

## DATA ACCESS

# Year after attacks files still guarded

By Jennifer LaFleur  
*Reporters Committee for Freedom of the Press*

After Congress passed the 1996 amendments to the Freedom of Information Act, government agencies began putting more data and documents on their Web sites.

But that trend reversed after Sept. 11, 2001, when several federal agencies moved to take down maps, databases and entire Web sites from the public domain citing security reasons. Journalists no longer could get to the data they had used before in news stories.

In particular the door was shut to environmental data, transportation maps, dam locations and other databases frequently used by reporters, community groups and citizens.

Instead of the usual reports and information on Web sites, many users instead found messages

*continued on page 14*

## SPOTLIGHT: EDUCATION

# Databases add depth to school stories

By Holly Hacker, *IRE and NICAR*

School's back in session, so as students tackle reading and math, journalists can get busy reporting. Whether it's teacher quality or test scores, budgets or bus rides, the education beat offers lots of issues worth exploring with computer-assisted reporting.

Nowadays, journalists need to be even more insightful and aggressive in their education reporting. The federal government and all 50 states have put public schools under growing pressure to help kids succeed in the classroom. Most notable is the No Child Left Behind Act signed into law by President Bush this year. It requires states to set academic standards for children at each grade level, then to test children to see if they're meeting those standards.

Some groups of children tend to perform better than others, but in the past those disparities rarely came to light. Now schools must break down test results by race, English proficiency, socioeconomic status and other variables

Schools and districts must publish report cards with test scores, graduation rates, teacher qualifications and other data. So for education reporters and CAR junkies, these mandates make for a veritable data-fest with lots of numbers, statistics and trends to scrutinize.

*continued on page 19*

## SPOTLIGHT:

FOR MORE ON EDUCATION SEE:

- Uncovering back-door vouchers, p. 8
- Measuring long bus rides, p. 10

# Data shows cronyism in schools

By Jason Grotto  
*The Miami Herald*

The principal of Little River Elementary School wears sneakers to work. Armed with a walkie-talkie and clipboard, she patrols one of the poorest schools in Miami-Dade County, monitoring classrooms, comforting children and assisting teachers. She also prepares grant applications and solicits corporate donations to buttress the meager resources available for her school, which scores at the bottom on Florida's Comprehensive Assessment Test.

The cigar-chomping lobbyist wears expensive suits to work. Using political connections with school board members and a powerful union chief,

*continued on page 17*

## Bits & Bytes

New and updated  
from the Database Library

### Consumer product safety (new)

When you mix everyday life and consumer products, the results can be hazardous to your health. For example, from 1990 through 2001 more than 300 people died after they tripped over a consumer product. Find information like this and detailed information about many consumer products in the newest addition to the IRE and NICAR Database Library's government database collection: the Consumer Product Safety Commission Injury and Death database.

Former NICAR data analyst Mike Sherry used the CPSC data to investigate elevator safety. His story, "Missouri Program to Check Elevators Fails to Deliver," ran in the July 7 *St. Louis Post-Dispatch*.

The nationwide data covering 1990-2001 includes information about potential injuries, deaths and investigations related to consumer products. Some of the products include children's toys, bicycles, swimming pools, ATVs (three- and four-wheelers), sports equipment, hair dryers, smoke detectors, playpens, hobby items, lawn mowers and playground equipment.

*continued on page 4*

## NICAR UPDATE

# Changes with Uplink, Database Library

By David Herzog, NICAR and Missouri School of Journalism

If you've been reading *Uplink* for a while, you've probably noticed some changes over the past year. We redesigned the newsletter and added new features to keep you current with the latest goings-on in computer-assisted reporting. For instance, we now highlight some of the best uses of analytic mapping software in "Mapping it Out." But we haven't thrown out those longstanding features, such as "Tech Tips," that our readers have found invaluable.

Likewise, you'll see some more changes in the issues ahead. One is that we'll coordinate *Uplink* more with *The IRE Journal*. That means by 2003 we'll have *Uplink* and the *IRE Journal* on the same release schedule. And we'll tie *Uplink* articles more closely to ones that run in *The IRE Journal*. So if we have an *IRE Journal* article about an investigative story, the corresponding issue of *Uplink* may have a related article that dissects the database work that went into the story.

We're also moving "Bits & Bytes" off *Uplink's* cover and onto page 2 to make room for one more front-page story. Some of these articles you'll see on the cover will take a look at issues that affect journalists doing computer-assisted reporting. This edition's story about restrictions to

data a year after the terrorist attacks of Sept. 11 is but one example.

*Uplink* has four new contributing editors who also are working as University of Missouri School of Journalism graduate research assistants in the NICAR Database Library. They'll write regularly for *Uplink* and help develop story ideas. The contributing editors are:

- Megan Christensen, who is a veteran of the database library, in her second year working on her master's degree and has interned at the *Detroit Free Press*.
- Jaimi Dowdell, another veteran of the database library who just started working on her master's degree this fall.
- Holly Hacker, who just left the *St. Louis Post-Dispatch* to work on her master's degree. She has reported for the *Ventura County (Calif.) Star*, the *Antioch (Calif.) Ledger-Dispatch* and the *Gazette Newspapers* in Montgomery County, Md.
- Stephanie Kang, another newcomer to the database library, has begun working on her master's degree and had interned at the *Los Angeles Times*.

In addition, Jeff Porter and I are swapping *Uplink* duties so Jeff can focus more of his attention on the work of the database library, where business has picked up considerably during the past few months. With this issue, Jeff becomes assistant managing editor and I become managing editor.

Thanks for reading *Uplink* and feel free to e-mail us at [uplink@nicar.org](mailto:uplink@nicar.org) if you have any suggestions.

## Census workshops scheduled for reporters, editors

IRE and NICAR have scheduled three workshops for journalists to learn how to use Census data.

Paul Overberg of *USA Today* and Steve Doig of Arizona State University will lead these sessions Oct. 5-6, at *USA Today* headquarters in McLean, Va.; Nov. 16-17 in Tempe, Ariz.; and Jan. 11-

12, 2003 in Columbia, Mo.

The workshops will include story ideas and tips for using the latest and upcoming data releases, including the American Community Survey and Public Use Microdata.

Visit [www.2000census.org](http://www.2000census.org) for more information and to register.

## CAMPAIGN FINANCE

# Building a database for tracking state money

By MaryJo Sylwester, *Center for Public Integrity*

State Secrets, an investigation of contributions and expenditures reported by state political party committees, started with a pile of paper reports that stood more than 15 feet high. That doesn't count thousands of records obtained directly in electronic format.

All told, there were nearly 400,000 records — and no easy way to put them together. The paper forms differed from state to state. Even the content in a single state didn't always remain consistent.

As a result, compressing all of that into one table of contributions and another of expenditures — and cleaning them — provided valuable lessons in how to build an effective database.

The Center for Public Integrity, in conjunction with the Center for Responsive Politics and the National Institute on Money in State Politics, published the findings of the yearlong investigation on June 25. (See the July-Aug. 2002 *Uplink* for more on political reporting.)

Our analysis showed that 46 percent of the money state political parties collected in the 2000 election was soft money from national party coffers. These transfers confirmed a commonly held perception that state parties are used to channel soft money and influence presidential and congressional elections.

## Paper to data

Even the most organized person would be taxed by the collection of hundreds of reports and the data entry. Planning how you're going to tackle the paper pile is crucial no matter the database size. If you don't set it up properly, you'll have bigger problems later.

There were some things that made life easier with State Secrets:

- I created an Access database where our staff members tracked our FOIA requests for each state, including phone numbers, Web sites, and notes. They also had a Word file for each state for more extensive notes and e-mails. I could run a query to quickly see what was missing.

- I pored over the campaign finance forms and listed the various features of each in an Excel worksheet. This showed me which fields were necessary for data entry and where the limitations were.

- We hired a data entry firm. This was, by far, the best move we made.

- We broke the work into manageable chunks. In this case, it was on a state-by-state basis. Once all the data entry, basic cleanup and checking was done on each of the states, we appended the files together. Some of the cleanup (like standardizing names) was better left until after we assembled the full table, however.

- We ran summary queries to compare the total expenditures to the summaries listed on paper reports. We found missing pages, duplicate records and even bad math on the part of the political parties that we would not have found without these queries.

In hindsight, I should have created ID numbers for each of the paper reports. The data entry house could have typed those IDs into the data to indicate which paper report each item came from. And the forms could have been stored in numerical order for easier access.

Instead, we used the name or date of the report (such as "Fourth Quarter report" or "Pre-Primary"), but there was so much inconsistency we had to spend several days standardizing the report names in order to do the summary queries mentioned above.

