January/February
2001

Uplink

CHILD WELFARE

Measuring foster care

By Jason Method
Asbury Park Press

Despite New Jersey's child protection agency's two-year, $128 million reform effort, children who are wards of the state are spending more time in foster care, group homes and institutions than just a few years ago. That was the key finding when the *Asbury Park Press* decided to find out how much, if any, progress the state had made in its reform goals.

Like in many states, the child protection agency is veiled under confidentiality laws. To make reporting more difficult, New Jersey has one of the weakest public access laws in the nation. DYFS officials turned down my request for some dozen statistical reports that would have documented the agency's performance.

Fortunately, I learned of a database kept by the federal government that can pierce through the confidentiality laws and bureaucratic resistance. The database tracks the some 600,000 children, in all but a handful of states, who are under state foster care or are in other state facilities. CAR journalists may use the database to evaluate state child protection agencies through the systemic criteria used by child welfare experts and to track that performance over time.

The Adoption and Foster Care Analysis and Reporting System contains detailed information on every child who has entered care in a federal fiscal year. The foster care data includes the date of birth, sex and race. It gives other important dates, such as the date the child last entered state care and the date the child was discharged. It provides reasons why the child was removed and indicates any physical disability.

AFCARS, which comes only in an SPSS or SAS format, also has variables tracking how many foster homes or other state facilities the child has been placed in, how many times they have been removed from their natural home, and what type of placement — such as foster home, group home or institution —the child was last reported in.

A second database that comes with the AFCARS data documents adoptions of former foster children or children placed by the state agencies across the country. It contains the same biographical and date information as the foster care database.

Names and addresses are excluded from the datasets and dates of birth are altered slightly for confidentiality reasons. But that

Continued on page fourteen

CENSUS 2000

Race problem solutions

By Steve Doig
Arizona State University

A year ago (see the January/February 2000 *Uplink*) I wrote about the looming problem of making valid comparisons of the racial and Hispanic origin data from 1990 and 2000. In that piece, I followed journalism's great thumbsucker tradition of raising an alarm but offering no solid solutions. Well, the problem is no longer looming – it's here, so it's time to focus on solutions.

The problem, of course, is that in the 2000 Census, respondents were allowed to check off as many racial categories as they wished, whereas respondents in 1990 and earlier could only pick one race to describe themselves. This meant that the ten possible 1990 race and Hispanic origin categories (ranging from "Non-Hispanic white" to "Hispanic Other") have ballooned to a total of 126 possible

Continue on page nineteen

Inside Uplink

This issue of *Uplink* contains a cornucopia of topics, ranging from census data to gambling.

Gambling

Barnet Wolf of the *Columbus Dispatch* used a variety of data sources to study gambling trends in Ohio.

SEE PAGE EIGHT.

First Ventures

Randy Travis of WAGA-TV in Atlanta tells about his station's first foray into computer-assisted reporting: a story that revealed convicted felons working as certified nurse aides in Georgia nursing homes.

SEE PAGE FOUR.

Jason Callicoat of the *South Bend Tribune* tells about his attempt to recreate a NICAR Bootcamp favorite, parking tickets and bad weather.

SEE PAGE FIFTEEN.

Mapping

John Maines of the *South Florida Sun-Sentinel* shares tips on setting up an inexpensive, simple mapping system that can handle any newsroom task.

SEE PAGE ELEVEN.

Uplink

January 2001

Volume 13, Number 1
A newsletter of the National Institute for Computer-Assisted Reporting

EDITOR
Brant Houston

MANAGING EDITOR
Mary Jo Sylwester

ASSOCIATE EDITORS
Mike Sherry
Amy Sherrill

ART DIRECTOR
Kerrie Kirtland

NICAR is a joint effort of Investigative Reporters and Editors and the University of Missouri School of Journalism.

NICAR services include hands-on newsroom training in computer-assisted reporting, special academic and advanced training in data analysis.

