* retirement incentive, allowing quicker replacement of top-scale teachers and principals with lower-paid beginners.

But other experts pointed out that the Chicago teachers and principals also get another perk, that few, if any, in private industry get – a pension that pays 75 percent of their base retirement salary.

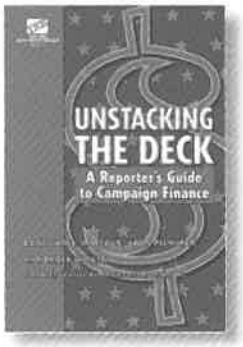
This project confirmed what Rossi and I had learned in the CAR boot camp. Getting the numbers is sometimes the easiest part of the story. Figuring out what they mean, putting them in context and then bringing those numbers to life is the hard part.

Contact Art Golab by e-mail at agolab@suntimes.com.

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

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

CAR TOOL

askSam tames the data clutter

By Bonnie Britt, *Freelance*

I worked in a newsroom where a clutter-averse assistant managing editor periodically pushed a big trash barrel through the aisles expecting reporters to toss documents. Packrats like me did not share his enthusiasm for a tidier newsroom.

Then I learned about askSam, an archiving and free-form database program that allows journalists to squirrel away information electronically – instead of letting it pile up on their desks. The program is a natural friend to journalists, historians, archivists, investigators, and packrats of every stripe.

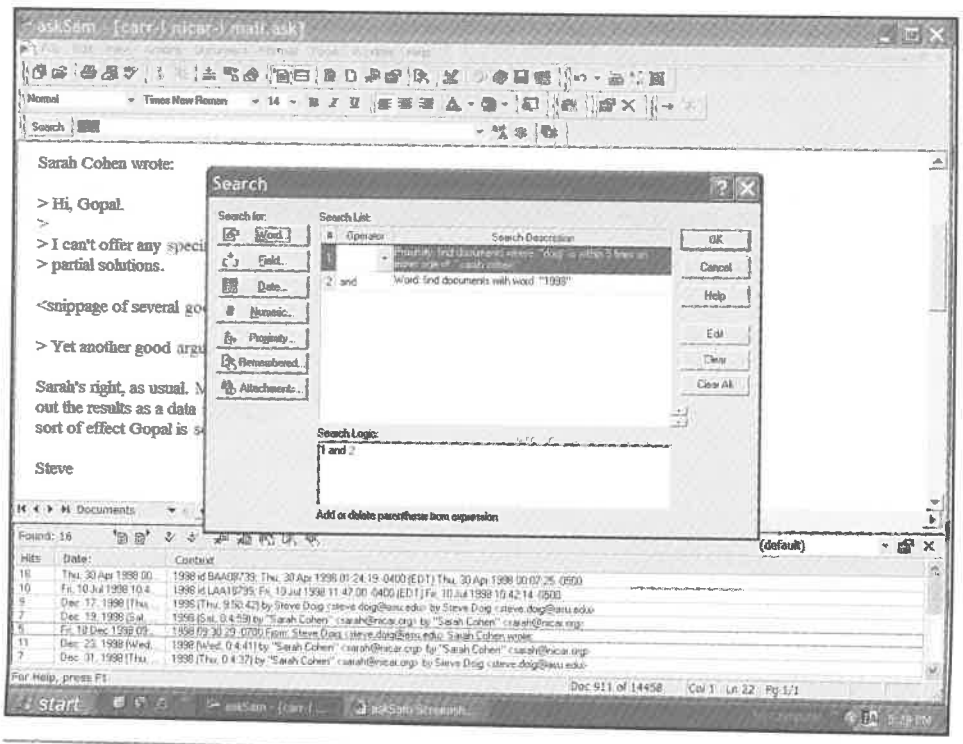
Using askSam is better than living with paper clutter because nothing gets lost or buried. And the program does not force you to shoehorn data into fields with set lengths. In fact, you can choose whether or not to import data into fields.

You can search, sort, and sift massive amounts of data (up to 16 terabytes), structured or unstructured, from an array of formats. Its speed and flexibility explain why askSam is popular among intelligence and law enforcement agencies whose personnel often think and work like journalists.

Using it

The U.S. General Accounting Office uses it with other software to collect and organize survey information. Senate investigators pursuing the Iran-contra scandal employed askSam's lightning speed to good advantage in the mid-1980's to ask informed questions of Oliver North in real time. The program enabled them to organize, catalog and present the volumes of evidence they had gathered during their investigation and subsequent Senate hearings.