I also would have assigned a code to each committee to be typed in addition to the name. We applied a code only after the information was typed — a difficult process because the committee names were never consistent.

## The scrubbing phase

Before embarking on data cleanup you should answer two questions: What standards do you want to meet? And how dirty is the data? (See page 16 for details.)

On the expenditure data, we decided it would be an ineffective use of our time to clean up all recipient names, but consultant names should be standardized so we could focus on those. On the contribution data, we scrubbed and scrubbed to make sure the contributor names were as clean as we could get them. We set a very high standard because we wanted to be positive that we identified the true top donors.

Figuring out the dirtiness of the data is the harder part. I'd recommend running queries, and making notes about the variations you find. Are there big chunks with consistent patterns that could be fixed using an update query? If you can pinpoint places to do mass updates first, that will save work later. Eventually, you'll have to do line-by-line cleanup no matter what, so do whatever you can to minimize that.

All of the suggestions I've listed here are beneficial no matter how small or large the database. Consistency and advanced planning will make your job much easier in the long run.

MaryJo Sylwester, formerly with the Center for Public Integrity and now with *USA Today*, can be contacted by e-mail at [msylwester@usatoday.com](mailto:msylwester@usatoday.com)

# MAPPING IT OUT

*The latest uses of mapping in news reporting.*

## Plotting school data

Public education agencies collect huge amounts of information about students, teachers and schools. And most of that data can easily be tied into a geographic location: a school building point, an attendance zone or even a census tract. So there are plenty of opportunities for journalists to use GIS analysis in education reporting. Here are some ways that journalists have been turning to GIS to help uncover trends in education.

### Shifting attendance

John Welbes, an education reporter for the *St. Paul Pioneer Press*, heard during a school board meeting last year that some Minnesota school districts had been losing students under the state's open enrollment system. That meant those schools also lost state aid money tied to enrollment.

Welbes obtained an Excel spreadsheet from the state that contained one row for every public school student. One column contained the identification number of the school that the student attended. Another contained the identification number of the school district in which the student resided.

Using that file, Janet Roberts, the newspaper's computer-assisted reporting coordinator, summarized the data by loss or gain for the region's school districts. She imported the table to ArcView and mapped the changes in the Minneapolis-St. Paul metropolitan area.

(See Map 1) The map showed that center city districts lost students; so did some in the suburbs.

### Charter schools

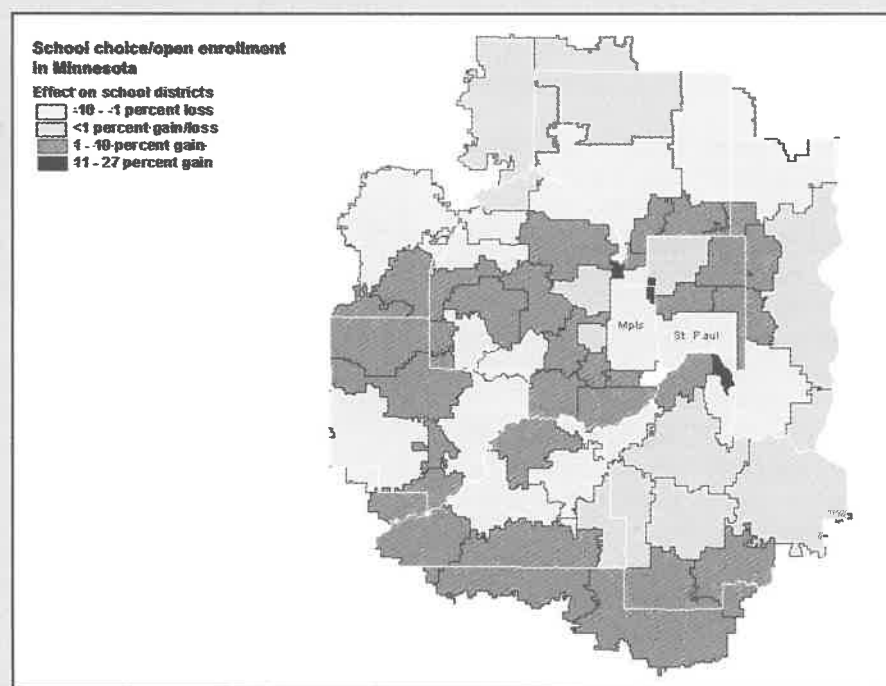
As part of its "Cheating the Classroom" stories, reporters at *The Miami Herald* wanted to see whether charter schools in Miami-Dade County were anything like their surrounding neighborhoods. (See the cover story in this edition of *Uplink* for a more detailed account of the series.)

*The Herald* had a data file with information about each school, including the school's street address. Using the geocoding function in ArcView, database editor Jason Grotto created a point map layer that showed the location of all the schools. Then he joined a table of school demographic information and another of neighborhood demographic data from Census 2000.

Grotto then selected nearby non-charter public schools in another map layer and found that many of the charter schools' student demographics were different from those in adjacent public schools. In addition, the GIS analysis showed that the charter schools failed to reflect the neighborhoods they were supposed to serve.

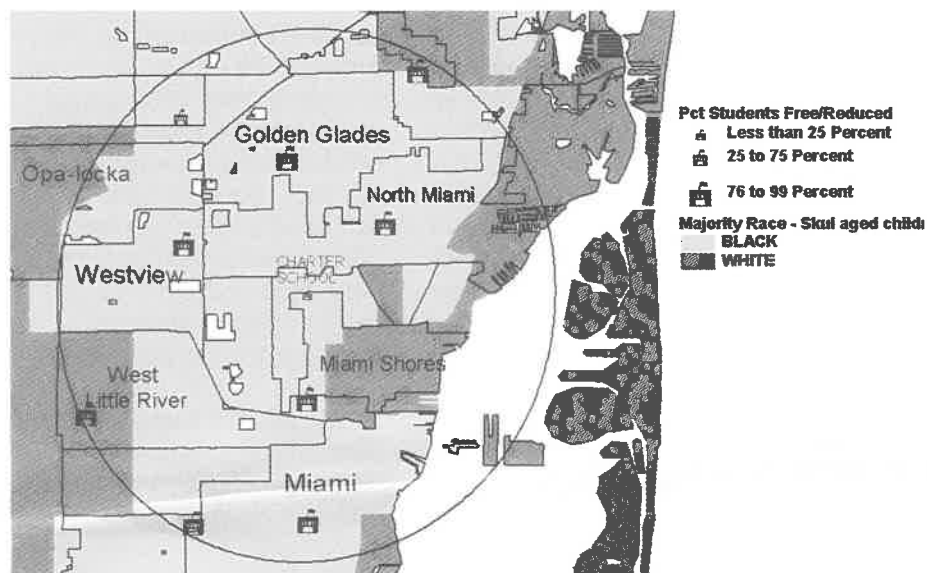
For instance, according to its contract with the Miami-Dade schools, Miami Shores/Barry University charter school is required to serve students within a 4-mile radius, and its population should reflect the community. But the mapping showed the charter was the only public middle school in that 4 mile radius where fewer than 25 percent of the students received free or reduced lunch. It also was the only school with a majority white student population.

(See Map 2).



Map 1: A *St. Paul Pioneer Press* map shows that school districts in the Twin Cities lost students under the state's open enrollment program.

## CHARTER SCHOOL GRAPHIC



**Map 2:** This map helped Miami Herald reporters see that a charter school, located in the center of the map, failed to reflect the student population in the surrounding community. The graduated size schools symbols show that the charter school had a lower student poverty level.