DIRECTOR OF PUBLICATIONS
Len Bruzzese

SUBSCRIPTION ADMINISTRATOR
John Green

Uplink is published every month by the National Institute for Computer-Assisted Reporting, 138 Neff Hall Annex Columbia, MO 65211. (573) 882-0684. Subscriptions are $40 for IRE members, $60 for nonmembers.

Postmaster: Please send address changes to NICAR. Send e-mail to jgreen@nicar.org

IN THE NEWSROOM

CAR newsletters

By Richard Dalton
Newsday

So you do computer-assisted reporting at your news organization and you've decided that with all the spare time on your hands, you might as well start an in-house newsletter to spread the CAR gospel to the rest of the newsroom?

Great. You've joined a handful of other newsrooms.

Even with a limited amount of time, you can produce a CAR newsletter that is both practical and inspirational. Here are some tips on creating CAR newsletters. CAR enthusiasts sent in some tips, while I culled others from reading newsletters.

For some examples of in-house newsletters, including Car Talk, Byte Me and Mouse Droppings, check out *www.kricar.com.*

This-Is-How-I-Did-It articles

The best way to provide both practical advice and inspiration is the TIHIDI article — the This-Is-How-I-Did-It piece. *Uplink* is chock full of TIHIDI articles, and for good reason. Such pieces show how you can replicate the story in your area and help you brainstorm. And the articles offer the Wow factor: "Wow, it's amazing what you can do with computer-assisted reporting."

Even with a limited amount of time, you can produce a CAR newsletter that is both practical and inspirational.

Even a small in-house newsletter should contain at least one TIHIDI article. And the best part is that you can delegate this task: Ask other staff members to write articles describing how they wrote a CAR story. If you're part of a chain, ask other reporters in your chain to contribute.

The piece should briefly describe the CAR story, indicate what software was used, what databases were analyzed and how they were obtained. A box accompanying the story should list some of the key fields in the database.

If CAR reporting is quite active in your newsroom, run a roundup of CAR stories.

CAR Q&A

Answer questions that reporters submit to the newsletter. Or if you help out another reporter in person, turn the help session into a Q&A for the newsletter.

Software Tips

Microsoft Excel and Access are the most commonly used software tools in CAR. Focus on those products initially, starting with simple tips: subtotals, importing data from HTML documents, etc.

Don't let 'em lose it

You know how easy it is to lose something amidst the mounds of papers on a reporter's desk. So, Jennifer LaFleur, of the *St. Louis Post-Dispatch* recommends printing the newsletter on obnoxiously bright paper, so it won't get lost in the shuffle. In her previous job in San Jose, she gave out binders for reporters to store tip sheets and copies of the newsletter.

Point out CAR resources

Highlight useful Web sites for finding anything from phone numbers to government data. What databases does your news organization have? Focus on a few, listing possible story ideas. List CAR resources reporters can turn to, including possibly yourself. Does your news organization offer training in Microsoft Excel or Access? List the schedule of classes. Or hold some classes yourself. Also point out books or videos on CAR software. Do you have tech-support people who can answer questions about software? If so, point reporters in their direction (and save yourself some time.)

Highlight Math Tips

Include tips on calculating percentage change, per-capita rates, inflation adjustments, etc.

How often to publish

Janet Roberts of the *St. Paul Pioneer Press* went from writing all of the articles to relying almost entirely on contributions from re-

Continued on page sixteen

Historical data sources

By Neill A. Borowski
The Philadelphia Inquirer

My squirrel-like habits of burying historical census data like acorns to be uncovered in the future began in the days of paper.

And my prized "database" is a booklet called Recent Historical Trends, published by the Philadelphia City Planning Commission. How "recent" are the trends? The report was published in 1966 and long ago fell apart (it's held together with a binder clip). But that report is packed with detail on population and race in parts of the city going back to 1930.

Should you be as obsessed? If you're so inclined, your newspaper could benefit greatly from historical local census data online – available well in advance of the flood of 2000 data.

