# Uplink

www.nicar.org

## IN THE NEWS

# Mine safety stories

By Brian Hamman
*IRE and NICAR*

In the days and weeks after the Sago Mine disaster in West Virginia, journalists used their computer-assisted reporting skills to dig more deeply into stories about mine safety.

The U.S. Mine Safety and Health Administration (MSHA) offers more than 20 years of accident, inspection, and violation data on its Web site that journalists can use to add context to a quick-hit story or as the basis of an investigation.

When they set out to study the violations at Colorado's mines, Katy Human and Jeff Roberts of *The Denver Post* found they had to begin with a more basic question: "Our first question was, do we have underground coal mines in Colorado?" Human says.

To get an overview of mining in Colorado, Human and Roberts turned to the data retrieval system provided by MSHA, available at *www.msha.gov/drs/drshome.htm.*

## SPOTLIGHT: MONEY & POLITICS

# Data lifts lid on DeLay money

By Jonathan D. Salant, *Bloomberg News*

Until he stepped down as U.S. House majority leader in September and then announced in January that he would not attempt to reclaim his position, Rep. Tom DeLay was arguably the most powerful member of the House of Representatives.

Even after he was rebuked three times in 2004 by the House ethics committee, DeLay had no trouble winning re-election to the No. 2 position in the House Republican leadership.

Only one member of the conference, Christopher Shays of Connecticut, called on DeLay to step down in the face of continuing ethical questions, including whether lobbyist Jack Abramoff

## SPOTLIGHT:

For more about Money & Politics see:
• Changes in campaign finance playing field, p. 8
• Network analysis of city's power lines, p. 10
• Data and resources for reporting, p. 11
• The March-April 2006 *IRE Journal*

## MEYER AWARD WINNER

# Advanced GIS analysis targets lost wetlands

By Matthew Waite, *St. Petersburg Times*

I hadn't been working in the St. Petersburg newsroom for more than a month when Craig Pittman, the environment reporter at the *Times*, said he had an idea he wanted to run past me. He wanted to measure the wetlands that had been destroyed in the past 15 years, but had no data and no idea how to do it. I told him it would probably require an analysis of satellite imagery.

"Can you do it?" he asked.

"Probably," I said, lying through my teeth. I had no idea if I could do it.

Eventually, I figured it out, and the answer to Pittman's question was not so simple. Between 1990 and 2003 – a time when the federal government pledged to stem the destruction of wetlands – about 84,000 acres of Florida's wetlands were replaced by urban uses; houses, schools, gas sta-

# Bits & Bytes

## New staff member

Megan Means (formerly Christensen) has rejoined the staff at IRE and NICAR as an assistant database library administrator. After working as a graduate research assistant in the Database Library from 2001 to 2003, Means spent a year at the *Detroit Free Press* through the Knight Ridder specialty development internship program, as a metro reporter and CAR specialist on the investigative team. Major projects included a mapping analysis of childhood lead poisoning in Detroit and a review of the Ford Crown Victoria's accident history. Most recently, she covered the K-12 education beat for the *Columbia* (Mo.) *Daily Tribune*. She is working toward completing her master's degree at the Missouri School of Journalism.

## Campus crime

ABC News Primetime reported in November that the University of Massachusetts had the highest rate of violent on-campus crime among the largest universities in the United States, based on an analysis of data by the IRE and NICAR Database Library.

The analysis also found that forcible sexual assault was reported more frequently than any other type of violent crime.

**IRE and NICAR analyzed 2002-03 Campus Crime Statistics data for the report. The analysis separated universities and colleges into four groups based on enrollment.**

The federal Jeanne Clery Disclosure of Campus Security Policy

## INSIDE NICAR

# Yes, we still say 'computer-assisted reporting'

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

A contributor to Poynteronline recently asked "Does Anyone Still Say 'Computer-Assisted Reporting'?" in a brief essay that suggested CAR has gotten as mundane as using the telephone to gather news.

Well, it hasn't and plenty of journalists still utter the words "computer-assisted reporting" when they are talking about analyzing data using spreadsheets, database managers and mapping programs for their news stories.

It's true that many journalists are tapping into Web tools to enhance their reporting, making good use of sites such as Google, reverse-lookup directories and the local assessor's online real estate tax rolls.

But many journalists know CAR means more than effectively scouring the Web for information. A journalist visiting the U.S. Mine Safety and Health Administration Web site would not learn that the agency lets many of the major fines assessed against mine operators go unpaid. But journalists who have obtained MSHA's enforcement data and analyzed it using a database manager discovered that newsworthy piece of information, among others.

That's what we still call "computer-assisted reporting."

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

## About our contributors ...

**Liz Carey** is a reporter for the Community Press family of newspapers based in Cincinnati. Previously, she was a network administrator for General Electric and a marketing director for OneNet Communications, a Cincinnati-based Internet Service Provider. She is a 2005 alumna of the Computer-Assisted Reporting Boot Camp.

**Brad Heath** is an enterprise reporter and CAR specialist at *The Detroit News*. Before that, he was an investigative reporter for the *Press & Sun-Bulletin* in Binghamton, N.Y.

**Gordon Hickey** is the editor of the projects team at the *Richmond Times-Dispatch*. He has been a reporter and editor for more than 25 years, working for *The Richmond News Leader, Milwaukee Sentinel* and the *Milwaukee Journal*.

**Aaron Kessler** is former deputy director of the Virginia Public Access Project, which tracks money in Virginia politics. He joined the *Richmond Times-Dispatch* on a freelance basis to perform the social network analysis for the "Who Runs Richmond" project. He is a graduate of the Missouri School of Journalism and a former researcher for the IRE Resource Center.

**Sarah Okeson** works at the *Journal Star* in Peoria, Ill., where she is a general assignment reporter. Her reporting has included covering local Army reservists in Saudi Arabia. She is an alumna of the 2004 Mapping Boot Camp.

**Jonathan D. Salant** covers money and politics for Bloomberg News. He is co-author of the "Follow the Money Handbook," published by the Center for Responsive Politics. He is a 1998 Boot Camp alumnus.

**Matthew Waite** is a reporter for the *St. Petersburg Times* who specializes in using CAR and geographic information systems (GIS). Waite is a 1999 Boot Camp alumnus.

**Derek Willis** is research database editor at *The Washington Post*, and runs The Scoop, a Weblog on investigative and computer-assisted reporting. He has worked for The Center for Public Integrity and *Congressional Quarterly*. Willis is co-author of the IRE Beat Book "Unstacking the Deck: A Reporter's Guide to Campaign Finance." He is an alumnus of the 1996 CAR Boot Camp and 1997 Intermediate Database Boot Camp.

# 2006 IRE CONFERENCE · JUNE 15-18

The best in the business will gather for panels, workshops and special presentations about covering public safety, courts, national security, the military, business, education, local government and much more. Visit www.ire.org/training/dallasfortworth06 for more information and updates.