A forensic document analyst at the U.S. Postal Inspection Service built a central repository in askSam of articles relating to death threats sent through the mail, forged signatures, counterfeit passports and related topics.



While askSam is valuable to people who toss unstructured notes into files, then do simple or Boolean searches for combinations of words or phrases, it also offers detailed reports based on field data queries. For example, you can set askSam to automatically recognize e-mail fields such as, To, From and Subject, or you can import e-mail without specifying any field. The database is searchable either way, but if you import e-mail with fields, you can ask for a report on all e-mail, say, from Ken Lay between two dates when the subject contains a particular word or phrase. You can search dates, fields, proximity, or explore numeric values and relationships, or do a global search and replace.

You can search through multiple askSam databases to keep track of projects or to compare statements of public officials over time. The program is easy to use once you've figured out what you want from it, and it's smart enough to adapt for use with Microsoft Access, Excel or LexisNexis.

Why would anyone want to import a spreadsheet into askSam? To more easily view and page through rows; to view it differently so as to spot errors; or to query it differently. Viewing a long row of data in Excel is easier to read when laid out on one page. Each succeeding row can be viewed, page by page, by pressing F9 for next (or ALT-F9 for previous.)

Stories captured from LexisNexis can be searched in a variety of ways, such as for combinations of names, alongside Web pages, stories cut and pasted from today's news, annotated screen shots, and other tidbits gathered and thrown into the mix.

The search results are displayed in a split window that is easy to navigate. The results show the number of hits per document, context, and the first line of

data. Click on any row to see the entire corresponding document.

Getting started

To begin using askSam, you open an empty database, name it and save it. You can immediately begin building a database in any of four ways:

• Drop documents into askSam by cutting and pasting them as plain text or, using the "paste special" option, as HTML or rich text. (You add graphics separately.)

• Import all or some documents, one file type at a time, from a Windows folder. You can opt to import e-mail attachments.

• Link any file in its native format. Linked files retain their formats and can be opened in askSam and edited.

• Attach a file in any format. An attached file becomes a part of the askSam database.

You can customize askSam databases using fonts, colors and other formatting. HTML and rich text documents will retain the formatting of the original files; text documents will reflect your choices.

You can also customize each database to control how dates are displayed and what search results are shown.

The biggest complaint of longtime askSam users is that table columns in HTML and Word do not align properly. Users have found a workaround: pasting table text into askSam to facilitate searching, and then attaching the original document to the database so it can be viewed with columnar data properly aligned.

The ability to share askSam databases with colleagues is a new feature. Last year, the company, based in Perry, Fla., began offering a free viewer, which is similar in function to the Adobe Acrobat Reader. Colleagues can use the free viewer to read and search – but not add to, or modify — existing databases.

The standard askSam program retails for $149 and is powerful enough for importing e-mail, or keeping track of the elements of a nonfiction book or news project. For power users who want more speed and flexibility, there is a professional version, which includes full-text indexing, for $395 retail. Indexing speeds searches of big databases. askSam Professional also includes an ActiveX control that can be accessed from Visual Basic and other programming languages to create customized applications.

The program's features include a tool to pack databases to keep them lean. The program also allows document deletion and creation, and remembers which documents and searches you've marked for recall.

For more about the program, visit *www.asksam.com*, where you can also download a 30-day evaluation version.

Contact Bonnie Britt by e-mail at editor@editorsforum.org.

readme.txt

CAR Tools focuses on software that's useful to journalists doing computer-assisted reporting. Recent issues of *Uplink* have taken a look at:

- ClusterSeer analytic mapping tool. (January/February 2004)
- SAS System Viewer (November/December 2003)
- ParseRat text parser (September/October 2003)
- OS X operating system (May/June 2003)
- Rematch extension for ArcView 3.x (March/April 2003)

If you have a favorite CAR tool that *Uplink* readers should know about, drop us a line at uplink@nicar.org.

Better Watchdog Workshops

Investigative Reporting on the Beat

Investigative Reporters and Editors Inc. and the Society of Professional Journalists, with funding from the Sigma Delta Chi Foundation, have joined forces to offer a series of workshops focused on doing investigative reporting while covering a beat.

The workshops, specifically for journalists at small- to medium-sized news organizations and those in bureaus of larger organizations, will emphasize the use of freedom-of-information laws and address juggling a beat while producing investigative and enterprise pieces.