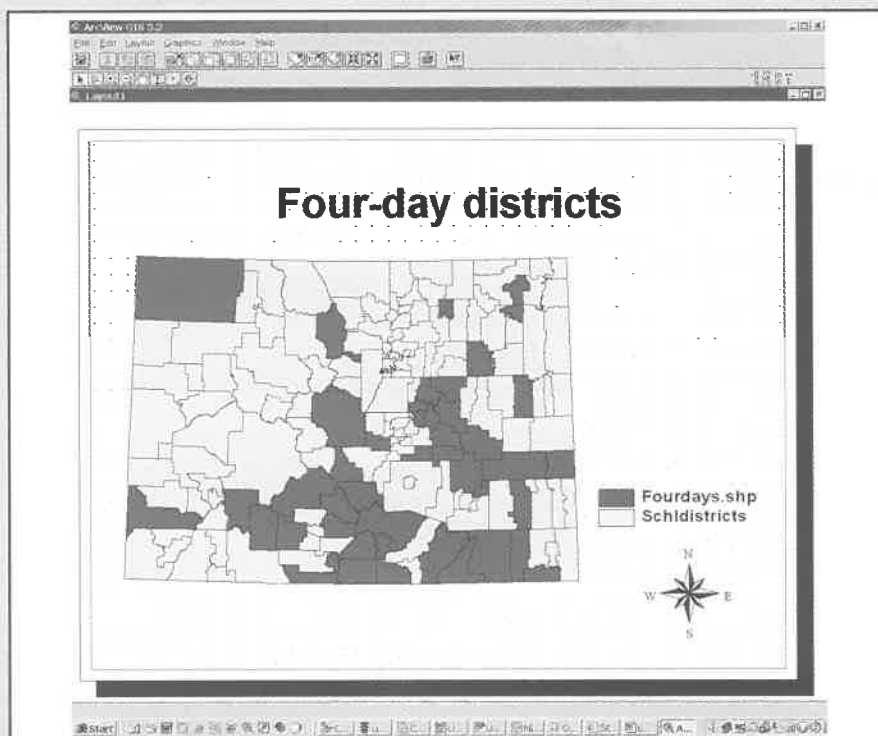
*Would you be willing to share a mapping example with fellow journalists? Send an electronic copy of the map along with details to David Herzog at [dherzog@nicar.org](mailto:dherzog@nicar.org).*

See the series on the Web at: [www.miami.com/mls/miamiherald/3024226.htm](http://www.miami.com/mls/miamiherald/3024226.htm).

## School week

The *Denver Post* used ArcView 3.2 for a story about Colorado school districts that switched to four-day weeks. (See Map 3) The map showed that the switch to a shorter school week mostly was happening in the rural parts of the state, where administrators were trying to save money.

Monte Whaley, one of the newspaper's education reporters, entered the information about the district schedules from the Colorado Department of Education into a spreadsheet. CAR editor Jeff Roberts then mapped the results using a school districts shapefile.



**Map 3:** The *Denver Post* created a map that showed how a switch to four-day school weeks mostly was happening in the rural parts of the state.



## Uncovering “back-door vouchers”

By James E. Wilkerson, *The (Allentown, Pa.) Morning Call*

Late last year, the superintendent of schools for the Roman Catholic Diocese of Allentown, Pa., told our education reporter Nancy Averett that the schools had received more than \$600,000 in tuition donations from local businesses under a new state tax-credit program.

She already knew about the program, but was surprised by the amount going to one diocese. Dubbed “back-door vouchers” by critics, the scholarship tax

was being funneled to religious schools with very little supervision or follow-up from the state.

The program, started in early 2001, gives businesses up to 90 percent state tax credit for money they donate to nonprofit organizations, which then distribute the money to students for tuition. Unlike full-blown vouchers, voted down in the state legislature three times in recent years, support-

vehement opposition.

## Background research

To do the job right, we had to examine more than 100 organizations approved by the state to receive scholarship money.

That's where the computer-assisted reporting work started. Nancy was familiar with Access, so I created an easy-to-use database that we could both use whenever our schedules allowed. Using the Forms Wizard, I created a simple data entry form that included the organization's name and other basic information, and provided space for us to add research notes, Web sites and other information we found.

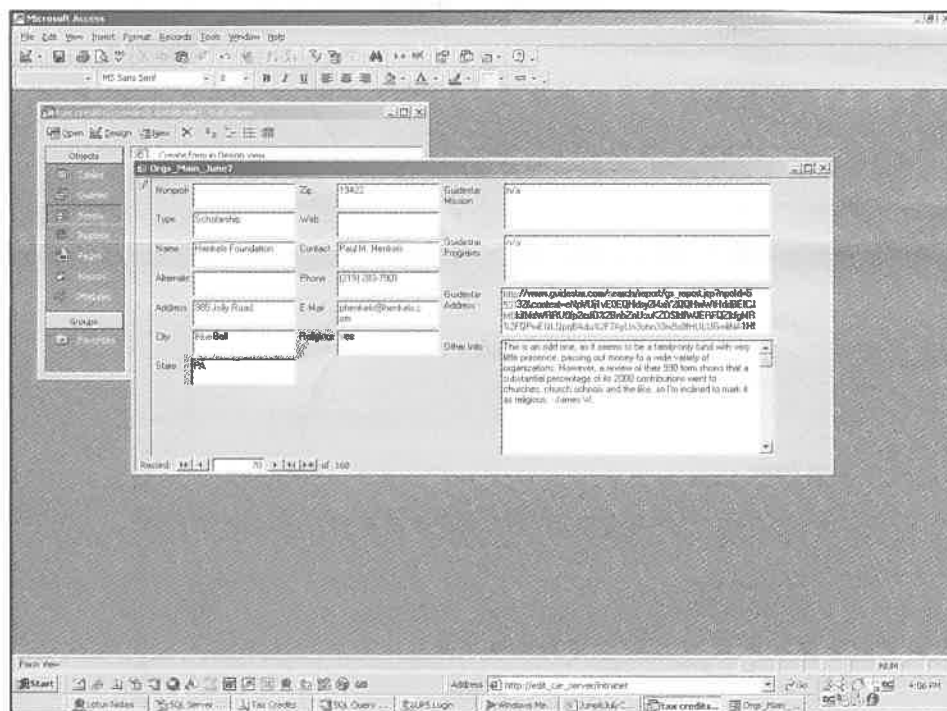
The state had lists of approved organizations on its Web site, including address, contact information and Web address. I copied those into a powerful text editor, UltraEdit, (available at [www.ultraedit.com](http://www.ultraedit.com)) and designed macros that parsed the five rows of data for each organization into a fixed-width text file that I later imported into Access. I added the fields from the table we had created in Access and we were ready to roll.

We relied almost exclusively on the Internet, Lexis-Nexis and our own newspaper archive for researching the organizations. Especially useful was GuideStar ([www.guidestar.com](http://www.guidestar.com)), a database containing the financial disclosure forms (Form 990) that nonprofit organizations are supposed to file annually with the Internal Revenue Service. In some cases we were unable to find information about an organization elsewhere, but the 990s on GuideStar provided details about how the groups were distributing money.

## Merging data

Within three weeks or so we had finished the bulk of our research. By that time, the state had given us a list of contributing businesses and the donation amounts that went to the nonprofits.

Unfortunately, the data was in two files, with the business name and address information in Word and the donation



credits were clearly funneling a lot of tax dollars to private, nonprofit schools. So Nancy came to me wondering if we could find out exactly how much.

Program administrators with the state Department of Community and Economic Development said they didn't know. In fact, there was a lot about the donations they didn't know. So we ended up having to build our own database, one that allowed us to show how millions of dollars in tax money

ers emphasized that the tax credit took no money directly from public school funds.

That argument seemed a little flimsy to us, however. The program diverted money away from the state coffers, money that could have easily gone to education. And, we suspected, most of that money went to private religious schools. In past legislative sessions, the potential of public money aiding private religious schools had triggered

amounts in Excel. I converted the Word document into a text file and again used UltraEdit to convert the information into

After running our findings by state officials and finding out that they had higher totals, we realized that we were missing some data. We put the story on hold until the year-end data became available in April. However, in the following weeks we kept plugging away at organization research and updated other data. When the final data became available we were ready to start writing.

Though the total contribution tally had climbed to about \$30 million, the share of

Our final story, which ran Aug. 4, showed not only that the lion's share of money was going to religious groups, but that the state was doing very little to track how the money was being used.

In fact, interviews with administrators at some of top contribution-getters in the state showed that much of the money was being distributed to students already in private schools, undermining the state's claim that needy students in poorly performing public districts were being targeted.

Along with the story, I built a Web lookup tool that allows readers a look at contribution lists. The story and query are available at [www.mcall.com/all-a1-5\\_voucheraug04.story](http://www.mcall.com/all-a1-5_voucheraug04.story)

James E. Wilkerson can be reached by e-mail at [james.wilkerson@mcall.com](mailto:james.wilkerson@mcall.com)

columns and rows. Once that was completed, I imported the fixed-width text and Excel files into Access and merged them into one table.

scholarship contributions going to religious organizations held steady, at about 75 percent, so our story did not change drastically during the delay.

Fortunately, with few exceptions, the nonprofit names were the same in all of the documents I was using. That allowed me to join on the name field with little clean-up.

When it came time to produce numbers, I had whittled the data down to two tables: Our research table, which included data about the organizations receiving money, and the business and donation table.

## Tallying results

I used SQL Server for my analysis, but the work could have just as easily been done in Access or any other database program. I joined the two tables on the organization name field and ran some simple Group By queries to look at the distribution of the donations.

In the end, we found that more than half of the \$17 million in donations had gone to scholarship organizations that primarily benefited religious schools. Looking only at scholarship organizations, about 75 percent of donations had gone to religious schools.