## Conference Hotel:
Renaissance Worthington Hotel, 200 Main St., Fort Worth, TX 76102

## Hotel Reservations:
To get the discounted rate of $146 plus tax, please make your hotel reservation by Monday, May 15. There are two ways to make a reservation:
**Online:** Visit https://marriott.com/reservation/availability.mi?propertyCode=DFWDT&mktc and be sure to enter this group code: INVINVA
**Telephone:** Call 800-468-3571 or 817-870-1000 and ask for the Investigative Reporters and Editors room block.

If you have hotel or general conference questons, please contact Ev Ruch-Graham, sr. conference coordinator, ev@ire.org or 573-882-8969. If you have registration questions, please contact John Green, membership coordinator, jgreen@ire.org or 573-882-2772.

## Host: The Dallas Morning News

## Primary Sponsor: Fort Worth Star-Telegram

**Sponsors:** Sunbelt Communications Company and Jim and Beverly Rogers, Chicago Tribune Foundation, The Poynter Institute for Media Studies Inc., National Press Foundation, Haynes & Boone LLP

**NOTICE:** You will be listed on the IRE Web site as an attendee. If you don't want to be listed, please e-mail John Green at jgreen@ire.org.

# REGISTRATION
You can register for this conference online at **www.ire.org/training/dallas/fortworth06** or by filling out this form and faxing it to **573-882-5431**. To register by credit card, you must have a VISA, MasterCard or American Express. If you are paying by check, please mail this form with your check to IRE, 138 Neff Annex, Missouri School of Journalism, Columbia, MO 65211. 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):

Card Number:      Exp. Date:

Card Holder Name:

Card Holder Signature:

Cancellations need to be sent via e-mail to jgreen@ire.org. There is a $50 processing fee for all cancellations until June 14, 2006. Refunds will not be given for cancellations after June 14.

To attend this conference, you must be an IRE member through July 1, 2006. Memberships are nonrefundable.

Pre-registration will close on May 31. Early-bird registration closes May 24.

**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** (includes online audio access for all taped panels post-conference):
____ **$165** 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.

**CAR DAY — optional:** Thursday, June 15
(requires additional fee)
____ **$50** Professional/Academic/Associate members
____ **$35** Student members

**BLUES BASH TICKETS:** Thursday, June 15 at 8 p.m. Advance tickets are $20 (beer, wine and soft drinks included). Tickets on site, if available, will cost more. Enter total number of tickets needed ___ x $20 = ___

**RECEPTION:** Sixth Floor Museum in Dallas, Friday, June 16 (Advance ticket required, 2 ticket max., only 350 tickets available) Enter total number of tickets _____ x $20 = _____

**SPECIAL NEEDS:**
❑ Do you require special assistance?
Specify
❑ I would like a vegetarian meal at the IRE Awards luncheon. (Must request when registering.)

**TOTAL $**

## Bits & Bytes

and Campus Crime Statistics Act requires colleges and universities to release campus crime and security procedures each year. The Database Library has the 2001-2003 data, which is available to journalists for a processing fee.

The dataset includes incidents of liquor and drug violations and possession of illegal weapons. A separate table tracks hate crimes, in which victims were picked out because of their race, gender, religion, sexual orientation, ethnicity or disability.

For more information about how the Database Library can perform analysis for your news organization, contact Jeff Porter, director of the Database Library, at 573-882-1982 or jeff@ire.org.

### Truck inspections

The condition of trucks, crossing our nation's highways by the millions, is a critical matter for protecting the public's safety. The Database Library has updated the U.S. Department of Transportation truck inspection dataset through 2004.

Journalists can review national or state truck inspection violations, by company or material carried to identify dangers to the public. Previous stories have delved into careless enforcement of regulations, poor safety records of trucking companies and lobbying by trucking companies to gain support for bigger, heavier trucks.

---

### FIRST VENTURE

# Finding surprises in drug arrest files

By Liz Carey, *Community Press (Cincinnati)*

Two of hardest things for me to wrap my mind around at the IRE and NICAR Computer-Assisted Reporting Boot Camp were how to get the data I wanted, and what to do with it.

Data analysis is more than taking apart a spreadsheet, or pawing through a database. Data analysis, we learned, is the middle step of a story. You cannot base a story on the data alone, and you cannot get the data unless you have the contacts to get it.

All of this became clear to me when I returned from the March 2005 Boot Camp and was chomping at the bit to find a story to use my newfound skills.

I bought the National Inventory of Dams database for my state from IRE and NICAR Database Library and searched through the IRE Web site for anything I could make a story out of, certain I could blow the lid off something in my Northern Kentucky beat, despite being at a weekly newspaper.

Little did I know, the story was right beneath my feet.

As part of my regular beat, I regularly wrote up police reports in the small towns I covered, transcribing them from paper to my laptop. I began to notice an increasing number of arrests on drug-related charges.
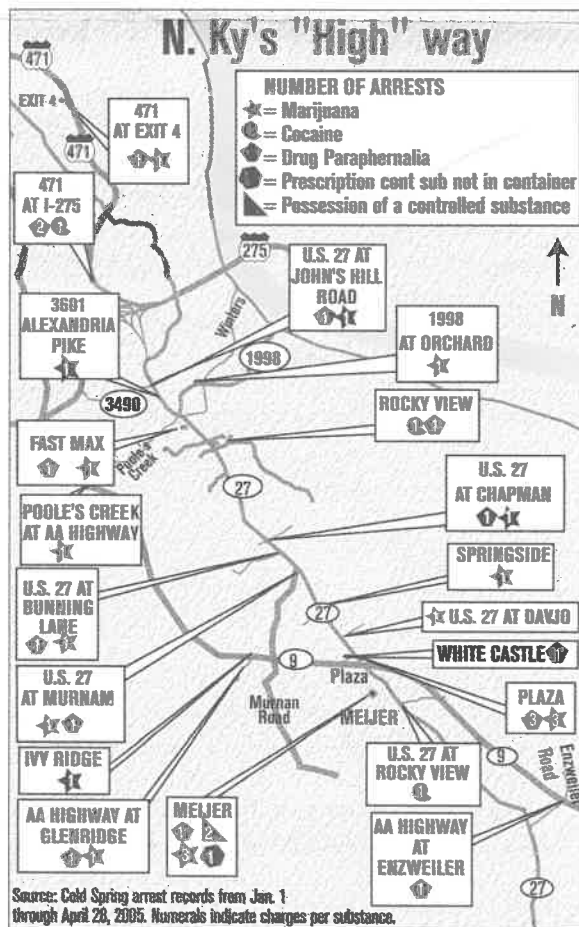
Cold Spring, Ky., was an example of this. Cold Spring, a city of just more than 4,000, was named the fastest growing city in Kentucky in 2004. The four-lane highway U.S. 27 runs through the middle of

the city, connecting the eastern half of Northern Kentucky to I-275, the bypass that circles Greater Cincinnati.