Electronic data archives should be easier to build than paper ones. But they aren't.

I discovered this recently when I began to ask around about the availability of 1970 and 1980 census data in computer form. A typical answer from one agency in Pennsylvania that hands out such data was "We don't have anything that old."

If you build a historical census archive for your area, you may have the only such archive around. You will be building something useful to reporters and editors that will be around long after you're gone.

However, you may find yourself behind closed doors – as I have been – reconciling an ArcView census tract map on your computer screen with an old paper map from 1950. Yet, imagine how powerful a graphic we'll have to accompany a story on population change in a given Philadelphia neighborhood: 50 years of change, from 1950-2000.

You may not want to get so nerdy that you include every census tract from 1960 in your archive. But you may want to have the populations of minor civil divisions (townships, cities, boroughs) or counties going back 50 or more years.

A few ideas to get started in your quest to pin the wonk-meter's needle:

• Visit your local library. Tucked away on shelves you'll often find dusty printed volumes of census results going back decades. Photocopy the pages and have the data input to build a database.

• Stop by the "United States Historical Census Data Browser" at *http://fisher.lib.virginia.edu/census/*. You can pull county-level data going back to the first decennial census in 1790.

• Purchase the 1980 and 1990 data in bulk from the Census Bureau. Go to *www.census.gov/mp/www/rom/* to check out pricing and availability. For example, the 1980 STF-3A file for the nation costs $1,500. However, the cost plunges to $150 if you're after only one state (the CDs contain several states each). The 1990 STF-3A file for the nation costs $4,340. Some single states fit on one CD at $100 each. Larger states, such as Pennsylvania, fill several CDs. Pennsylvania, for example, is a three CD set costing $225. Texas is a four-CD set costing $300. The PL 94-171 file for 1990 costs $550 for the nation (10 CDs) or $100 for a single CD.

"Your newspaper could benefit greatly from historical local census data online - available well in advance of the flood of 2000 data."

• Buy the 1970 block-level data from the National Archives at *www.nara.gov/nara/electronic/cen1970.html*. The price ranges from $90 for a state to $1,315 for the nation.

• Purchase 1940 through 1970 tract-level data from the National Archives at *www.nara.gov/nara/electronic/bogue.html*. This dataset, called the "Bogue File," was created by keypunching paper data as part of research at the University of Chicago. Not all areas of the country are represented, although the largest cities are in the data. Each year is $80 on nine-track tape.

• Buy the CensusCD from GeoLytics Inc. (*www.geolytics.com/*). National data for a full year is available on one CD. This private firm's slogan is "Making Demographics Easy to Use!" and using the CensusCDs is far easier

Continued on page seven

Sources for historical census data:

United States Historical Census Data Browser, *http://fisher.lib.virginia.edu/census/*

U.S. Census Bureau, 1980 & 1990 data, *www.census.gov/mp/www/rom/*

National Archive, tract level data, *www.nara.gov/nara/electronic/bogue.html*

GeoLytics Inc., *www.geolytics.com/*

CIESIN, *www.ciesin.org/*

OSEDA, *www.oseda.missouri.edu/.*

Felons and nurses

By Randy Travis
WAGA-TV, Atlanta

We've worked long-term investigations here before, but nothing compares to my station's first venture into computer-assisted reporting. As it turned out, the story was well worth the wait. Even if it took nearly a year.

Our investigation began with an issue some reporters in other parts of the country have already tackled: people with criminal records working in nursing homes.

What intrigued us here in Georgia is a unique twist in our law. The state certifies nursing home workers, called certified nurse aides (CNA). Then, Georgia nursing homes must do a criminal background check before they are hired. But the law doesn't preclude the nursing home administrator from hiring a convicted felon. With staffing shortages so critical, we suspected CNAs with criminal records were still being hired. Turns out, we were right.