**Business Leadership Organized for Catholic Schools (BLOCS)**

222 N. 17th Street  
Philadelphia, PA 19103  
Type: Scholarship  
Web site: <http://www.blocs.org>  
Contact Name: Regina DiGilio  
Phone Number: (215) 387-0590  
E-mail Address: [RDIGILIO@blocs.org](mailto:RDIGILIO@blocs.org)  
Primarily Benefits Religious School or Schools? Yes

Contributions as of July 2002

Company	Date Received	Organization	Amount
Aluminum Bank	03/04/02	Business Leadership Organized for Catholic Schools (BLOCS)	\$10,000
Advanced Home Health Care, Inc.	06/03/02	Business Leadership Organized for Catholic Schools (BLOCS)	\$1,000
Austrian J. Decker Associates, Inc.	08/09/01	Business Leadership Organized for Catholic Schools (BLOCS)	\$1,500
Armstrong Sports and Entertainment Services, Inc.	08/15/01	Business Leadership Organized for Catholic Schools (BLOCS)	\$20,000
Woodfield Mutual Service Bank	12/09/01	Business Leadership Organized for Catholic Schools (BLOCS)	\$15,000
Bank of the Americas, Inc.	12/03/01	Business Leadership Organized for Catholic Schools (BLOCS)	\$4,000
Comcast	05/22/01	Business Leadership Organized for Catholic Schools (BLOCS)	\$10,000
DVI Personal Services	09/15/01	Business Leadership Organized for Catholic Schools (BLOCS)	\$10,000

**Pennsylvania Nonprofits Ordered by Total Contributions**

Click on the business or organization name for more details

1	<a href="#">Scholastic Opportunity Scholarship Fund (SOS)</a>	SO	Pittsburgh	PA	\$1,982,864
2	<a href="#">The Mennonite Foundation, Inc.</a>	SO	Goshen	IN	\$1,200,053
3	<a href="#">Business Leadership Organized for Catholic Schools (BLOCS)</a>	SO	Philadelphia	PA	\$1,152,090
4	<a href="#">Diocese of Scranton Scholarship Foundation</a>	SO	Scranton	PA	\$1,142,977
5	<a href="#">Henkels Foundation</a>	SO	Blue Bell	PA	\$776,499
6	<a href="#">Neumann Scholarship Foundation</a>	SO	Harrisburg	PA	\$718,386
7	<a href="#">Eastern Pennsylvania Scholarship Foundation</a>	SO	Allentown	PA	\$682,840
8	<a href="#">Pittsburgh Jewish Educational Improvement Foundation</a>	SO	Pittsburgh	PA	\$623,622
9	<a href="#">Philadelphia Youth Network, Inc.</a>	EIO	Philadelphia	PA	\$621,736
10	<a href="#">STAR Foundation</a>	SO	Erie	PA	\$582,455
11	<a href="#">Foundation for Jewish Day Schools of Greater Philadelphia</a>	SO	Philadelphia	PA	\$517,796
12	<a href="#">Abington Friends School</a>	SO	Jenkintown	PA	\$462,000
13	<a href="#">Junior Achievement of South West PA</a>	EIO	Warrendale	PA	\$452,083

## SPOTLIGHT: EDUCATION

# Measuring long bus rides for rural children

By Scott Finn and Eric Eyre, *The Charleston Gazette*

During the past decade, West Virginia education officials closed one of every five schools in a massive consolidation drive. State officials said the closings were necessary because of dwindling enrollments in rural areas.

Parents and consolidation opponents said rural children were forced to ride the bus four hours a day or longer because of the closings. State officials said most children rode the bus for only a short period of time. Each side relied on anecdotal evidence to back their claims.

We started with asking a simple question: "How long are children's bus rides in West Virginia?"

We thought that someone, somewhere within the educational bureaucracy would have attempted to answer that question. Instead, we discovered the state Department of Education had ignored a 1998 state law telling them to study the issue of lengthening bus rides.

Even if education officials had followed the law, West Virginia's antiquated record-keeping system made it almost impossible to answer the question. In other states, school systems keep computerized records of bus routes and riders. But most of the school districts in 35 rural counties did not. In fact, bus drivers themselves kept most of the records by hand.

So we had to construct our own database of 1,569 bus runs in rural counties. We started our project intending to study all bus rides in the state, but we soon discovered that many urban districts also did not keep computerized busing records. We decided to focus on the state's rural counties, in part because that's where the longest bus rides occur, and to keep the data entry manageable.

The process, from FOIA requests to data entry to double-checking, took us eight months. But the resulting ability to write with authority made the effort worthwhile.

In "The Long Haul," published in August, we found the number of children who rode the bus more than two hours a day doubled during the 1990s. Also, we found that two-thirds of bus routes carrying elementary children exceeded state guidelines for a "reasonable" bus ride – an hour per day or less. The average student at Clay Elementary, a typical rural school, spent 37 minutes on the bus each way, nine minutes longer than the average Los Angeles commuter.

Our data helped us locate the best areas for conducting interviews and finding stories. In one county, elementary school students counted the 46 bolts on the roof of their bus to while away their hour-and-40 minute ride to school. Teachers at one high school slipped cups of coffee to 15-year-old students to help them stay awake in class. Parents of one 4-year-old dressed him in his sleep so he could enjoy a few more minutes of sleep before his one-and-a-half hour ride.

## Gathering the data

Through the state Freedom of Information Act, we obtained records for bus runs for the state's 35 most rural counties. (Those counties were eligible for federal school construction funds because they were the most sparsely populated of the state's 55 counties.)

At first, several county school systems fought our efforts to obtain the bus

Microsoft Excel - uplinkbustimes

File Edit View Insert Format Tools Data Window Help

G41 =COUNTIF(G2:G36,">0:30")

	A	B	C	D	E	F	G	H	I	J
1	Bus#	County	begin	el end	ms end	hs end	el dif	ms dif	hs dif	
22	152	Monroe	7:05	8:03	8:03	8:15	0:58	0:58	1:10	
23	154	Monroe	6:50	7:55	7:55	7:45	1:05	1:05	0:55	
24	155	Monroe	7:01	7:43	7:43	8:15	0:42	0:42	1:14	
25	156	Monroe	7:08	7:55	7:55	8:25	0:47	0:47	1:17	
26	157	Monroe	6:25	7:55	7:55	8:25	1:30	1:30	2:00	
27	158	Monroe	7:11	7:45	7:45	8:10	0:34	0:34	0:59	
28	159	Monroe	7:17	7:50	7:50	8:14	0:33	0:33	0:57	
29	160	Monroe	7:15	8:05	8:05		0:50	0:50		
30	161	Monroe	6:55	7:55	7:55	7:36	1:00	1:00	0:41	
31	162	Monroe	7:18	7:46	7:46	8:17	0:28	0:28	0:59	
32	163	Monroe	7:05	7:45	7:45	8:15	0:40	0:40	1:10	
33	164	Monroe	6:55	7:50	7:50	8:15	0:55	0:55	1:20	
34	166	Monroe	7:25			8:06			0:41	
35	167	Monroe	6:48	7:40	7:40	8:15	0:52	0:52	1:27	
36	168	Monroe	6:48	7:45	7:45	8:15	0:57	0:57	1:27	
37										
38						average	0:49	0:49	1:08	
39						median	0:51	0:51	1:10	
40										
41						over	30	21	17	
42						total	34	34	29	
43						percent	88.2%	61.8%	58.6%	
44										

Ready



routes. Some claimed it contained private information, so we asked them to redact any personal information, such as children's names. Other school systems wanted to charge us large sums for the printing. A couple of counties required us to drive up to three hours to their headquarters to obtain the records.

We created an Excel spreadsheet that included when each run started, when it stopped, and how much time children rode inbetween.

The bus runs represented the longest possible time a child might ride a bus. The run began when the first child stepped on the bus, and ended when the children were dropped off at an elementary, middle or high school. The bus run times did not include the time children stood waiting for the bus to arrive. These were ideal times, not taking into account bad weather or traffic delays.

### Calculating with Excel

To determine how long each bus run was we subtracted the start time from the drop-off time. We then used Excel's Count If function to count only the runs that exceeded state guidelines. We divided that by the total count to determine the percent of runs over the guidelines.

The average bus ride could not be tabulated for most counties because the schools don't track when individual children board their bus. Clay County is one of the few that tracks exactly when each child boards the bus and when they are dropped off. Because we had the data, we were able to determine that the average Clay Elementary student spent 37 minutes on the bus, more than the state guideline of 30 minutes. The average middle school student was on the bus for 34 minutes, the average high school student for 38.

We checked our results with transportation directors in all 35 counties, both through phone calls and by sending a copy of their county's data to them

through the mail. We made corrections to the database based on their responses. In some cases, we spent half an hour going over a particularly complicated bus schedule with a transportation director.