The city is hardly an urban center, boasting a generally white, upper-middle class population, with home values in the upper-$250,000 range.

I was noticing more frequently that regular traffic stops in Cold Spring were turning into drug arrests. Additionally, shoplifting charges in the four major grocery stores were increasing and, more often than not, the things being stolen could be used in methadone production.

I decided to take closer look. I didn't find a smoking gun, but an existing trend that was being investigated by the police. It



N. Ky's "High" way

Source: Cold Spring arrest records from Jan. 1 through April 20, 2005. Numerals indicate charges per substance.

---

is evidence of the importance of source relationships.

We started by requesting all of the police reports since January 2000. At first, the request was met with a modicum of disbelief. In short, the department clerk told me I was free to go through all 5,000-plus handwritten reports and take from them whatever information I wanted.

But the police chief told us that, since 2003, those records were kept in a database, and he would be willing to give them to us on CD.

That data dump took more than an hour, not in processing time, but in negotiation over each data category and whether or not it should be divulged. The process meant sitting with the chief and discussing each category in depth, sometimes giving in, sometimes citing public record law. There was no charge for the data.

Some categories of data were easily determined such as age, offense, date of offense and location of offense. Information about name, address, arresting officer and Social Security number had to be thoroughly discussed to determine how essential the information was to our story, and whether or not it violated anyone's privacy. In the end, we were able to get everything, except Social Security numbers, in a comma-delimited text file which we imported into a Microsoft Excel spreadsheet.

I work in a Mac-based newsroom and Excel was readily available. In addition, I worked with Excel in the corporate world before becoming a reporter and it was easy for me to work with the data that way and turn over a story quickly, instead of relearning the Structured Query Language also covered in Boot Camp.

The result was a detailed report of every arrest or citation in the previous three years that was easily sorted and manipulated for analysis.

That analysis started with breaking down information by year. Each year's arrests and citations were then categorized by type of citation. All those not falling into a drug category were eliminated from the table. Then, data was sorted per citation.

Each citation was sorted by name of offender. Offenders were checked against one another to make sure that they were not multiple charges on the same citation. Additionally, citations were sorted according to ZIP code to determine where the biggest number of offenders was coming from.

The data proved that most offenders lived in the county, contrary to law enforcement professionals' belief that the majority of the offenders were coming from outside the county.

Data also revealed offenders came more from rural areas of the county,

# Covering Congress, states getting trickier

By Derek Willis, *The Washington Post*

Using computer-assisted reporting to cover congressional elections used to be so much simpler.

The Federal Election Commission's FTP site was the first stop, where pretty much all of the campaign finance data you would ever need resided.

Now, thanks to 527 political committees, state PACs and the 2002 McCain-Feingold campaign finance law, things are more complicated.

It is no longer enough to limit your data gathering and analysis to the two or more candidates in a given congressional or statewide race. Do that in 2006 and you could miss the big story, or at least trail the rest of the pack on it.

While the FEC is still the starting point for federal races, it is not the finish line. Those using CAR for elections should assess the contests they will be covering to see whether the old boundaries between federal and state races still apply. If your state has a hot race, then you will need to know how to gather data from multiple sources and integrate it into a single system.

Consider the election for Virginia's attorney general in 2005 (Virginia is one of a few states that have significant off-year elections). While Republican and Democratic candidates raised millions for this statewide post, the GOP candidate got more than a third of his money from a single source: the Republican State Leadership Committee. Because the RSLC is a 527 organization reporting to the Internal Revenue Service, Virginia regulators and journalists knew only that it had given a bunch of money to a state candidate. They did not know where that money came from and, thanks to the IRS reporting schedule, they would not find out until after the election.

Elections in 2006 will see more frequent disclosure from 527s, which follow a reporting schedule roughly equivalent to the FEC in even-numbered years. And like the FEC, the IRS releases its data in bulk — meaning that CAR folks will have to download the entire set of 527 electronic filings and then find the committees they are after. If you have never done this before, a little practice wouldn't hurt. Unlike the FEC, dealing with the IRS data doesn't require any reformatting of fields (something you can avoid by purchasing FEC data from the IRE and NICAR Database Library), but it does lump original and amended filings in the same compressed file.

Updates are available every Sunday at *http://eforms.irs.gov*, and The Center for Public Integrity also has a regularly updated site that provides downloadable contributions and expenditures for 527 committees (*www.publicintegrity.org/527*).

In federal races, some advertising activity by 527 committees must be reported to the FEC, thanks to the McCain-Feingold law. These expenditures, known as "electioneering communications," occur when an organization not registered with the FEC airs a television or radio ad that mentions or identifies a federal candidate within 30 days of a primary election or 60 days of the general election. Such ads fall under the scope of the law only if they are targeted to a candidate's constituency; if the Sierra Club runs an ad about a New Jersey candidate in Arizona, it would not meet the criteria.

The good news about electioneering communication reports is that they include very specific information about ad buys, including the stations and amounts involved, and reports are due within 24 hours of the ad's initial appearance. The

bad news is that those reports are filed by fax only. For journalists outside Washington, D.C., that means waiting on the FEC to scan and post images of those faxes, which sometimes are difficult to read. It also means that there is no database of such activity, unless you decide to build one (The Center for Public Integrity did so for the 2004 elections).

This year, state political parties will face their first real test since McCain-Feingold, since many of them received an artificial boost during the 2004 presidential campaign. State parties now find themselves having to raise and spend mostly federal or "hard" money, since even the most generic party activity is presumed to have some benefit to House and Senate candidates on the ballot. So if your state has a looser reporting schedule, you may find that the state parties' federal accounts file more often and contain the bulk of the money raised and spent. In some cases, journalists will find more details in the federal data than is required on the state level. Beware of duplicative reports — reconciling state and federal data can be a tricky proposition. When in doubt, call the treasurer or the party official in charge of filing the reports for an explanation.

In most respects, the FEC hasn't changed its data-handling process much; the data it makes available via FTP is updated every Monday morning (usually before 5 a.m., for those interested in running Scheduled Tasks in Windows or cron jobs on other operating systems. These files contain all of the contributions associated with a two-year election cycle (the current cycle began the day after the 2004 election).

There are ways, however, to avoid working with the entire set if you are only focusing on a handful of committees. All FEC committees, except for Senate candidates and the two parties' senatorial fundraising arms, file electronic reports to the FEC, and those filings can be downloaded and used to create databases. The advantages of doing so are two: timeliness (you can access a filing as soon as it appears in the FEC's system) and the presence of expenditures, which the FEC does not keypunch or release in

bulk form but are available on a filing-by-filing basis.

Expenditures are easily the least-covered aspect of campaign finance, and sometimes the most important. They can show the tactics used by campaigns (direct-mail fundraising or big, flashy events?) or provide home addresses for campaign workers. Particularly interesting are the expenditures of candidates who face little or no opposition – what do they spend their money on?