"You'll learn enough in the first 15 minutes to keep you busy for a month."
Kevin McGrath, *The Wichita Eagle*

Workshop is scheduled for:
May 1-2, 2004 – Muncie, Ind.
May 22-23, 2004 – New Orleans
Nov. 6-7, 2004 – Oklahoma City, Okla.

For more information, visit
www.ire.org/training/betterwatchdog

To request a workshop for your area, contact Executive Director Brant Houston at **watchdog@ire.org**.

OPEN SOURCE

Xandros: Switching from Windows

By Aron Pilhofer, *Center for Public Integrity*

Mydoom was the last straw.

It was Tuesday, Jan. 27, just past 9 p.m. I was staring at dozens of Microsoft Outlook e-mail messages infected with copies of the Mydoom virus, trying to figure out how, despite security precautions bordering on paranoia, one of those little buggers had managed to infect my Windows XP system.

I was firewalled and patched. My virus software was up to date. I had never used the Outlook preview pane, read suspect messages or opened unknown attachments. Yet, I had the virus.

It was then and there that I decided I was done with Windows. I was going to make the switch to Linux as my main desktop once and for all.

I have been working with Linux for years now – first as a newsroom intranet and data server, and then as a secondary operating system on my home computer. But even though I had adopted open source software for much of my CAR-related work, I never did make the permanent switch to a Linux desktop.

A few years ago, I was close to the brink of sanity trying to deal with yet another utterly inexplicable Windows ME crash and burn. But then Microsoft released the truly excellent Windows XP, which was fast, easy and (most importantly) stable.

Windows XP, however, isn't secure and I'm tired of having to hold my breath every time I hit the "send/receive" button in Outlook. A properly maintained Linux box, on the other hand, is just as stable as XP and virtually immune to the kinds of attacks plaguing Windows users. I have considered other options, such as Apple OS X (which is based on a core operating system similar to Linux). But even though I love the software, I am not willing to pay a premium for new hardware.

Thus, my transition to Linux began.

Which distribution?

There are literally thousands of versions of Linux, some are well known but most are fairly obscure. Some are produced by corporations, such as Mandrake, SUSE or Red Hat. Others, such as Debian, Slackware or Gentoo, are designed by enthusiasts as a labor of love.

Xandros lives up to its promise and then some.

I had experimented over the years with most of the versions I've mentioned, so I was pretty familiar with their respective strengths and weaknesses. Because I was going to be using Linux for both home and work, I had some fairly specific criteria in mind. I needed a distribution that is:

- Easy to administer and maintain
- Able to run a wide range of software, both open source and proprietary
- Well-documented and well-maintained

My first choice was Red Hat (See a review in the September/October 2003 *Uplink*), but the company recently decided to focus on corporate desktops, and handed over development of its desktop distribution (now called Fedora) to volunteers.

I tried Fedora, but the installation froze – not a good sign. I then read about a distribution called Xandros, which was designed to appeal to former Windows users. The deluxe version even comes packaged with software that allows users to install and use Microsoft Office and other Windows applications.

Best yet, Xandros is based on the Debian version of Linux, one of the oldest, most respected and reliable versions available. It also features a slick software management tool called apt. I paid $89 for the deluxe version and downloaded the installation discs and documentation. A few hours later, I was on my way.

Installation

Xandros has one of the smoothest installations available. Xandros recognized nearly all my computer hardware during the installation – something that was a first in all my previous operating system installations.

I wanted to add Xandros while preserving both my existing Windows installation and another version of Linux I had already put on the machine. I also had just added some fairly cutting edge hardware, which can cause installation problems for any operating system, but especially Linux.

To my surprise and delight, Xandros handled my hardware without a hiccup.

My New Home

Booting Xandros for the first time revealed a Linux environment that has been tweaked as much as possible to look and feel like Windows. Double clicking opens a directory or launches a program. Xandros automatically loads additional hard drives or partitions – including Windows partitions – into a top-level directory named "disks."

Xandros goes to great lengths to hide its Linux underpinnings. When booting into or shutting down from Xandros, you see a graphic, and not the typical stream of gobbledy-gook messages Linux usually spews onto the screen.

The last, and biggest, difference in the Xandros deluxe version is the inclusion of Crossover Office, which is a set of utilities that trick Windows applications into running under Linux. They are optimized, as the name suggests, to work best with Microsoft Office, and my copies of Microsoft Access, Excel and PowerPoint 2000 installed flawlessly.

One criticism of the desktop is that it does not look as smooth, crisp and clean as Red Hat or Windows XP.