We started with separate open records requests to the state Department of Human Resources (DHR) and the Department of Corrections (DOC). DHR maintains a database of all certified nurse aides, both past and current. DOC keeps track of all criminal convictions, both felony and misdemeanor. We got the complete DOC database without any trouble, and no charge. Our first roadblock came in getting what we needed from DHR.

We felt Social Security numbers provided the most accurate way to match up names between the two databases. No problem, DOC told us. They gave us everything. But DHR balked at the idea of providing us what they considered private information. Our state's open records law does not directly address this issue. It took four months of phone calls to the attorney general's office and DHR before we finally convinced them the information was subject to the open records law. The disk cost us only $10. But by then, our plans for a May sweeps series were gone.

Dirty data

It didn't take long, though, for us to realize that the hard work hadn't even begun. The data from both sources were dirty. In one table, the Social Security numbers were separated by dashes. In the other, the numbers weren't separated at all. First and last names were in separate fields in one ... but combined in the same field in the other.

That's why we turned to NICAR. They were able to clean up the data and do the match between databases. Of the 73,410 nursing aides, NICAR found 5,595 had a criminal record. But there was more work to do.

The CNA database included people who hadn't been certified in years. We didn't want them included in this investigation. Secondly, some of the CNAs committed their first crime after they were certified. Can't blame a nursing home for that. Yet they still showed up as a match.

So, we began eliminating matches. Using Microsoft Access, we first kicked out all CNAs whose certification expired before 2000. That gave us a list of current criminal CNAs. Then we took out all CNAs who were convicted after their certification date. That eliminated all those CNAs who had clean records when they were hired, but committed a crime later on. We wound up with 219 felons and 2,942 probationers currently allowed to work in Georgia nursing homes. Now it was time to find them.

We felt Social Security numbers provided the most accurate way to match up names between the two databases. No problem, DOC told us.

Georgia does not keep track of where certified nurse aides actually work. So it was impossible for us to say how many of the criminal caregivers actually had access to the elderly. The DHR database did include the CNA's address at the time of certification. So, we started there. Thanks to the Web site healthgrades.com, we used those addresses to get a list of every nursing home nearby. Then we started calling asking to speak to some-

Continued on page five

The WAGA-TV Story "Criminal Caregivers," will be available soon in the IRE Resource Center at *www.ire.org/resourcecenter/*

Data: certified nurse aides **database**
Source: Georgia Department of Human Resources

Data: criminal convictions
Source: Georgia Department of Corrections

Software used: Microsoft Access

See sample slices of the data WAGA-TV used on the NICAR Web site, *www.nicar.org/techtips.html*

Continued from page four:

Felons

one to verify employment. When we got a hit, we confirmed it by Social Security number. We focused on the small communities that didn't have as many nursing home choices. We found more than a dozen criminal caregivers that way.

Crimes at nursing homes

The other part of our investigation looked at whether criminal caregivers are actually more likely to commit crimes once they get inside the nursing home. The state Department of Human Resources also maintains an abuse registry. When a CNA harms a resident, they wind up on the registry. We merged the registry with the Department of Corrections table and found 82 criminals. Many of them committed the same crime inside the nursing home that they had on their record when they were hired: burglary, assault, theft, etc.

It took us even longer to track down those victims or their families. By law, they're not listed in the abuse registry. That meant contacting local police departments and asking for incident reports of those crimes. But often, the nursing home didn't bother to call the police. Months later, we managed to find family members to react to our investigation. They had no idea the nurse aide who had harmed their elderly relative was a criminal before he was ever hired.

Our series finally aired November 8 and 9, 2000, nearly a year after we made our first request for the data. The investigation is expected to prompt a change in the law when the state general assembly meets in January.

Unfortunately, in an ironic twist, state lawmakers may also vote to make Social Security numbers exempt from the open records law. If you're planning a similar investigation, fight for that key data. Without it, our investigation would certainly have taken even longer.