For purposes of the stories, junior high schools were classified as middle schools. Buses for special education students, who often ride much longer than other students, were excluded from the database.

We also relied upon two surveys of county school transportation directors by the state Department of Education, one in 1992 and the most recent in 1996. The survey asked transportation directors to estimate how many children rode the bus for longer than state guidelines allow.

Those surveys are imperfect, because they rely upon the impressions of the directors and not hard data. For example, one county's survey had exactly 600 students on the bus for one time increment, and exactly 700 for another increment, leading us to believe the times were estimates, not the detailed study state officials led us to believe. If anything, we believed they tended to underestimate the amount of time children spend on the bus, based on comparisons to our spreadsheet.

So far, the story has resulted in a promise from the state transportation director to computerize all bus schedules and complete its own study about the length of bus rides.

We are planning more stories this fall on the damaging effects school closings have on the state's youngest and poorest students.

To see the whole series, visit [www.wvgazette.com/news/Closing/](http://www.wvgazette.com/news/Closing/).

Contact Scott Finn by e-mail at [sfinn@wvgazette.com](mailto:sfinn@wvgazette.com).

Contact Eric Eyre by e-mail at [ericeyre@wvgazette.com](mailto:ericeyre@wvgazette.com).

## Annual Computer-Assisted Reporting Conference

— MARCH 14-16, 2003 —



### Panels that touch on every beat:

Census, crime, education, local and state government, freedom of information and transportation

### Hands-on computer classes:

Spreadsheets, database managers, mapping, statistical software and more

### For hotel reservations call:

866-837-4148 and ask for the Investigative Reporters and Editors room block before Feb. 14

The Westin Charlotte  
601 S. College St.  
Charlotte, NC 28202

### Sponsors:

The Charlotte Observer  
NBC News

### REGISTER ONLINE AT

[www.ire.org/charlotte](http://www.ire.org/charlotte)  
or call 573-882-2042

## HOMELAND SECURITY

# Rating cities' readiness to handle attacks

By Mike Fish, *CNN.com*

Just how prepared are major U.S. cities to face a terrorist attack? Can we assume places like New York and Washington, D.C., are more up to speed than, say, Milwaukee or Columbus?

The questions were being posed by editors at CNN.com as another possible follow up to events of Sept. 11. Like several news organizations, we had completed a computer-assisted reporting project that focused on the airline industry and security at the

that produced "How Prepared Is Your City?" on the Web site, a 30-minute special report on "Wolf Blitzer Reports" and also a TV package by CNN Homeland Security correspondent Jeanne Meserve.

The early challenges were plenty. First, we were determined to produce a thoughtful package that would both enlighten and address people's concerns, yet at the same time wouldn't serve as a blueprint for would-be terrorists and criminals. We found this

whelming and time consuming a task. And when we considered 25, we struggled to justify eliminating important cities like Miami and Atlanta that barely missed making the cut.

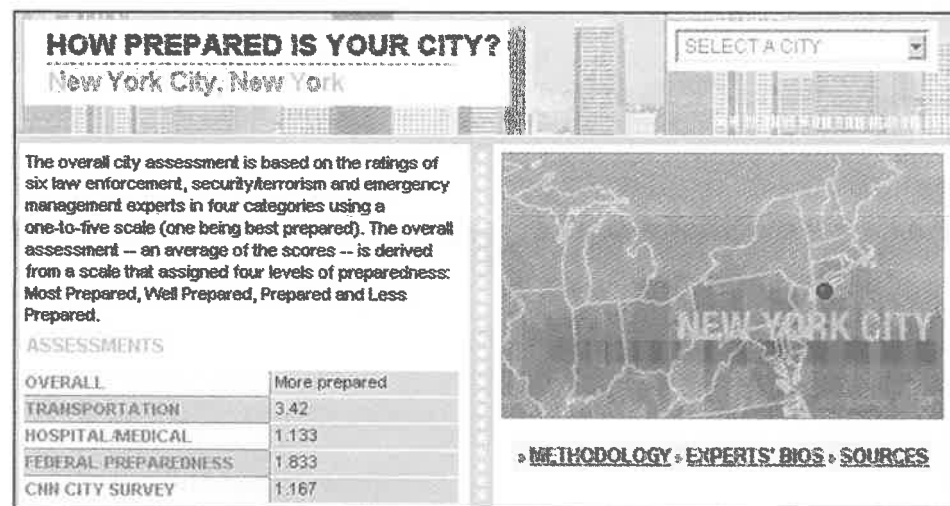
Ultimately, we reviewed national crime statistics reports and developed our top 30 cities by the criteria that cities had to have a population of at least 350,000 and also employ at least 1,100 full-time police officers. Even then, to include as many cities as possible, Dallas/Fort Worth and San Francisco/San Jose were combined rather than evaluated individually. We made sure to present separate data on each city in our final report and, as it turned out, our final rankings for these cities would have been the same no matter which way we did it.

That bit of information would prove helpful in addressing challenges from Fort Worth and San Jose officials, who argued that their respective cities had their own individual issues and didn't care to be evaluated with their neighbors.

Once we were comfortable with our list of cities, we dug for every shred of relevant data and public record that was available. The analysis would focus on four crucial categories: transportation (access to and from emergency sites); hospital/medical; federal preparedness (an evaluation of funding and training programs); and emergency management.

True preparedness is often deemed best achieved where the local, state and federal governments — as well as members of the private sector — work as partners. But we recognized a challenge in preparing for emergencies was also getting access to available resources.

The Centers for Disease Control and Prevention, we found, awarded \$120 million to states during the past two years to upgrade public health's capacity to respond to bioterrorism. But only four cities received separate grants, including New York (\$1.14 million), Chicago



country's busiest airports, and there was a desire to similarly rank, or at least gauge, the performance of major cities.

Terrorism experts, who by all appearances had multiplied in number after Sept. 11, were insisting on two things: no city is immune from risk and no one is fully prepared to handle the fallout of an attack. So folks couldn't help but worry about their own safety as well as suddenly wondering if their city was schooled to deal with everything from biological to chemical threats.

To assess the preparedness of the largest cities, CNN.com undertook a computer-assisted reporting project last fall

also to be the top concern of the experts who were interviewed or assisted us in evaluating the cities.

This ongoing dialogue with a team of experts proved invaluable, as they often had first-hand knowledge of the preparedness efforts and vulnerabilities within a number of major cities. We also were able to seek their input on the statistical data, federal grant money and such criteria that would best help in evaluating the work done by city officials leading up to Sept. 11.

### Narrowing the cities

OK, but how many cities could we review in depth? We thought initially 50, until deciding it would be too over-

(\$902,089), Los Angeles (\$850,275) and Washington, D.C. (\$570,317).

Finding, analyzing and weighting the value of consistent, statistical data that could be applied, in our four categories, across the 30 selected cities, would be a key to making this study work.

## Surveying experts

Early in the project, however, it became apparent that statistics alone were not going to tell us the full story of a city's preparedness. Some of the information that would be valuable in evaluating cities, such as details about pharmaceutical stockpiles, is appropriately guarded.

The experts stressed the need to weigh intangibles, such as how the uniformed services work together as well as the level of commitment from the private sector to participate in training. So while accumulating the data in each of our four categories, CNN.com also provided local emergency management directors with a 10-question survey to evaluate staffing, training and such key questions as:

- Who does your city's emergency management director report to?
  - A. Mayor's office/city or county manager
  - B. Police department
  - C. Fire department
  - D. Other
- The number of your full-time staff dedicated solely to emergency management?
  - A. More than 10
  - B. 5-to-10
  - C. 1-to-5
  - D. 0
- In the current fiscal year, the amount received by your department for terrorism planning, training and equipment?
  - A. More than \$1 million
  - B. \$500,000-to-\$1 million
  - C. \$100,000-to-\$500,000
  - D. Less than \$100,000

All of the data accumulated by CNN.com was later shared with the

local emergency management directors, who were offered the opportunity to comment and report any corrections. In most cases, it required multiple follow-up calls and faxes — but it was crucial that officials confirm the accuracy of data that would be used in the city evaluations.

Eventually, we would provide our empirical data to six law enforcement, security/terrorism and emergency management experts asking them to rate the cities in each of the four categories, using a one-to-five scale (one being best prepared). The Disaster and Terrorism Assessment Survey given the experts was a 22-page document that included key potential vulnerabilities of each city, general city information such as the number of police officers per 10,000 residents and then specific data compiled within our four categories.