In very limited testing, I can report that all three seem to work nearly as well under Linux as they do under Windows. The only obvious differences are aesthetic. Menus seem a little darker under Linux, and the menu fonts seem smaller.

One criticism of the desktop is that it does not look as smooth, crisp and clean as Red Hat or Windows XP. Fonts can appear blocky, but that is a minor gripe.

Simplicity

In terms of software management, Xandros is outstanding. The distribution has a very simple, but powerful, package manager. It can be used to install updates from the Xandros Web site, or add new software from any number of sources, including standard Debian repositories on the Internet. Think of it as the Windows Update and Add/Remove Software utilities all in one.

The manager operates in simple and expert modes. In simple mode, users can install only the software that has been packaged for the Xandros installation as well as any security or software updates the company makes available on its Web site. This is probably more than enough for most users, but there are a few software packages (PostgreSQL, Webmin) that were excluded from the standard installation.

Switching into expert mode provides access to software stored in repositories across the Internet, and the ability to add more of your own.

Installing an application is usually a simple matter of a single mouse click. Not only does Xandros download the application, but also it automatically handles the problem of figuring out what other programs or libraries are required and downloading those as well.

Most of the popular Linux distributions use a package management system based on one developed by Red Hat a few years ago (the Red Hat Package Management system, or RPM). RPM, to be blunt, stinks. Installing applications other than those packaged with the distribution under RPM can be a long, frustrating, and often-fruitless experience.

Xandros's apt-based package manager is amazing. I cherry-picked more than a dozen non-standard applications – a recipe for disaster for RPM-based distributions. Xandros dutifully and flawlessly installed each one without a hitch. That is impressive.

Final Thoughts

Xandros lives up to its promise and then some. It is Linux for those who want to wade rather than dive into a new operating system.

Some Linux purists will not like the modifications Xandros has made. But for folks who don't come from a Unix background, and don't have a computer science degree, Xandros is just possibly the best of both worlds.

Contact Aron Pilhofer by e-mail at apilhofer@publicintegrity.org.

Computer - Assisted Reporting

BOOT CAMPS

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

- **May 16-21**
- **Aug. 1-6**

Editor Boot Camp

Special three-day workshop tailored to the needs of newsroom managers in both print and broadcast.

- **April 23-25**

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

MOVING?

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

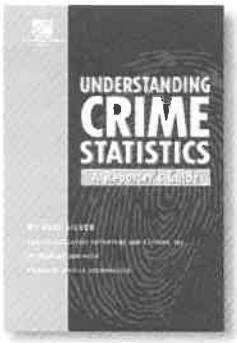
Drop us a note at
jgreen@ire.org

Investigative Reporters and Editors, Inc.

REQUIRED READING
For Your Newsroom

Understanding Crime Statistics
A Reporter's Guide, by Kurt Silver

Know the data and reports available to add perspective and authority to your crime reporting. Learn how to use the data and avoid common pitfalls while writing responsible crime statistics stories.



ORDER TODAY!
Call 573-882-3364 with your VISA or MasterCard
-OR-
Visit our Web site at www.ire.org/store
for online ordering or order form downloads
IRE MEMBERS: $15 each • NONMEMBERS: $25 each

Drugs

continued from page 1

made public for the first time, provided key information for the reporting.

First, the paper records.

I obtained nearly 3,000 drug overdose autopsy case files, most of them from 2001 and 2002, from Florida's 24 medical examiner districts. I examined each file, looking for the name of the treating physician of the deceased. Some files contained "medication inventory" sheets, which logged all of the pills found at the death scene with the name of the doctor, drug, the pharmacy that filled the prescription and the date filled, but others didn't.

I typed the names of victims, doctors and drugs into a Microsoft Access database on my laptop. It was often as frustrating as it was time consuming. In many cases, pills were found at the death scene in bottles, but nobody recorded the name of the doctor who had prescribed them, for instance.

After visiting morgues for several months my death file grew and a pattern emerged: each examiner's district had a handful of doctors with abnormally high numbers of deaths that had gone unnoticed.

I wound up with more than 40 doctors with four or more patient deaths for 2001 and 2002. Sixteen doctors had prescribed pills found at the scenes of eight deaths or more, mostly in that same two-year period. Some had deaths spanning several medical examiner districts.

Nobody had done this type of research before, even though state medical licensing officials believe a pattern of overdose deaths in a doctor's medical practice warrants investigation. Nor had anyone matched doctors with high numbers of patient deaths against Medicaid prescription drug billing data. I got the chance to do that after Florida Attorney General Charlie Crist agreed to make the billing records public at no cost.