Randy Travis can be reached by e-mail at rtravis@fox5atlanta.com

TECH TIP

Enterprise joins

By Ron Nixon
IRE and NICAR

Enterprise joins – matching databases that weren't designed to be matched – is one of the most powerful features of relational database managers like FoxPro, Access and Paradox. Using enterprise joins, reporters across the country have successfully matched databases of school bus drives against people with driving convictions, found dead people voting in elections, felons with hunting license, and deadbeat parents in state government.

When seeking databases, reporters should always think about ways they could join databases from different agencies or across different fields of study to get stories. Unfortunately, databases rarely have identifying codes like Social Security numbers or some form of employee id. In cases where you don't have this type of information you have to come up with proxies. Some identifiers could be birth dates or address like the person's street number, city or zip code.

Although enterprise joins can produce great stories they can be fraught with danger. A wrong match can give you incorrect information. You could, for example, get a match for a deadbeat parent working at the agency that collects delinquent childcare payments. But since you only matched on the person's name, the join could be wrong.

Always remember that a name is never a unique identifier. How may Ron Nixon's are there or John Smith's? There are probably thousands. Another thing to remember is that the names have to match in both. Ron Nixon is not the same as Ronnie N. Nixon to the database, even though they might be the same people. You will have to standardize the names to get the matches you want.

Finally, always verify, verify, and verify that the name you have is really the person you think it is. The join is just the start. Once you have that the real reporting begins. This not only ensures that you get the story right, it also protects you and your paper from a potential lawsuit.

Ron Nixon can be reached by e-mail at ron@nicar.org

Stories

For more on matching felony and caregiver databases, take a look at these stories available from the IRE Resource Center, (573) 882-3364

Story # 15823.
The Record (Hackensack, N.J.) reports that "New Jersey has allowed numerous thieves, drug dealers, and violent offenders to work as home health aides - no questions asked. In nearly every county, *The Record* found criminals — fully certified by the state — working alone in the homes of cancer survivors, the elderly, and the infirm, their pasts hidden from vulnerable patients."

Story # 15102.
The Denver Post **examines four suicides that occurred at a state mental health institute. Abusive employee reports led into findings of their criminal histories. Many of these employees were on work-release programs.**

Valuable accident data

By Hurst Laviana
The Wichita Eagle

Wayne Louden has been involved in 37 traffic accidents since Jan. 1, 1990, making him the most accident-prone driver in Kansas. Thomas Spicer has been involved in 27 wrecks, 21 of which resulted in inattentive driving tickets.

Louden and Spicer were key players in a series of stories about accident-prone drivers in Kansas. The series was based on a Kansas Department of Transportation 1990-1999 database that covered 703,000 traffic accidents involving 1.1 million drivers.

Acquiring the database was a struggle, but it has proved to be one of the most valuable data sets in our newsroom. It has yielded stories about the growing number of car-deer accidents in the state, the reluctance of drivers in the nation's "beltless belt" to use seatbelts and the fact that accidents involving police chases doubled in Kansas during the 1990s. The data also allows us to quickly generate lists of accidents that have occurred at any intersection in the state.

Acquiring the database was a struggle, but it has proved to be one of the most valuable data sets in our newsroom.

But the biggest reader reaction came from the story about accident-prone drivers. Many were surprised to learn they were sharing the roads with 20 Kansas drivers who had 10 or more accidents in the 1990s. And some of those drivers had dozens of traffic convictions.

How did they keep driving?

It turned out that the Kansas Division of Motor Vehicles hasn't been enforcing a law that allows it to suspend the license of any driver convicted of three moving violations in a 12-month period. Instead, the DMV only takes action after the fifth conviction.

It also turned out that Spicer's inattentive driving tickets – as well as 70,000 other inattentive driving citations issued each year by Wichita police – don't even count against a driver's record. Inattentive driving is a ticketable offense in many Kansas cities, but state law does not prohibit it. So the state doesn't count it.