Such a scale is recognized by, among others, the National Institute for Computer-Assisted Reporting, as a common grading system within academia, the media and law enforcement. The U.S. Department of Justice provides the same scale for a self-assessment by states applying for domestic preparedness grants. The most comprehensive was hospital/medical, which included 18 items ranging from physicians and emergency outpatients per 10,000 residents to whether the state employed a public health veterinarian and the number of CDC-sponsored "disease detectives."

While reviewing the data, we stressed the importance for the experts to rely on their own experience with the cities. So factored into the evaluation was their awareness that some cities face a greater risk because of such things as location, commerce, landmarks and hosting major events.

The final overall assessment, using Excel spreadsheet software, of the 30 cities was based on the simple average of the experts' ranking, with CNN.com weighting the hospital/medical category 1.5 times because of its importance and the volume of relevant data. Cities were judged as

being most prepared, well prepared, prepared or less prepared.

In the end, only New York, having earlier experienced attacks on the World Trade Center, was judged most prepared. We did a separate story looking at New York's performance and the lessons learned.

The major knock against New York came in the transportation category, where experts cited the potential for gridlock — a factor in getting responders to the scene as well as transporting the injured and ill to medical facilities.

Several other large cities also fared poorly under transportation — notably Boston, Philadelphia and Detroit. The panel judged them overall less prepared, joined by Las Vegas, Milwaukee and New Orleans. CNN.com shared the rankings with city officials before publishing and produced a separate story detailing heightened preparedness as well as steps taken by the individual cities since Sept. 11. Our analysis also found:

- While cities are better prepared than prior to Sept. 11, significant improvement is still needed in public health, emergency response, evacuation planning and inter-agency communication.
- Las Vegas and New Orleans, two major tourist/convention cities, are located in states that do not employ public health veterinarians, who could be helpful in the early detection of a biological crisis. Nor did the cities have a CDC-sponsored disease detective.
- Cities faring best are those like New York who routinely host major events or traditionally face natural disasters, such as Florida and California cities.

The two-month project produced eight stories for the Web site. Web site visitors could rank the cities on different variables and see how their cities compared to others.

Mike Fish can be contacted by e-mail at [mike.fish@turner.com](mailto:mike.fish@turner.com)

# Guarded

continued from page 1

saying that sites were down or information was unavailable.

The U.S. Army's Redstone Arsenal posted this message: "Due to the threat against our nation, our way of life, national security, and because I was told to do so, portions of the Team Redstone homepage have been temporarily blocked from public viewing."

Since that time, some of the information has been returned to government Web sites, much continues to be unavailable and additional information is either being removed or not updated. In addition, proposed legislation and rulemaking could restrict information such as pipeline locations, chemical plant information and other "sensitive" types of data.

Where reporters could once get maps showing pipeline systems, the Office of Pipeline Safety Web site now has this message: "The Office of Pipeline Safety (OPS) has discontinued providing open access to the National Pipeline Mapping System (NPMS). Recent events have focused additional security concerns on critical infrastructure systems."

The agency went beyond removing the information from its Web site. The information is now available only to pipeline operators and local, state, and federal government officials.

## Tracking closures

OMB Watch, a nonprofit organization in Washington, D.C., concerned with freedom of information issues, has tracked federal agencies' removal of information from their Web sites since last fall.

"We saw a pretty dramatic shutdown," said Sean Moulton, senior policy analyst with OMB Watch. "We haven't seen much of a reversal of that. If it didn't reverse immediately...it became stuck amid the debates."

In December, OMB Watch sent Freedom of Information Act requests to federal agencies asking them to list what information they had removed from the Web sites.

The Federal Energy Regulatory Commission refused to comply with the request saying that the mere disclosure of such a list would be problematic. Early removal of information from the agency's Web site was based on size whereby "oversized" documents were removed from the public reference room.

"Literally tens of thousands of documents that were 'FOIable' a year ago or that you could just go get are now exempt from FOIA," Moulton said. "With no change in FOIA and no change in court rulings." Rather, the changes were based on interpretation.

The Federal Aviation Administration responded to the OMB Watch request by saying that they have not taken anything down, Moulton said. Yet the Web site where FAA enforcement data was once available now has a message stating: "The Enforcement Information System (EIS) is not available at this time due in part to security considerations." (See article on page 15.)

Moulton said that one of the most responsive agencies to their request was the Environmental Protection Agency.

"They showed that stuff was taken down and then put back up," Moulton said.

According to Odelia Funke, chief of EPA's Policy and Program Management Branch, the agency went back and did an inventory of database and identified where there were issues. In the end, much of the information was returned to the agency's site. However, access to some chemical information and direct connect access to the agency's Envirofacts Data Warehouse remain unavailable.

"Agencies that took down stuff wholesale are having difficulty deciding what to put back up," Funke said during a symposium of the American Society of Access Professionals in September.

## Other efforts

Moves to take down information were not isolated to federal agencies. Several states removed information from Web sites as well. Pennsylvania removed environmental data from its site; a memo to New York state agencies urged them to review sensitive information; and New Jersey removed chemical information from its Web site.

Efforts have been made by states to exempt terrorism meetings and homeland security agencies from state public records acts.

An Oct. 12 memorandum from Attorney General John Ashcroft told federal agencies to carefully consider "disclosure determinations" under FOIA.

A March 19 memo from White House Chief of Staff Andrew Card ordered federal agencies to reexamine information about weapons of mass destruction. A second memo prepared at Card's request tells agencies to look at three categories: classified, reclassified and declassified, and "sensitive, but unclassified" information. The guidance states "the need to protect such sensitive information from inappropriate disclosure should be carefully considered, on a case-by-case basis."

New Homeland Security legislation also includes tighter restrictions on information dissemination.

In June, Sen. Christopher "Kit" Bond, R-Mo., introduced the "Community Protection from Chemical Terrorism Act," legislation that would restrict access to chemical plants' risk management plans.

Environmental reporters and citizen groups have used that information to assess the risk of chemical plants in their areas. And although the information is no longer online, it is available in EPA reading rooms around the country. However, when a chlorine gas cloud spewed from a rail car near Festus, in Bond's home state, reporters did not have at hand access to the risk management plan for the company where the spill occurred.

But risk management plans do not provide overly detailed information about a plant, Moulton said. "It doesn't say where it is stored, nor indicate what it is stored in."

"There is a risk when industry is not under the watchful eye of the public," Moulton said.

In July 2001, Ralph Haurwitz and Jeff Nesmith did a series of stories in the *Austin American-Statesman* looking at pipeline safety around the country.

As a result of the series, federal agencies with pipeline oversight have increased efforts to update regulatory activities and rulemaking, Haurwitz said. In addition, the Texas Railroad Commission proposed its largest fine ever against a company responsible for an accident in Abilene, Texas.

Much of the data, including pipeline incidents and company financial information, is still available.

But other information concerning pipelines in environmentally sensitive areas is no longer available.

"Today, if a reporter wanted to do a particular analysis dealing with these areas, it would be difficult to do it because the information is no longer public," Haurwitz said.

The potential for closure of these and other types of "critical infrastructure information" worries journalism organizations.

"I think if the government gets its way, it's going to be an awful lot harder to do what we've been doing," said James Bruggers, president of the Society of Environmental Journalists and environmental reporter for *The* (Louisville, Ky.) *Courier-Journal*. "I'm not sure what the overall benefit is. The fact that a lot of this information it has been made public has made communities safer."

Jennifer LaFleur can be contacted by e-mail at [jlafleur@reporterscommittee.org](mailto:jlafleur@reporterscommittee.org).

## FAA withholds data

By Jeff Porter, *IRE and NICAR*

A year ago, right after the Sept. 11 terrorist attacks, IRE and NICAR provided copies of the Federal Aviation Administration enforcement database to well over 100 news organizations. Story after story based on the data provided glimpses of security problems in airports across the United States.

IRE and NICAR had planned to update the database that October, with a routine download from the administration's FTP server. Three days after the attacks, though, the Database Library downloaded the database so journalists could have the most recent information available.

That was the last time the FAA data update was routine.

Shortly after the update, the administration pulled the data from its FTP server, citing security. The data has yet to resurface in its entirety.

After the Database Library's March 29 Freedom of Information Act request, the FAA supplied a copy of the database on June 22 — minus any mention of security or hazardous material violations.

The database still lists actions in some categories, such as flight operation or maintenance problems. As for security or hazardous material violations, the FAA argues that it's none of the public's business.

The official response from Joseph Tintera, FAA's manager of the regulatory support division, says, "sensitive security information has been withheld ..." citing regulations on whether the information is "detri- mental to the safety of persons traveling in transportation."

On July 18, IRE filed a letter asking for that denial to be reconsidered.

IRE argues that since the information resided on publicly accessible Internet servers from at least November 2000 until October 2001, the FAA had set a precedent. Even prior to that, the agency had provided the data on CD.