After some negotiation, Medicaid fraud unit investigators sent me a Microsoft Access database on CD-ROM. It contained data for 8.8 million Medicaid pharmacy claims, which amounted to more than $346 million in drug billings for 2000-2002.

The database contains two tables, one that was 1.49 GB, the other slightly smaller. The data included sales of all "controlled" drugs – 844 varieties of narcotics and other medications. Each drug has a five-character code that serves as a unique identifier. The drugs range from brand name painkillers like OxyContin that cost Medicaid more than $7 a pill to generic tranquilizers and sleeping pills that cost just pennies per dose.

I spent many hours rummaging through this huge data set.

The database contains fields for the name of the prescribing doctor, license number of doctor, year of the prescription, county of the doctor, amount paid, number of Medicaid recipients treated and the pharmacy that billed for the drugs.

Initial queries using Access showed fewer than 3 percent of the state's doctors were prescribing two-thirds of the narcotics dispensed to people on Medicaid. Five doctors had prescribed $2 million or more worth of narcotics over three years, much of it OxyContin.

I spent many hours rummaging through this huge data set, doing queries and sorting results to find other patterns. Scrolling through, I spotted the name of a doctor I knew

Tech tip...

Curl-ing apart Web documents

By Derek Willis, *Center for Public Integrity*

Here's a pet peeve you might recognize: you find a database you need on the Web, but can't download it with a single click. Instead, the records are contained on separate pages, sometimes behind a search form that lets you see only a few entries at a time. Hitting Control-S several hundred times isn't my idea of a great day.

But there's good news: there are tools that can pry loose all that data without much trouble or effort on your part. My favorite is curl, a command-line utility for Windows, Unix, Linux and Mac OS X. With curl, I can download batches of files with a consistent naming pattern, such as those pesky .ASP or .CGI files that contain numbers in the URL string.

Curl is a free command-line utility, meaning that you run it from the terminal in Linux, Unix or Mac OS X, or from the DOS prompt in Windows. If your system doesn't have curl installed, you can get it from *http://curl.haxx.se*. Be sure to bookmark curl's basic documentation (*http://curl.sourceforge.net/docs/manpage.html*); it comes in handy.

Here's an example of using curl: a few months ago, Jeff South, a Virginia Commonwealth University professor and former database editor at the *Austin American-Statesman*, posted to the NICAR-L listserv asking if anyone had an easier way to put together a list of California lobbyists.

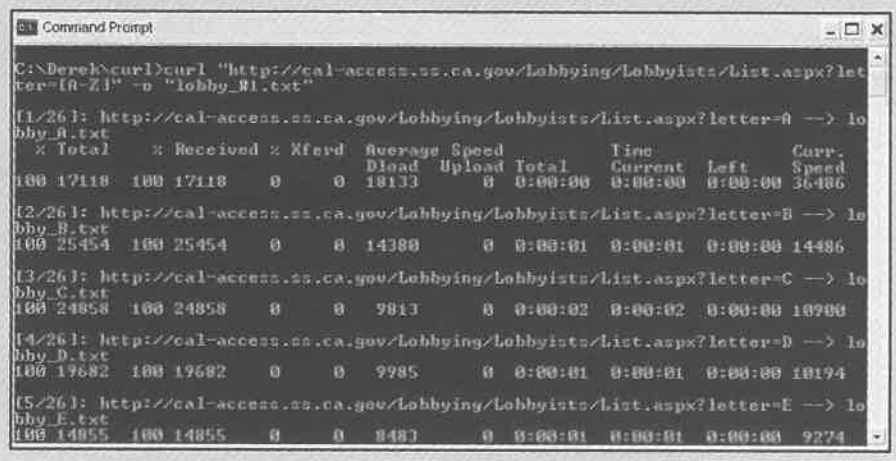
The official state Web site (*http://cal-access.ss.ca.gov/Lobbying*) has a list of current lobbyists but only shows one page for each letter of the alphabet, meaning that you'd have to save each one and then put them together before beginning the task of separating the records from the HTML around them.

Curl was built for just this kind of task, since it offers powerful ways to automate downloads based on a consistent pattern. The URL for lobbyists with a last name beginning with "A" is *http://cal-access.ss.ca.gov/Lobbying/Lobbyists/list.aspx?letter=A*, and the only character that changes from one file to the next is the final capital letter.