Doing the stories

The seeds for the story were planted more than a year ago when Crime & Safety team leader Jim Lewers asked KDOT for five years of accident data. KDOT mailed him copies of the 1990 through 1996 annual traffic accident reports.

Lewers tried again, this time asking for a comprehensive electronic database with detailed information about each accident. KDOT's response: for $51,710 we could have the data – minus the drivers' names and accident locations. KDOT, which wanted 5 cents a line for just more than 1 million lines of data, said giving us the names and accident locations could result in the information being used against it in court.

KDOT held its ground until February, when a judge ruled that the agency was improperly withholding railroad crossing accident data from another Kansas newspaper, the Garden City Telegram. KDOT eventually gave us the accident data – names and all – for $87.11.

Crunching numbers

The file arrived in the newsroom on a CD-ROM in the form of a 550-megabyte Microsoft Access database, which included separate tables for such variables as drivers, accidents and vehicles. My three-gigabyte hard drive was no match for that much data, but we did turn around some quick stories using pieces of the data.

On March 22, the day the Kansas House of Representatives approved a bill designed to cut car-deer accidents by increasing the number of out-of-state hunting permits, we sliced the deer accidents from the accidents table and loaded them into an Excel spreadsheet.

Excel's Chart Wizard produced an interesting graphic that showed a steady rise in

Continued on page seven

"Bad Drivers," is available online at *http://web.wichitaeagle.com/content/wichitaeagle/2000/10/02/special/driverslist1002.htm*

It is also available in the IRE Resource Center, story #16839.

For more on the *The Wichita Eagle's* story see the January/February 2001 issue of the *IRE Journal.*

Software: Microsoft Access and Excel

Data: Kansas traffic accidents, 1990-1999
Source: Kansas Department of Transportation

Continued from page six:

Accident data

the number of car-deer accidents. It also revealed a sizable blip every November, when the males were out seeking mates, and a smaller blip every May, when females were out looking for places to give birth.

Our lead: "Cars, trucks and buses crashed into more than 10,000 deer in 1999, a record for the state."

Once a bigger hard drive was installed in my computer, we set out to find the accident-prone drivers. We needed the extra room in order to link the tables so we could track such variables as injuries, fatalities and accident causes.

A single "group-by" query on the driver's license field generated a list of hundreds of Kansans who had been involved in three or more accidents during the decade. The process was complicated by the fact that Kansas changed its driver's license numbering system in the mid-1990s, so most drivers had two driver's license numbers.

Other "group-by" queries on the firstname, lastname and DOB fields turned up another batch of candidates. We ended up working with a list of 1,960 drivers who had five or more accidents in the 1990s. Their 11,165 accidents left 50 people dead (including 14 of the accident-prone drivers) and 5,171 injured.

We eventually decided to focus on the six drivers who had 12 or more serious accidents during the decade.

Doing the legwork

The next step was to ask KDOT for hard copies of the accident reports for the six drivers. We also asked local law enforcement agencies for copies of reports for non-injury accidents involving less than $500 in damage – accidents that aren't reported to the state.

The written reports, which often had narratives and diagrams, included more detailed information than the KDOT database. Based on this information, reporters interviewed drivers.

The series ran in late September, and a week later we put the list of drivers with five or more accidents on our Web site with a disclaimer noting that the listed drivers were not necessarily at fault in any of the accidents.

The only negative reaction from readers came on the following Wednesday, when a woman who works at City Hall called to say our on-line information was wrong – that she hadn't been involved in six accidents. When given the dates and locations she retreated, then insisted that none of the accidents was her fault.

Hurst Laviana can be reached by e-mail at hlaviana@wichitaeagle.com

Continued from page three:

Data sources

than rummaging through the Census Bureau's data disks. However, queries using the CensusCD set were designed to go through the front end.

You can extract bulk data, but it can be cumbersome. If you can't afford mapping programs, the CensusCD+Maps product does a solid job of making thematic maps (the maps can be exported to ArcView Shape and MapInfo MIF files).