An additional argument: The request didn't seek records that reveal details of inspections, investigations and alleged violations and findings, but rather only general information, allowed to be disclosed.

Indeed, even before the terrorist attacks, the FAA held back enforcement actions from public release for a year.

Aside from security violations, the FAA denying information about problems with hazardous materials is particularly troubling.

Even before Sept. 11, journalists have used the enforcement database to look at hazardous material violations. In 1999, *The* (Cleveland) *Plain Dealer* published a series examining hazardous materials on airlines.

In July, IRE wrote to the agency: "The public should have access to this information to be informed whether government-regulated entities ... have abided by federal regulations and how government agencies have applied regulations. Failing to provide information to the public concerning mishandling of hazardous materials disregards the public interest and can be detrimental to the safety of persons traveling in transportation."

At this writing, the agency has yet to respond.

Jeff Porter can be contacted by e-mail at [jeff@nicar.org](mailto:jeff@nicar.org)



# Tech tip...

## FoxPro program cleans, standardizes names

By MaryJo Sylwester, *Center for Public Integrity*

Because of the massive size of the State Secrets database (see page 3), we couldn't do cleanup in Access and I was the only person in our office with software powerful enough to handle the data (Visual FoxPro). So I came up with a way to have others help with the cleanup without having direct access.

I exported chunks of the data to Excel files, and had the others put the changes in a new field. Then I wrote a FoxPro program using the Excel copy down functions to apply the changes to the main data file. For further information, read on, then download an Excel file and FoxPro program file that shows this in action (you can open the FoxPro file with NotePad) from [www.nicar.org/techtips.html](http://www.nicar.org/techtips.html).

For example, here's a list of names that needed to be standardized. In this case the first name field needs to be updated to the combined name:

ID	LAST	FIRST
VA556	Smith	John R
VA557	Smith	John
VA558	Smith	John & Jane

CITY	STATE	NEW FIRST
Reston	VA	John R & Jane
Reston	VA	John R & Jane
Reston	VA	John R & Jane

Then I added two new columns, putting the following pieces of code in them and copying down the length of the data:

```
REPLACE ALL NEWFIRST WITH FOR ID=
```

So in this example, the two pieces of code are in columns G and H. The data (as shown above) is in columns A through F. In the next blank column, you concatenate together various pieces to write the code for a simple program. Here's what it would look like for this example:

```
=G2&" "&H2&" "&F2&" "&F2&" "&H2&" "&A2&" "
```

ID	LAST	FIRST	CITY	STATE	NEW FIRST
VA556	Smith	John R	Reston	VA	John R & Jane
VA557	Smith	John	Reston	VA	John R & Jane
VA558	Smith	John & Jane	Reston	VA	John R & Jane

The result for the first line should look like this:

```
REPLACE ALL NEWFIRST WITH "John R & Jane"  
for ID= "VA556"
```

To explain: The ampersand (&) is the tool to tie the pieces together. The first set of quotes provides a space between the word "with" and the "John R & Jane." The next set of quotes (it's four double quotes) is how you tell the computer you want to include the quotes in your answer — something that you need in this case to tell the computer that "John R & Jane" is a text string.

When you copy down the concatenated formula, you get the appropriate ID numbers for each line without a lot of repetitive typing. Then copy just the field that has your answers and paste it into a PRG file in Fox Pro. It will look something like this:

```
REPLACE ALL NEWFIRST WITH "John R & Jane"  
for ID= "VA556"  
REPLACE ALL NEWFIRST WITH "John R & Jane"  
for ID= "VA557"  
REPLACE ALL NEWFIRST WITH "John R & Jane"  
for ID= "VA558"
```

You can use this same trick to apply changes to more than one record at a time by looking for keywords. For example, we used this for coding the expenditures. I created a list in Excel of possible keywords that we might find listed in the purpose field (most of the list came from a group-by query on the purpose field).

We had previously decided what code each keyword would fall under, so I assigned those in the Excel file. Here's what it looked like:

MAIL	A4
SALARIES	A2
HOTEL	A1
AIRFARE	A1
OFFICE SUPPLIES	A5
POLLING	C4
DIRECT MAIL	C6

Then, in the Excel file, I added the FoxPro commands, but this time using wildcards. Here's what the command looks like when completed:

```
REPLACE ALL PURPOSE WITH "A4" FOR  
"MAIL" $ PURPOSE
```

The dollar sign symbol tells FoxPro to find the term "MAIL" anywhere in the field called PURPOSE.

The coding program ended up being more than 400 lines long because I had so many possible keywords. I kept the Excel file throughout the coding process, and that proved to be worthwhile when we decided to tweak codes.

Instead of having to sort through the whole FoxPro program file, I simply went back to the Excel file, sorted or filtered it to find the codes I wanted (there might be anywhere from two to 50 lines for any given code) and made the changes. Then I re-copied the formula and pasted it into a new program file. It took maybe five minutes to make hundreds of changes.

# Cronyism

continued from page 1

he scored nearly \$300 million in school business for his clients, including a questionable contract that landed him a \$4 million payday.

During an eight-month investigation of Miami-Dade County Public Schools, investigative reporting team colleague Ronnie Greene and I used an array of databases and computer-assisted reporting techniques to document how taxpayer dollars often benefit lobbyists, contractors and former school board members rather than the district's 365,000 children. Our goal: to bring strong government reporting into the classroom.

The *Herald's* review of a decade's worth of district spending helped explain why morale is low and how politics pervades the school system.

We used three key databases to follow the money as well as Internet resources, including tax documents, Securities and Exchange Commission filings, property records and state corporate records. We also used mapping software and SAS (a statistical and database package) to explore the district's burgeoning charter school movement. Coupling the CAR work with traditional reporting, we were able to present the most comprehensive picture ever of how the struggling district does business.

## A triangle

Miami-Dade's public school system has a \$4.1 billion budget, about the same as county hall. But unlike county hall, the district had no independent inspector general with power to probe spending. It was a perfect niche for our newspaper to fill.

Conceptually, our approach to following the money was a triangle. First, we needed to know with whom the district did business and the money involved. Second, we had to track which lobbyists companies hired to score contracts. And finally, we had to know how much

campaign money lobbyists and companies contributed to school board members. Much of the data, we knew, only was on paper.

We requested a decade's worth of paper campaign contributions from the county's elections department. Turning 10 years of paper into a database was not cheap, but it was indispensable. The bill ran about \$4,000. We sold it to our editors by pointing out the ongoing benefits of having the information. Once we had the database, we loaded it into SQL Server and created ASP applications that allow reporters to search contributions on our Intranet. It's a resource we continue to utilize. Now when a big contract goes before the school board, reporters can search for corporate officers and lobbyists to see how much they've doled out. And because there are fewer records, updating the database each quarter is relatively cheap.

Although lobbyists abound at county hall, we found companies that get business from the district hire a small group of lobbyists with powerful connections. As one of the lobbyists put it, the school system was "the best-kept secret in town." So we created another database from scratch using the district's lobbyist registration forms. Again, we looked at a 10-year period, this time inputting most of the records ourselves.

With the lobbyist database, we zeroed in on the three busiest contract-pushers and found a former superintendent who lobbied his ex-employees months after leaving the system. Because the district keeps most of its electronic information in arcane mainframe databases, obtaining the final point of our triangle wasn't easy. When we requested a copy of the purchasing database, they came back with a price of more than \$6,000 and said it would take a month to process. There was no way we could sell that to our editors. So we negotiated. Turns out the school district's IT department had a stripped-down subset of the data for auditing purposes. It cost us about \$300 and included all the important information: vendor name, item de-

scription, number of units, cost of per unit, total amount of purchase, date and a purchase order number we could use to trace electronic records back to paper.

I also learned a cardinal lesson. Be flexible and open to ideas from people who handle the information you seek. Just be sure to get a record layout of any database you're interested in first. Before you can negotiate, you have to know what's in play.

Once our triangle was complete, we could connect school board members, contracts and campaign contributions. When documenting how the cigar-chomping lobbyist made \$4 million off a questionable insurance contract, we were able to quantify his influence.

## Web resources

From the outset of reporting the series, we learned that a nonprofit organization, founded by a former school board member, made millions from the district by serving high-risk children. One of our first moves: going to [www.guidestar.com](http://www.guidestar.com) to find the company's federal nonprofit financial disclosure forms, a.k.a. Form 990s. GuideStar contains a wealth of information for journalists and is often featured in NICAR training. By coupling the tax filings with property records, we learned that the former school board member was the nonprofit's landlord and charged the charity \$1 million in rent during a two-year period.