Such reporting can yield great stories, but if you go this route, remember that you'll need to keep track of the filings, because amendments appear often. I've found that it's useful to track filings themselves, and you'll probably want to do the same if you plan to maintain the latest and greatest version of a committee's data.

The electronic filings end with the file extension ".fec", but they are delimited text files that can be imported into Microsoft Excel or other programs. But you'll want to be familiar with the layouts and keep in mind that each filing contains all of the transactions associated with that report: contributions, expenditures and summary figures, all stacked on top of each other.

A campaign finance junkie would do well to keep checking the FEC's FTP site regularly, however; the agency often will place custom collections of data there that might prove useful.

One last tip, which has been a personal interest of mine: don't forget the metadata. In campaign finance, the metadata means things like when a committee reports, how many amendments it files or even the length of the reports. Such details are worth tracking because they can provide insight into the operations of the campaign. Multiple amendments filed on the same day are a sure sign that either the campaign's accountant found a mistake or the FEC's auditors did. Either circumstance is worth a look, and may yield a good story.

Contact Derek Willis by e-mail at willisd@washpost.com.

**OPEN SOURCE**

# Simplifying database full-text searches

By Brad Heath, *The Detroit News*

One of the most important things you can do in your newsroom is spread the wealth of information. But finding a way to let other reporters and editors get the data they need, whenever they need it, is a challenge. The easiest way to do it is on your company's intranet – but most of the tools are cumbersome or expensive.

Here is my solution.

OpenInfo is a collection of intranet scripts you can use to run simple, full-text searches across dozens of newsroom databases simultaneously. You can add data to the site with a few minutes of work, instead of spending hours writing new search forms. You can customize it to help people in your newsroom find information most relevant to them, with easy searches and automatic cross-referencing of records. Users who want to go deeper will find forms they can use to run advanced searches or write data queries (without having to know Structured Query Language).

Best of all, it runs entirely on open-source platforms (an Apache Web server, a MySQL database server and the PHP scripting language). If you already have the hardware, you can get the whole package set up in an afternoon, at no cost. And because it relies on MySQL to do the heavy lifting, it can easily accommodate searches of millions of records. (I have used it on one server with almost 70 million records spread across 60 tables.)

Download the OpenInfo file from *www.nicar.org/downloads*.

Once you provide OpenInfo with basic information about what kinds of data are stored in your table, it does the rest: building search forms, a query wizard and integrating the records into a global full-text search.

Once you have the scripts set up (more on that in a minute), searching is very straightforward. The home page offers a single form that will search across every database that you incorporate into OpenInfo, using syntax similar to what you are accustomed to using for searching the Web or your electronic archives. (Figure 1) Because the search runs against a full-text index, you don't need to know the order in which your search words appear, or whether they're in the same field in the database.

In addition to simple searches, OpenInfo gives you an almost unlimited ability to drill into and cross-reference your data. Take, for example, my own voter registration information. To find it, I'd start by searching for Heath, Brad* (yes, OpenInfo supports wildcards), then click on the matching results to see the full record. (Figure 2). OpenInfo shows all the fields incorporated into the database (unless you have chosen to hide a few), and makes the contents of each field clickable. Clicking any of the links will take you to a pop-up guide that will help you cross-reference that information against other OpenInfo records, Google, or any other online database with an accessible URL syntax (Yahoo! News, your archives, your photo assignments, etc.).

If you want, OpenInfo can also do some of this work for you. For example, you can teach OpenInfo how to assemble the fields in your database to make a complete address. If you do, the script displays an embedded map for each record (as well as linking to Google Maps), and gives you tools for cross-referencing the address against other records. It will also automatically check for other records in that table with a matching address and show you a list. That means when you look at my voter registration record, OpenInfo automatically finds my wife's registration, too. (Figure 3). You also can have OpenInfo mash
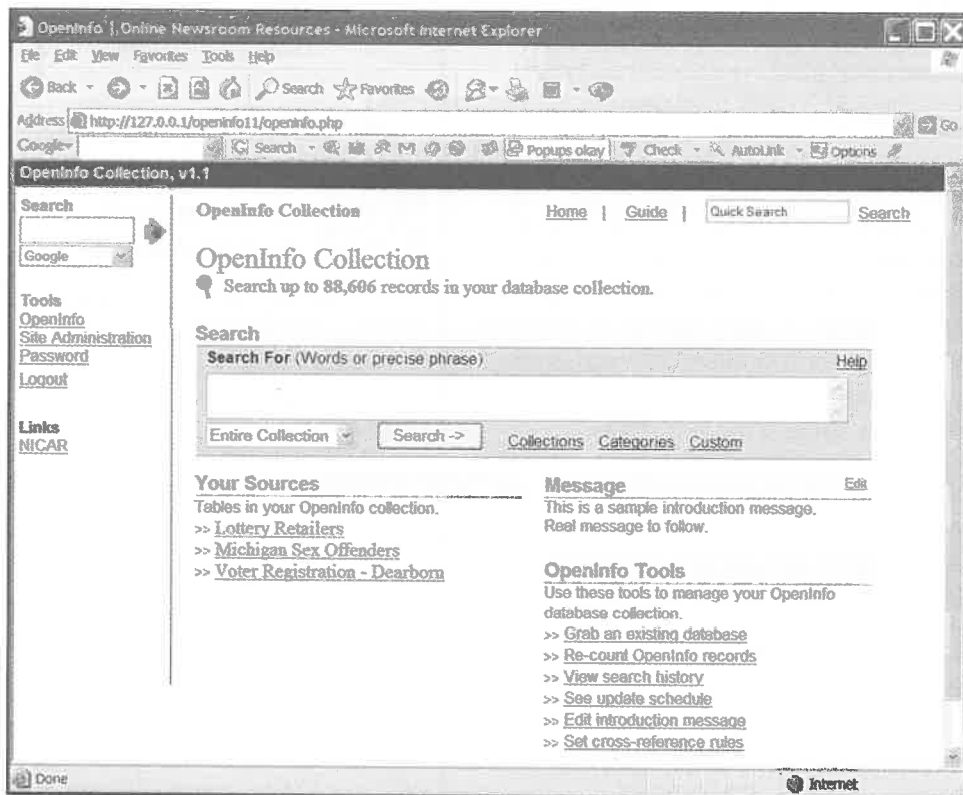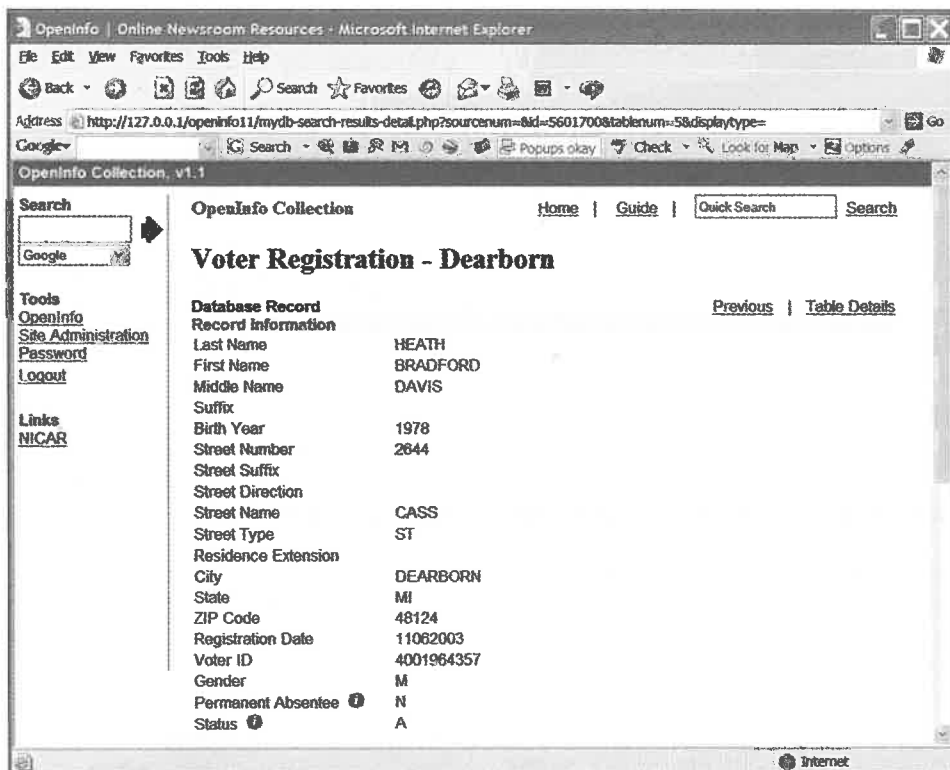