Costs: CensusCD+Maps (to block group geography), $500; CensusCD Blocks, $1,000 for nation; $500 for a single state; CensusCD 1980, $800 each; $400 for a single state. The company also has announced a CensusCD1970, although the introduction date and pricing weren't available at press time.

• Other useful sites for census data include Center for International Earth Science Network (CIESIN) at *www.ciesin.org/* and Office of Social and Economic Data Analysis at the University of Missouri (OSEDA) at *www.oseda.missouri.edu/*. Also check the site for your state data center *(www.census.gov/sdc/www/)* or state, regional and local planning agencies.

Neill Borowski can be reached by e-mail at nborowski@phillynews.com.

Computer-Assisted Reporting Boot Camps
Missouri School of Journalism
Columbia, Mo.

- **March 25-30, 2001**
- **May 20-26, 2001.**
- **July 15-20, 2001**
- **August 5-10, 2001**

CAR Boot Camps give journalists a jumpstart in computer-assisted reporting techniques. These unique seminars train journalists to acquire electronic information, use spreadsheets and databases to analyze the information and to translate that information into high-impact stories. In addition, the institute then provides follow-up help when participants return to their news organizations.

More information, including a registration form, is available at *www.ire.org/training/bootcamps.html*, or call (573) 882-0684

GAMBLING PART I

The math on gambling

By Barnet D. Wolf
Columbus Dispatch

Sometimes the germ of a project begins with something very common.

"Ohio Hedges Its Bets," a three-day series on the gambling industry published Oct. 27-29 in *The Columbus Dispatch* really started when a Business News copy editor began pooling money from staff members to drive two hours to Indiana and buy Powerball tickets.

One writer pointed out that ever since the multistate lottery games began in states adjacent to Ohio, the staff had rarely pooled its money for the Ohio Lottery.

What was happening in Columbus, we thought, probably was happening all across the state, not only with the lottery but also with other types of gambling.

Just how big is the legal gambling industry in Ohio and the nation, we wondered, and what impact have casinos, multistate lotteries and other forms of betting in neighboring states had on Ohio?

Setting out to find data on the subject, we found that some statistics would be easy to uncover, because they were collected by the states. A large amount of data also was compiled by the casino industry.

State lottery

For instance, some computer-assisted research was used to plot the impact of new border-state gambling options to the decline in Ohio Lottery (*www.ohiolottery.com)* ticket sales.

By using the state's own lottery sales numbers for several years, and then breaking them into ZIP codes, the figures could be plotted on a map.

The Ohio Lottery Commission provided an electronic database of sales for the past several years. That database was pretty thorough, with a breakdown of sales by each vendor (gas station, convenience store, etc.). Each vendor listing had a street address as well as a city and ZIP code for the business where tickets are sold.

That made it fairly easy to analyze with Microsoft Access. We basically grouped and totaled the sales figures by ZIP code, then moved that into our ArcView mapping software. Once we had a list of sales by ZIP code, it wasn't tough to map that across Ohio.

The results were compelling. They showed that the sales decline was much more severe along the state's borders than in the middle of Ohio.

The state also keeps statistics for bingo games and horse racing, two other large gambling sectors.

Out-of-state casinos

But other conclusions were much more difficult to determine, particularly trying to find out how much money Ohioans gambled at out-of-state casinos.

The result was a time-consuming exercise that required searching for financial and marketing data from public and private local, state and national sources, and then applying old-fashioned math to the numbers.

The findings could not be precise because of the limitations of the data available, but the approach produced a reliable estimate.

Fortunately, *The Dispatch* had the assistance of the Buckeye State Poll, conducted by Ohio State University in conjunction with the newspaper, to help serve as a secondary source to back up some of the numbers.

The questions posed by the poll, conducted in April, looked at the gambling opinions and habits of Ohioans and where — if anywhere — they gambled.