Privatizing public education is big business in Florida, and Miami-Dade County is no exception. Edison Schools Inc. was the first private management company hired to run a public school in Miami-Dade. It also partnered with the powerful teachers union to win approval for nine new charter schools. Yet our newspaper never took a hard look at the company. We did not know whether Edison did a better job educating children or how much money it made from the district.

Using school audits and SEC filings available on [www.sec.gov/edgar/](http://www.sec.gov/edgar/), we some found answers. For example, al-

continued on page 18

# Cronyism

continued from page 17

though the publicly traded company claimed they had yet to make a profit, SEC filings showed Edison had lent its CEO \$8 million to buy stock just before landing a huge contract in Philadelphia. When we asked how much of the school's budget went to the company, they said between 8 percent and 12 percent. Yet district audits showed the number was more than 20 percent.

Searching for documents online isn't the sexiest CAR maneuver, but it's still an important avenue of inquiry.

## Veil of competition

We used more complicated CAR tactics to explore charter schools, which are built and managed by private companies using public money. Charter school proponents claim they create better schools by forcing districts to compete for kids. If parents and students aren't pleased with the public school, they can take their tax dollars and move to a charter. In a state where schools are underfunded, this is a big issue.

Among Miami-Dade's charter advocates: a former superintendent who now runs the country's second-largest charter company and a former school board member doing consulting work for the same corporation.

First, we were interested in what the charters look like. We used a database from the state to determine the schools' demographics, including race, ethnicity and income level. We found that although 70 percent of public elementary school students are eligible for free or reduced-cost lunch, about 40 percent of charter students are.

Meanwhile, the charter schools were claiming they score better than public schools on Florida's standardized test. Building off great school test analysis done by the *Palm Beach Post*, we used a statistical method called multiple regression to show a more complex picture. We ran regression methods in SAS called STEPWISE and FORWARD and found the higher a school's percentage of students eligible for free or reduced lunch, the lower its test scores. When we took the percentage of low-income students into account, many of the

## readme.txt

Read *The Miami Herald* "Cheating the Classroom" stories on the Web at [www.miami.com/mld/miamiherald/3024226.htm](http://www.miami.com/mld/miamiherald/3024226.htm).

charters did worse than public schools.

Reading charter studies in education journals, we came across one that used mapping software to compare the demographics of charters to adjacent public schools. From the state statutes we already knew that there was no charter money earmarked for transportation, so most attracted students from nearby.

We used mapping software called ArcView to geocode the address of every school in the county and joined schools' demographic info and 2000 Census data to the maps. Geocoding turns street addresses in data tables into points on a map. Then we drew circles around the charters to capture nearby public schools and neighborhood demographics. The result: Charter schools were vastly different from adjacent public schools and many looked nothing like the neighborhoods they supposedly served.

All charter schools must sign a contract with the district outlining which students they planned to serve and how. For example, one contract stated that it would serve students "within a reasonable distance, not greater than four miles from the school" and "achieve a racial/ethnic balance reflective of the community." Yet our analysis revealed that while the neighborhood was 80 percent black, the school was 26 percent black. And, although other schools within four miles had populations that were 75 percent or more low income, the charter school had less than 25 percent. We helped put a strong focus on which "community" the charter served.

Jason Grotto can be reached by e-mail at [JGrotto@herald.com](mailto:JGrotto@herald.com)

## 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.

#### Workshops are scheduled for:

Nov. 16 — Fort Lauderdale, Fla.  
Nov. 23 — Wichita, Kan.  
Feb. 22, 2003 — Phoenix, Ariz.  
March 22, 2003 — St. Petersburg, Fla.  
March 29, 2003 — Columbia, Mo.  
March 29, 2003 — Minneapolis, Minn.  
April 5, 2003 — Cleveland, Ohio  
April 12, 2003 — Spokane, Wash.  
April 12, 2003 — Long Island, N.Y.

For more information, visit [www.ire.org/training/betterwatchdog](http://www.ire.org/training/betterwatchdog).

To request a workshop for your area, contact Executive Director Brant Houston at [brant@ire.org](mailto:brant@ire.org).



# Education

continued from page 1

## Stories, education, data

That's just one idea of how to examine education. As this issue of *Uplink* shows, more ideas abound.

- James E. Wilkerson of *The* (Allentown, Pa.) *Morning Call* describes how he built a database that showed that a new tax-credit program in Pennsylvania funneled millions of tax dollars to religious schools with little state oversight. Instead of helping needy kids in failing public schools, as intended, much of the money went to children already in private schools.

- Scott Finn and Eric Eyre of the *Charleston* (W.Va.) *Gazette* discuss how they built a database of school bus schedules in rural West Virginia. Their investigation found that thousands of children rode school buses at least two hours a day, in violation of state laws and guidelines.

- Jason Grotto of *The Miami Herald* explains how he used a variety of databases to expose questionable spending by Miami-Dade school officials that benefited lobbyists, contractors and former colleagues but did little to help students. In one case, a lobbyist made \$4 million from a single deal.

## Data at all levels

Fortunately, journalists can find lots of education data on a local, state and national (in a few cases, even international) level. To find helpful federal data, try the National Center for Education Statistics ([www.nces.ed.gov](http://www.nces.ed.gov)), a branch of the U.S. Department of Education. The most comprehensive database is the Common Core of Data. It has basic information on every public school and school system in the United States, including student enrollment and demographics, finance data, teacher and staff counts, and graduation and attendance rates. Files are updated annually and can be downloaded from the Web.

The data can help answer lots of questions: How much do school districts

spend on each student, on average? How much education funding does the state provide? What is the teacher-pupil ratio? Which school systems have a high percentage of minority students, and how has that changed over time?

For higher education data, NCES has the Integrated Postsecondary Education Data System. It gives basic information on colleges and universities, including student enrollment, tuition charges, degrees offered, admission requirements, number and type of degrees awarded, finance data and faculty salaries. These files, too, are updated yearly and can be downloaded.

Another branch of the Education Department, the Office for Civil Rights ([www.ed.gov/offices/OCR](http://www.ed.gov/offices/OCR)), keeps data that let reporters look for possible inequities in schools and districts. For example, the OCR tracks enrollment in gifted and talented classes, special education and Advanced Placement courses by race and gender. It looks at discipline rates for students with disabilities, and participation in sports by gender.

Don't forget the U.S. Census Bureau ([www.census.gov](http://www.census.gov)), which tells how many kids go to public and private school and how much education adults have. It's also possible to look at demographic data by school district.

## Digging for state info

Many state education departments keep useful and more current data on schools and school districts. Find out what information the state collects, when it's collected, and how it's stored. Some states put downloadable files on their Internet sites. For instance, the California Department of Education ([www.cde.ca.gov](http://www.cde.ca.gov)) has files on enrollment, dropouts, graduates, teacher qualifications and credentials, high school graduation requirements and English proficiency.

If you can't find what you're looking for, call the state. They might have it, just not online. When I covered education in Missouri, I could download files from the state's education Web site

([www.dese.state.mo.us](http://www.dese.state.mo.us)) with test scores and the percentage of low-income students by school. For other data, such as dropout or graduation rates by school, I called the state's data experts and had them either e-mail me the files or post them on their FTP site.

Those who are familiar with spreadsheets and databases might want to try more sophisticated approaches to covering education. Many newspapers across the country don't just report school test scores; they take family income and other variables into account. This can help level the playing field between schools in affluent neighborhoods vs. those in poor ones, since family income has such a strong effect on test scores (the poorer the children, the worse they tend to do on tests).

Holly Hacker, formerly an education reporter for the *St. Louis Post-Dispatch*, can be contacted by e-mail at [holly@nicar.org](mailto:holly@nicar.org).

## readme.txt

### Resources

The IRE Resource Center provides stories and tipsheets to help journalists. Search the story and tipsheet databases at [www.ire.org/resourcecenter](http://www.ire.org/resourcecenter). Call 573-882-3364 to order these and other investigative stories and tipsheets.

### Tipsheet

"Measuring Equality in Schools:" Doug Haddix of *The Columbus Dispatch* explains how to analyze a school's racial composition, income levels, and changes in attendance by using a district's student and teacher computer records. The tipsheet shows how Haddix used data from the district accounting system to look at spending. *The Dispatch* published a four-day series called "Dividing Lines" in June 2000 with the material. (See Tipsheet #1518)

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 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 Tiefenbach

**Copy Editor**  
Pia Christensen

**Contributing Editors**  
Megan Christensen  
Jaimi Dowell  
Holly Hacker  
Stephanie Kang

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, tip sheets and  
data analysis.

**Editorial**  
573-884-7711

**Subscriptions**  
573-882-2772

**Director of Publications**  
Len Bruzese  
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.*