Curl has the ability to iterate through a range of letters or numbers, so I wrote the following command:

```
Curl "http://cal-access.ss.ca.gov/Lobbying/Lobbyists/list.aspx?letter=[A-Z]" -o "lobby_#1.txt"
```

This command tells curl to begin with the initial URL ending in A, download each one, and save it to



Curl in action: At top, the Curl code for grabbing Pennsylvania charities data. The bottom window shows the Curl code for getting data about California lobbyists.

Principals

continued from page 4

superintendent pay. He revealed that top-paid superintendents were getting retirement incentives that had the potential to cripple the taxpayer-backed state pension fund.

Our competition, the *Chicago Tribune*, borrowed that angle and applied it to elementary teachers.

So Rossi suggested we look at principals.

Our data had the name, salary, position and district information for all school employees in the state, but the state board had omitted the names of their schools.

We asked them to provide another disk with the school name information, which they did, and forged on, running simple queries for salaries of high school, junior high and elementary principals. To do that we selected the codes used for each position, which were provided in a separate text file that also contained the record layout.

The top ten ranged from $179,000 to $247,000, which was definitely very interesting. But the question arose, how do they make the money? Most were at or close to retirement age. Was it just another case of retirement incentives inflating pay?

The raw numbers only gave us total salaries.

We had to get the breakdowns the hard way, by calling each district. We made our jobs easier by focusing only on the top ten salaries.

But the district superintendents and business managers at first took advantage of the fact that we didn't know which questions to ask. They gladly broke out retirement incentives because they could make a case that this saved them money by enabling them to hire less-experienced, lower-paid replacements.

But accumulated sick pay, vacation pay and other benefits that we had no idea existed all were lumped into one category.

Gradually, by comparing notes between districts, we learned to ask for more specific breakdowns from everyone and kept calling until we got them.

Getting the numbers is sometimes the easiest part of the story.

As a result, the Microsoft Excel spreadsheet we used to keep track of the salary breakdowns kept growing wider.

Three facts quickly emerged. All of these principals had retired, and in many cases retirement incentives had pushed them into the top pay brackets.

But only suburban principals got the retirement incentives.

Chicago principals got a different going away present. They were allowed to cash out nearly a year of sick days and apply a large portion of those days to their service time, thereby boosting their pensions.

We still weren't sure this was a story.

To put it in perspective, we called school districts in New York and Los Angeles. After wading through their complex retirement packages (nothing is simple in educational bureaucracies) we learned they were far less generous with accumulated sick pay than Chicago.

And a compensation expert at the Illinois Chamber of Commerce told us such benefits were unheard of in private industry.

So we finally had a story.

School and union officials defended the system, saying it provides an incentive for teachers (who eventually become principals) to show up every day and thus saves money spent on substitute teachers. It also rewards Chicago teachers who stick it out until retirement and don't make as much money as suburban teachers.

And it serves as a *de facto* retirement incentive, allowing quicker replacement of top-scale teachers and principals with lower-paid beginners.

But other experts pointed out that the Chicago teachers and principals also get another perk, that few, if any, in private industry get – a pension that pays 75 percent of their base retirement salary.

This project confirmed what Rossi and I had learned in the CAR boot camp. Getting the numbers is sometimes the easiest part of the story. Figuring out what they mean, putting them in context and then bringing those numbers to life is the hard part.

Contact Art Golab by e-mail at agolab@suntimes.com.

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

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

had died years ago and that gave me the idea to see if others who were deceased or had their licenses revoked were still generating pharmacy bills. They were.

When I matched the top billing pharmacies with prescribing doctors, I noticed that a small group of pharmacies had derived most of its Medicaid billings from prescriptions written by one or two doctors, a situation state officials later said raised the possibility of kickback schemes.

But the most significant finding came from checking the top prescribing doctors against the overdose database I created from autopsy and police reports.

It turned out that many, though not all, of the highest Medicaid prescribing doctors were the same ones with abnormally high numbers of patient deaths. For example, 16 of the top prescribers had a combined 61 patient overdoses resulting in death in two years.

Many of these doctors also had been disciplined previously by the state for offenses that included drug trafficking and trading drugs for sex.