Figure 1

Figure 2

together other fields from the record that you think might be helpful for users looking to find related information. For example, it might be helpful to have one link that will search based on a person's first name, last name, city and state.

For users who want more power, OpenInfo has more advanced search functions. You can search individual fields in a database, or use a built-in query grid to search through or summarize your data and export the results to Microsoft Excel (users who want to do even more can download full data tables as dBASE files). Users can create custom groups of sources, or search based on the type of information associated with a field, such as names and addresses. If you want, you can allow users to create their own

databases on the fly, for their own use or to add to the public collection of sources. OpenInfo tracks searches and record views so you can find out quickly what is most popular with your newsroom.

Setup is easy and inexpensive. To start, you need a MySQL database server, an Apache Web server and the PHP scripting language, all of which are open source. (If you don't already have these, you can add them fairly easily by downloading XAMPP, which installs all three at once, from *http://sourceforge.net/projects/xampp*.) After that, unpack the script files into one of the folders on your server. Then sign on to your database server as a root user and run the setup.sql file included with the scripts. (You might need to make a few other adjustments to your server, depending on how it's configured. A complete readme file is included with the scripts.)

Once the scripts are configured, you need to load your data tables into MySQL. (Because it's open source, you can find a variety of free tools to help with this; one is at
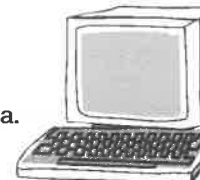


Figure 3

# Wetlands

tions, malls, etc. That figure formed the backbone of an investigation that revealed how poorly the U.S. Army Corps of Engineers has enforced the Clean Water Act.

What did it take to do it? Two college courses, $5,000 in software, 220 GB of hard drive space, editors with an abundance of patience and faith and about 10 months of free time.

If I haven't scared you off yet, and you think you might try your hand at satellite imagery analysis, here's how we handled the problem of detecting change with imagery.

The first step in any satellite imagery analysis is to remember Philip Meyer, the father of pricison journalism and a professor at the University of North Carolina. Meyer has been telling us that we should use the scientific method for journalism. With satellite imagery analysis, it is a survival method. You need to clearly understand your problem and thoroughly

understand your data before starting your analysis.

Why? Lots of reasons. If you don't understand what you are after, you could spend months and money analyzing data, only to come up with nothing. If you misunderstand your question, you might buy the wrong data. That's right: More often than not, you are going to be buying data. Some of it is expensive. You do not want to spend hundreds of dollars on Landsat imagery only to realize you needed to spend thousands on SPOT data.

To help you understand your question, you need to understand what imagery can show you. For instance, Landsat data has a resolution of 30 meters – one pixel in the image equals 30 meters on the ground. If you're looking for individual houses, it's the wrong data. If you are looking at how whole subdivisions have changed the landscape, you're in the right place.

You also need to know what data you need in your imagery. Imagery can contain just visible light – three bands, red, green and blue – or it could contain slices of the infrared spectrum. You need to know what the bands can tell you and which bands you need for your analysis.

It's harder than it sounds, and could fill an *Uplink* article by itself.

It might seem esoteric to talk about being sure you understand your own question, but it's not. It's critical. I learned the hard way. We originally set out to find how many acres of wetlands had disappeared between 1990 and 2003. After months of work, I had an answer, but it was tens of thousands of acres beyond what any of our reporting had found. So I looked at what we had – a map showing where wetlands had increased and decreased between 1990 and 2003. In a few minutes, I found the problem – tides. In one map, the tide was out, in another, the tide was in. So thousands of acres of saltwater marshes were gone, but not really. Other changes were time differences – one image was taken during a wetter month than another. All over Florida, natural changes – rains, tides, etc. – affect the size of wetlands.

Our question was too simplistic. What we were really interested in was how many acres of wetlands man had destroyed permanently.

More on that in a minute.

Obtaining data is both easier and more difficult in satellite imagery analysis. It is easier because there are only so many places you can get it. It is harder because the ways in which imagery is stored are a little arcane and, most of the time, costs money.

Before you buy data, though, ask around. You might be surprised who has some and who might be willing to share. I found a complete 2003 dataset through a state agency, which made it free to me under Florida's public records law. I found a complete Landsat dataset for the late 1980s through the University of Maryland's Global Land Cover Facility. They make imagery from all over the country freely available to anyone for download. The upside? It's free. The downside? It may not be the time period you want, or they may not have your area.