I used the Florida Department of Health's online physician profiles and old binders with documents summarizing disciplinary actions taken against doctors. The profiles include a category for "deceased," and helped to identify dead doctors who were still generating bills. The profiles and the CD order form can be found on the Web at *www.doh.state.fl.us/MQA.*

I also used the Internet-based Medicare and Medicaid exclusion list and compared by hand every Florida doctor who had been kicked out of Medicaid but still generated billings. (See *http://exclusions.oig.hhs.gov/search.html.*)

Another data set that proved invaluable came from the Florida Agency for Health Care Administration, which oversees Florida's $12 billion-a-year program. I e-mailed a list of the codes for all of the controlled drugs to the agency's Medicaid pharmacy manager, and he had the state's Medicaid contractor provide at no cost a Microsoft Excel spreadsheet with the drug, doses purchased, cost of drug and number of units for 2000 through September 2003. This data showed that dosage levels of certain drugs, including Xanax bars and methadone, had soared.

A project like this can be done in states that grant access to autopsy reports, though getting the Medicaid data might take some tough negotiating and even a lawsuit.

It's worth the expense and effort. Prescription drug overdose deaths are soaring nationwide and in many states authorities believe Medicaid helps fuel the illegal street market in these pills. States facing huge cost increases in Medicaid simply can't afford to allow fraud schemes to go unchecked.

Here are some tips for taking the reporting beyond the data analysis:

• Contact the families of people who die. While many won't welcome a call from a reporter, others will help you get documentation such as pharmacy receipts showing the dates victims were prescribed pills and the amounts.

• Seek the advice of legitimate pain management doctors and pharmacists at their offices or stores. Many will be helpful.

• Whenever possible, review the complete file from a medical examiner's death investigation. Many times an investigator's notes will have the phone number or comments from the doctor who prescribed the drugs. Phone numbers also are critical in making sure you have the right physician.

Fred Schulte can be contacted by e-mail at fschulte@sun-sentinel.com.

readme.txt

Knight Ridder's investigative team recently reported on the dangers of doctors prescribing drugs for treatments that haven't been approved by the U.S. Food and Drug Administration. See the March/April 2004 *IRE Journal* for more details.

Reporter Chris Adams used the FDA's Adverse Event Reporting System, available from the IRE and NICAR Database Library, in the investigation. Adams said that the database was useful because he could query it by the type of drug and by the physician's reason for the prescription. The database includes information about adverse reactions to prescription drugs.

Adams recommended allotting plenty of time to work with the data and to seek help as needed from organizations that use the data regularly.

In some cases, a doctor and the drug producer independently file reports about an adverse reaction, Adams said, so it's important to check for duplicate records.

Adams supplemented his analysis of the AERS data with prescription data provided by private companies.

For more information about the database see *www.ire.org/datalibrary/databases/AERS.*

Finding

continued from page 1

statewide tool that is supposed to help locate missing children.

The investigation began in July, with just a simple search of data made available on the Internet through the state's Web site. A link to the Indiana State Police Web site led to the Clearinghouse list, containing the names of 467 children supposedly missing 30 days or more. The list, which was in fixed-width format, was easily imported into a Microsoft Excel spreadsheet for analysis. Four columns of information contained the names of the missing children, their birthdays, the dates they were reported missing and the agencies handling their cases.

First, I sorted the data and kept only the records involving agencies within our circulation area, placing those in a separate spreadsheet. I used Microsoft Access to count the number of cases handled by each law enforcement agency in the state.

That search showed two of the major cities in our readership area – Hammond and Gary – led the list in terms of the number of children reported missing by single agencies. At the very least, we could run a story showing that Hammond, with 56 cases, and Gary, with 60 cases, each had nearly twice as many children on the list as the state's largest city of Indianapolis, which had 32 cases.

Next, I created new columns in the statewide and regional spreadsheets and wrote formulas to calculate the number of years or months each child had been listed as missing. I also calculated their current ages and the ages at the time they disappeared. The results showed cases of children who had been on the list for a number of years, including a Hammond boy listed as missing for 10 years following an alleged abduction in a custody dispute.

After looking at the data, two lines of reporting emerged: one on why cities in our region topped the list and the other on the stories of children who had been missing for a long time.

Those areas of focus changed slightly following my first interview with the supervisor of the Hammond Police Department's juvenile division. During the interview, the police sergeant agreed to track down more information on the cases of Leslie Richardson, reportedly missing for 10 years, and of Talion Mansfield, who had been listed as missing for two years. I was also told that if I wanted copies of the original missing children's reports filed by parents or guardians, I would have to file an open records request through the city's legal department.

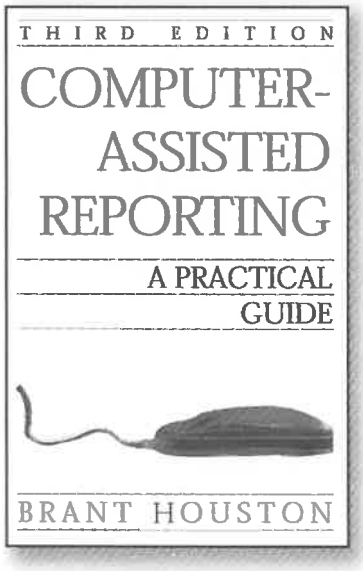
After the interview in Hammond I was driving back to our main office in Munster when my cell phone rang. The sergeant was on the other end and he explained to me that after making a couple of phone calls in the Richardson case, he determined that the boy – reported missing at the age of 4 and now 14 – was no longer missing. It appeared, the sergeant said, that the boy's parents had never called back to report that the custody dispute leading to the report had been resolved. The sergeant added that parents have the responsibility to contact the authorities once children are found and police merely take the reports.

Shortly after I returned to the office my cell phone rang again. This time the sergeant reported that the other case I had inquired about involved a child who also was no longer missing and had not been for quite some time. Again, the sergeant blamed the original complainant on the report for failing to close out the case with the police.

At this point, red flags regarding the list's accuracy and effectiveness began waving furiously. In a groundbreaking series of stories that ran nearly 20 years ago, *The Denver Post* debunked inflated claims national child abduction totals. (For a copy of the Pulitzer Prize-winning stories contact the IRE Resource Center at 573-882-3364 and ask for Story No. 1061).

I filed public record requests for more than 100 missing children reports from Hammond and Gary and within three

COMPUTER-ASSISTED REPORTING



A Practical Guide, by Brant Houston

New Edition!

The most practical guide to the skills journalists need to succeed has been updated to expand on the highly regarded basic skills instruction and offer more advice on how to write better news stories.

ORDER NOW

By phone: Call 573-882-3364 with your VISA or MasterCard

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

By mail: Send check to IRE, 138 Neff Annex, Missouri School of Journalism, Columbia, MO 65211

Plus postage:
Media rate – $4 for the first book, $2 for each additional book

$32 MEMBERS • $39 NONMEMBERS

www.ire.org/carbook

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

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

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

Uplink

Uplink Info

A newsletter of the National Institute for Computer-Assisted Reporting

Editor
Brant Houston
brant@ire.org

Managing Editor
David Herzog
dherzog@nicar.org

Asst. Managing Editor
Jeff Porter
jeff@nicar.org

Art Director
Lisa Triefenbach

Copy Editor
Pia Christensen

Contributing Editors
Christina Caron
Megan Clarke
Jaimi Dowdell
Holly Hacker
Andrea Lorenz
Tim Ragones

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

Director of Publications
Len Bruzzese
len@ire.org

Advertising Coordinator
Pia Christensen
pia@ire.org

Subscription Administrator
John Green
jgreen@ire.org

Subscriptions
IRE members $40, nonmembers $60

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

Postmaster: Please send address changes to IRE-NICAR.

weeks began sifting through each one, gleaning contact numbers for parents and calling them one by one. I used people-finding tools on the Web to track down some of the parents who had moved or whose phone lines had been disconnected.

That helped me find another 14 cases from the region involving children who had been listed as missing for years or months but who had actually only been missing for days or hours. In each case, parents said that the police had never contacted them for follow up after the parents filed their initial report. And in six of these cases, parents said they had called the police to tell them their children had been located.

I also called parents whose children remained missing. Those parents reported that police had also failed to follow up in the cases of their children. We even followed one woman who walked the streets of Chicago's South Side, tacking posters of her daughter to utility poles because she was frustrated the police did nothing to help locate her daughter.

Combing through all of the original reports helped identify other angles for the series. Runaways accounted for most of the cases on the list. That became the main angle for part two, which outlined the problems and profiled three chronic runaway children. For part three, we focused on a Hammond woman whose daughter had been abducted by her ex-husband and taken to Egypt in 1993, never to return.

The impact of the story was felt almost immediately. Before the series even ended, local state legislators called upon the governor and the state police for a probe of the Clearinghouse list and its inaccuracies. The matter remains under review, and a northwest Indiana congressman and a state lawmaker are considering legislation that would require police agencies to regularly follow up on missing children's cases.

Contact Marc Chase by e-mail at mchase@nwitimes.com.

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: David Donald, ddonald@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