Legend
Created Wetlands
Natural Wetlands

0   0.125   0.25   0.5   Miles
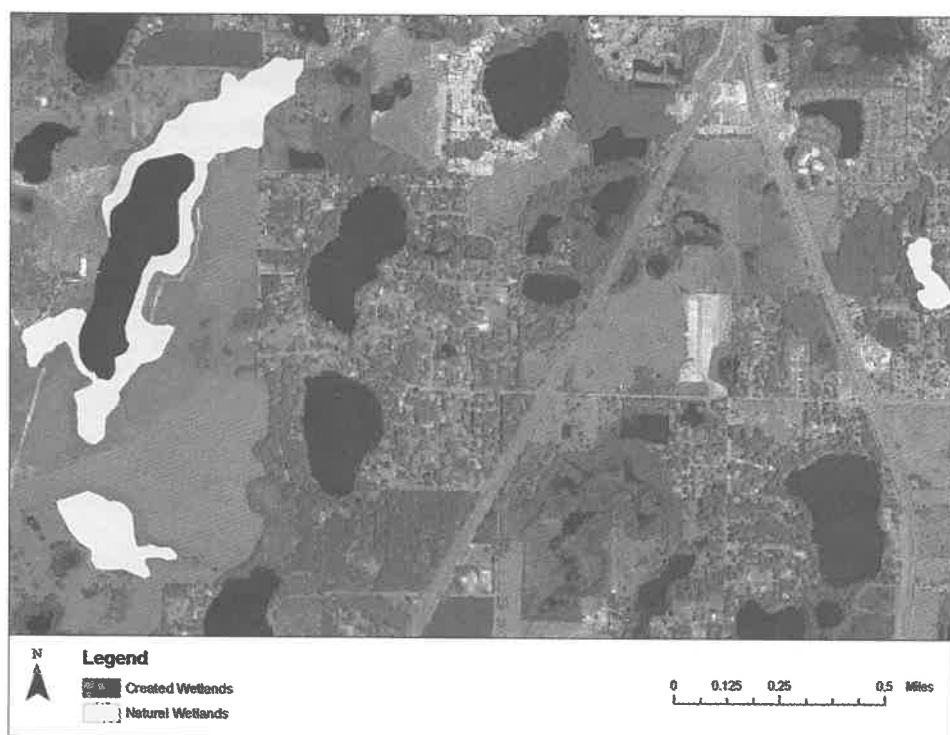
If you need something more detailed than Landsat, it gets more difficult. For instance, Space Imaging, the people who run the Ikonos satellite, will sell you 1-meter high-resolution imagery, but it's expensive. Many states and the U.S. Geological Survey do aerial photography at 1 meter – USGS will sell it to you or your state might give it away.

You might be able to avoid paying for imagery, but you won't duck the software price tag. And it's a doozy. Leica ERDAS IMAGINE , the industry standard software tool, is $3,500 last time I checked. Leica makes an extension for ArcGIS called ArcView Image Analysis. It is $2,500, and it is the one we went with, though I regret that now.

Comparing IMAGINE to Image Analysis is like comparing SAS to Excel. The power difference is huge, and you will outgrow Image Analysis before you're done. There are other software packages, but you don't even want to know how much they cost.

With software and data, we set out. The analysis I devised was based on a half a dozen academic papers I read that attempted something similar to the same thing. It's intentionally simple, and it's called unsupervised classifications.

(The detailed methodology with bibliographic information is available at www.sptimes.com/2005/05/22/State/In_depth_methodology.shtml.)

In unsupervised classifications, the software takes pixels and groups them together in to classes – clusters of pixels of close to the same values. You can set the number of pixel classes you want. We went with 100 to give us maximum flexibility. The trouble with unsupervised classifications is that the computer creates the classes but you have to identify what the class is.

How do you do that? You click. A lot. Using maps of ancillary data -- land-use maps, road maps, anything you think will help – you determine that all the Class 10s fit inside ponds, so you

identify all the Class 10s as ponds. We kept it simple, identifying things as wetlands or not wetlands.

After you get some practice, one Landsat scene might only take a day's work. Until then, it can take days. But with 14 Landsat scenes covering Florida, it was extremely tedious.

After classifying all the imagery twice – once for each year – you can then compare them. Image Analysis has a change detection algorithm built into it. The algorithm will compare two raw images, but the work that goes into comparing raw images is extremely complex and far beyond this article. With classified images, the algorithm will detect where it was once one thing and is now another.

From there, the analysis gets more complicated. To answer my question – to see where wetlands had been replaced by urban uses – I had to rely on another dataset, a satellite-derived map of urbanized areas. It was based on the same data, so I just totaled up the acres of area that had been wetlands but were now urbanized areas. To do that, I used another extension, Spatial Analyst.

Spatial Analyst, a $2,500 extension to ArcGIS, works with raster data, which is what satellite data is. Spatial Analyst has powerful tools that are like Excel for raster data. There are functions and formulas you can write to get the information you need. The syntax is a little tricky, but the ESRI support reference is very good.

Our analysis was, in the end, easily the most complicated work I've ever done. If you decide to try this, I strongly urge you to first take college courses in this. It is well worth the effort and a tremendous addition to your CAR toolbox. But it is complex data and not something you can learn in a tipsheet or an afternoon class.

I took two remote sensing classes, which are offered by many large universities. Many of these introductory classes are often offered for those who are getting geography degrees, and the advanced courses are often graduate level.

You can read Our Vanishing Wetlands online at www.sptimes.com/wetlands.

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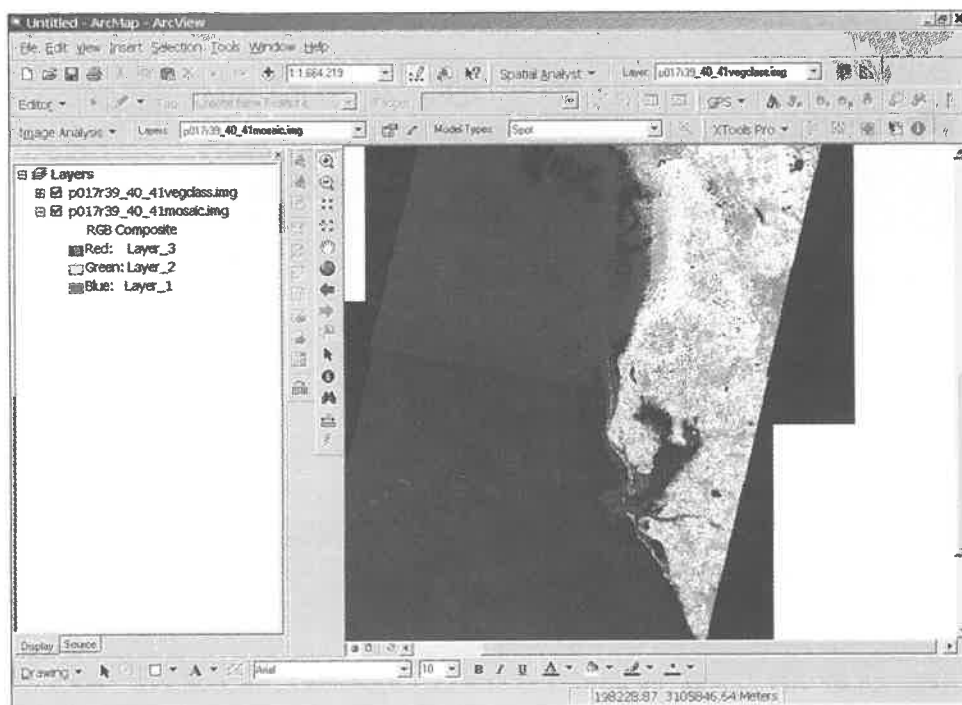
# Open

www.mysqlfront.de.) After you have data on your server, sign in to OpenInfo as an administrator, go to the main page and follow the "Grab an existing database" link. You'll be prompted to choose a database and table, then to enter a name and description of what information is stored in the table. Next, OpenInfo reads the structure of your data and asks you to fill in basic information about the table, such as the field that contains a unique identifier for each record and the field you want to use as the main display line for a record. (Figure 4) You also need to decide which fields you want included in the full-text search. (Note that this applies only to text fields, and because of limitations in MySQL, you can choose, at most, 16 fields.) You will be able to add additional information later, including adding detailed annotation to fields, specifying sort orders and setting up rules to create addresses and quick links.

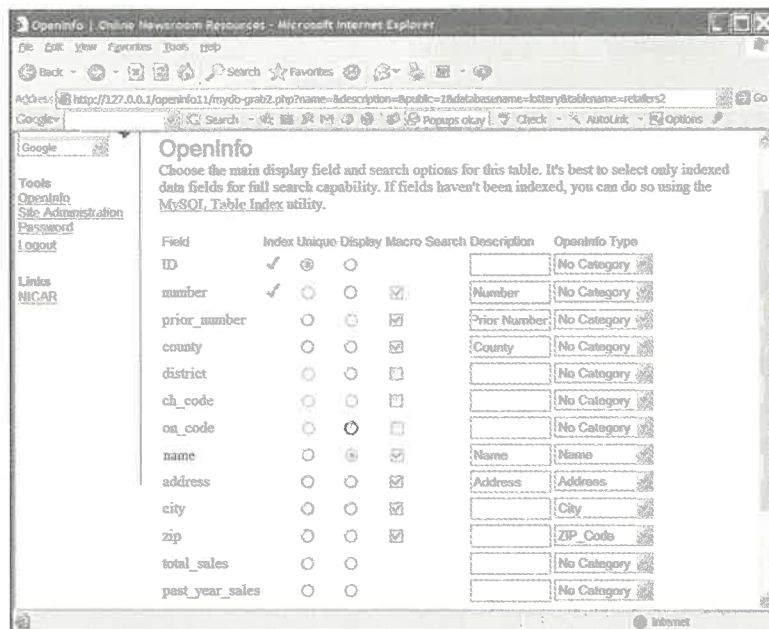OpenInfo is yours to play with and im-



Figure 4

prove. If you don't like the way it works, change it. If you have a suggestion for a feature you like, let me know and I'll work to incorporate it. Among the improvements I'm already polishing up (they'll be ready soon) are a system that picks out relevant records based on previous searches and the ability to search based on how a word sounds, not just its literal spelling.

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

# Mine

an employment table with details about the number of employees, the amount of production, mine type and other information, and an incident table with details about each injury and illness, including a narrative of the incident and information about the experience, shift and job title of the person involved. Data is available for 1983-2004.

By joining the employment table with the injury table, it is possible to calculate an injury rate for each mine, as well as a national average. Once you have identified mines of interest, however, you will need to turn to the data retrieval system to view citations and fines.

Much of the attention following the Sago Mine focused on the time it took the first rescue team to begin work. Ward reported in *The Gazette* that a rescue team did not enter the mine until almost 12 hours after the explosion.

To add depth to the story, Ward used a nationwide list of mine rescue teams available from MSHA from the mine rescue link off the home page, or *www.msha.gov/MineRescue/MAP/ASP/minerescuehome.asp*. The data comes as an Excel file with a row for each rescue team nationwide. The columns include number of members, the type of activities, the number of monthly training hours, and whether the team participates in mine rescue competitions.

By combining the rescue team data with other MSHA data, Ward was able to report there was only one rescue team for every four mines nationwide.

Human and Roberts caution journalists to be careful to note what dates are included in their analysis. Since MSHA updates its data retrieval system frequently, summary information such as violations and accident totals can change throughout the reporting of the story.

To get acquainted with the terminology and regulations associated with the mining industry, Ward recommends several sources, including the Pennsylvania Department of Environmental Protection mining glossary (*www.dep.state.pa.us/dep/deputate/minres/dms/website/training/glossary.html*) and the U.S. Code of Federal Regulations, which is available from the MSHA home page. Ward also recommends obtaining underground mining textbooks through interlibrary loan at your local library.

In addition to the textbooks, some colleges offer mining programs. The faculty there can provide another resource. Tomich drew on the experience of a local mining professor to help him explain the differences between the amounts mines were fined.

Contact Brian Hamman by e-mail at
brianh@ire.org.

# Connections

network analysis at *www.ire.org/sna*).

When the data was imported in UCI-NET as a text file and drawn for the first time in the program's NetDraw graphing application, it proved to be quite a monster. Social network types sometimes refer to this as the "hairball" – and it was quite a hairball.

Despite the visual mess, the analysis functions of UCINET allowed us to calculate who within our hairball held the most power in terms of a number of factors, such as "closeness," "centrality" and other terms which basically all are variation of the same question: who wields the most influence?

But displaying the information visually so that the relationship could actually be seen and understood within the network was our goal. And that was simply impossible with close to 1,000 ties.

While there are ways to work with the data in UCINET, we found it easier to pare down the data in FoxPro, where our master table resided, and then import the results back into UCINET for analysis.

As the peripheral relationships were slowly peeled away, we could begin to see the web of connections that formed the true core of Richmond's power structure. The results were striking. The boards were even more inter-

locked than we had originally surmised. A couple of organizations shared so many directors that they were almost indistinguishable from each other. And a core group of the city's powerful and wealthy seemed to have their hands in all the right pots.

Richmond's "good-old-boy" network was coming to light. While many people had already suspected who the major players were, even they were surprised to see just how intertwined the relationships were.

We still faced a few challenges. The most notable one involved Richmond's mayor.

Richmond had just elected a mayor, the former governor of Virginia, L. Douglas Wilder. He had come in bearing the torch of reform, claiming he would clean up city hall after years of corruption and abuse. But in doing so he had advocated for a change in the city's charter that would grant the mayor sweeping new powers. With those powers in hand, his influence was being felt across the city.

While Wilder did not fit into the network of interlocking boards, to totally ignore him would not be a viable option either for a project focused on power brokers within the city. So to bring him into the mix, reporters gathered data on campaign contributions to his mayoral race, and entered them into the database.

At first, we tried simply to include Wilder as his own "organization" – the Wilder

Campaign, so to speak. Those already in the 1,000-record main data table who had contributed to Wilder were given a new connection – to the Wilder node.

However, it became evident this approach was not going to work. First, because it was easier to make campaign contributions than to be appointed to a board of directors, Wilder's prominence in the network was exaggerated when added to the mix in UCINET. He had more ties than even the most connected directors. And even more crucially, the logic of the visual network map was that a tie between a person and a board or foundation translated to "is a director of." The ties to the Wilder campaign did not mean the same thing.

To solve this, we instead decided to display our nodes differently, showing whether the node was an organization, an individual who gave to the campaign, or an individual who did not contribute. In UCINET you can create data attributes that NetDraw can later use to create distinct symbols. We used blue squares for organizations, red circles for individuals who contributed to Wilder and green circles for all other individuals.

The result was a network map that showed a substantial buy-in from the city's connected and powerful to Wilder's campaign.

After some other final changes, the visual "spider web" representation of the city's interlocking directorates helped guide even more reporting for the project, and the months of digging by reporters came to fruition in a Sunday package of stories that had even some longtime residents shaking their heads at just how pervasive the good-old-boy network still was.

The network graphic that appeared in the paper was also accompanied by a series of interactive images on the *Times-Dispatch*'s Web site.

Contact Gordon Hickey by e-mail at ghickey@timesdispatch.com.

Contact Aaron Kessler by e-mail at kessler.aaron@gmail.com.
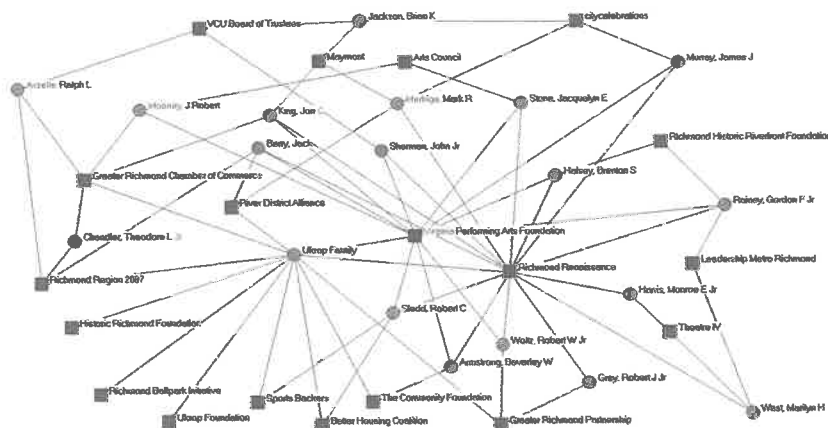
# Mapping

*continued from page 7*

into a Microsoft Excel spreadsheet and summed them. I did a similar query and table for the block groups that were covered by the sirens. This showed that the areas not covered by the sirens were mostly rural with few people. Only about 5,400 people in the county, or about 3 percent of Peoria County residents, were not within range of the sirens.

I also geocoded the county's nursing homes and schools. I got a list of schools from the county superintendent of schools and got nursing home data from *www.medicare.gov*. Not all of the addresses showed up on the map, so I went back out with my GPS device and got the coordinates for the schools I couldn't plot the addresses for. My initial geocoding showed that two schools
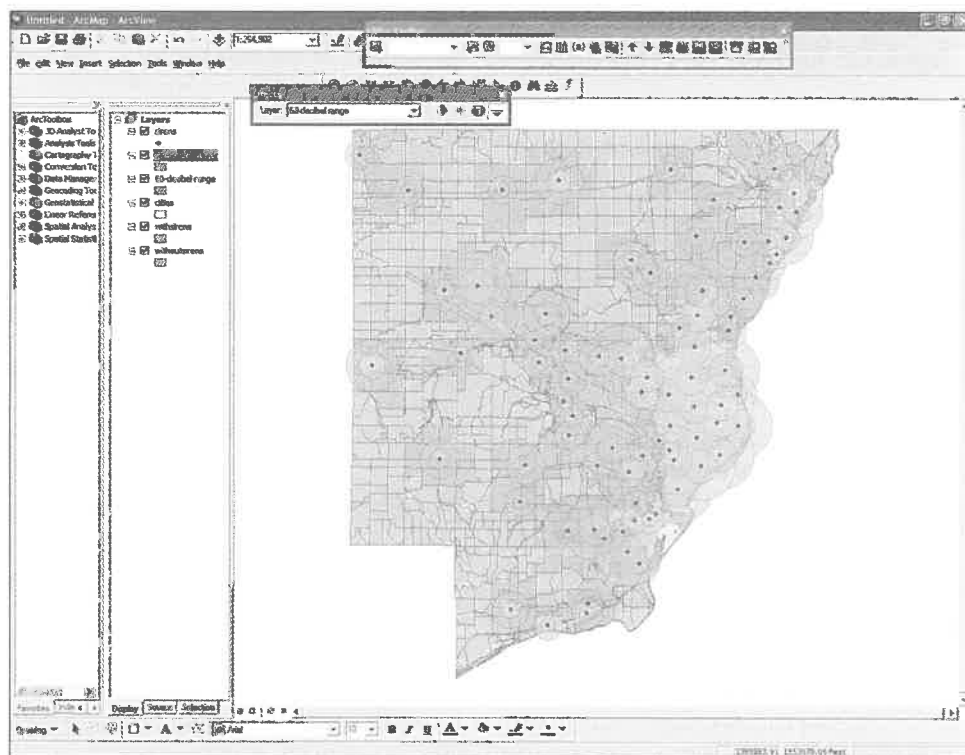


# Drug arrests

*continued from page 5*

than from the more urban areas. Those arrested were more likely to be mid-20s to mid-30s, instead of college-aged children coming from Cincinnati or nearby Northern Kentucky University as many law enforcement officials previously thought.

Because of space considerations, much of that information was not used in the final story.

After culling the data down to the essential information, saving each spreadsheet along the way, we were able to track the increase, however slight, over the three years covered in numbers and types of drug arrests and citations.

That is where the story started, not ended.

At the same time, our law enforcement sources were telling us that area police officers were looking into the same things we were.

As it turned out, police officers in Campbell County, Ky., were working to clamp down on drug use along the same stretch of highway where we had noticed problems.

A six-week sweep netted 70 busts for drug possession and use. It was the anecdotal evidence that helped make the story come alive.

Highland Heights Police Chief Carl Mullen, who organized the sweep, said the problem stemmed from users traveling south from Cincinnati, as well as from residents from every municipality in the county.

"I was surprised at the number arrested; we all were actually," Mullen said. "We knew it was a problem. But we didn't know how big of a problem it was."

Mullen said arrests included pregnant women, social workers, students and even mothers with their children in the car with them.

"It was all walks of life. It's not isolated to any class," he said. "We had one where a girl had told us that she at one point

had a $400-a-day habit. When she was approached, she didn't have any drugs on her, but she admitted to us that she had an addiction problem."

One officer talked about arresting a woman for shooting up heroin in her car while driving her children to day care.

Our examination of the trends showed that while there was a 17 percent increase in the number of arrests made for drug-related offenses in Cold Spring alone, there was a staggering 146 percent increase in arrests made for possession of a controlled substance.

Campbell County police continue to work together, nearly nine months later, on controlling the flow of drugs through the area.

In the words of Mullen, what happened over those six weeks is just the tip of the iceberg.

"We got 70 people," he said. "How many others do you think are out there?"

Contact Liz Carey by e-mail at lcarey@communitypress.com.