Gambling revenues

States maintain decent statistics when it comes to revenues from gambling — after all, tax money is derived from it — so that's always a good place to start. Most states also provide the most basic of their lottery data online.

In a state like Indiana, which has riverboat gambling, a gaming commission maintains statistics about slot machines and gaming tables. West Virginia's devices — video slots only — are overseen by the Lottery Commission (*www.state.wv.us/lottery*).

Sometimes the casino information is easily accessible. For instance, Mississippi's gaming commission *(www.msgaming.com)* has data on its Web site that gives the percentage of casino business from residents of other states.

For the most part, however, states maintain data only on the amount of money wagered and lost by individual casinos. Once you obtain those results, it takes a little leg-

Continued on page nine

Data sources:

West Virginia Lottery Commission, *www.state.wv.us/lottery*

Mississippi Gaming Commission, *www.msgaming.com*

Las Vegas Convention & Visitors' Bureau, *www.lasvegas24hours.com*

Indiana Gaming Commission, *www.state.in.us/gaming*

The Columbus Dispatch series, "Ohio Hedges Its Bets," can be viewed online at *www.dispatch.com/news/special/bets/day1/frilead.html*

It is also available in the IRE Resource Center, story #16974

Continued from page eight:

Gambling

work to get numbers for a specific state.

Indiana has riverboat casinos, and its gaming commission (*www.state.in.us/gaming/*) reports monthly on the amounts of money gambled at and won by each casino for slot machines and table games.

The commission's annual report — also online — includes the total numbers for the previous year at each casino.

For all intents and purposes, the amount won by the casino is the amount that is lost by its customers.

The result was a time-consuming exercise that required searching for financial and marketing data from public and private local, state and national sources, and then applying old-fashioned math to the numbers.

The legwork comes in when you need to find out how much of that money came from Ohioans. According to several statisticians and business experts, the easiest method for determining those amounts is to get market-share data.

Most of the border-state casinos were willing to divulge the percentage of their customers who visit from Ohio. When the market share was not provided by a casino, it was determined from other sources, ranging from estimates by state officials to a study that one university conducted by counting license plates in casino parking lots.

Las Vegas

Calculating the Ohio gambling numbers from Las Vegas was a bit more complex.

The Las Vegas Convention and Visitors Bureau *(www.lasvegas24hours.com)* is a good source of information on the amount of money won by casinos in the city, as well as airplane traffic and visitor profiles.

According to the bureau, 87 percent of people who visit Las Vegas gamble. Those who gamble spend an average of $559 per visit.

So, how many of those people are from Ohio?

As with other airports, McCarran International Airport (*www.mccarran.com*) maintains statistics regarding the true origin and destination sites for commercial flight passengers traveling to and from Las Vegas.

Using these numbers, it's possible to get a total number of Ohioans coming to Las Vegas via commercial flights (since the number will be for arrivals and departures, it must be divided by two to get only arrivals).

By taking the Ohio number against all commercial passengers, it's possible to get a percentage number of Ohio passengers that then can be applied to the charter flight statistics, in which origin destinations are not calculated.

Add up all those passengers and then multiply it by .87 (remember, 87 percent of all visitors gamble). Then multiply that number by $559 (the average lost per visit) and the total is the amount that Ohioans lost gambling.

Since the amount of money lost is about 8 percent of all money bet at the casinos, the total that Ohioans bet can be ascertained by multiplying the amount lost by a multiple in the neighborhood of 12.5 percent.

This calculated number of Ohioans gambling in Las Vegas meshed with the numbers determined in the Buckeye State Poll. Thus the amount of people gambling at casinos in other locations can be determined by using the poll.

Adding it all up, *The Dispatch* was able to estimate conservatively that Ohioans bet $11.7 billion and lost $1.8 billion on legal gambling in 1999. And a large chunk of that — all that money gambled outside the state — resulted in tax revenues for states other than Ohio.

Our approach in tackling the complex topic of gambling can be applied to other subjects in which precise data does not exist: With a little creative thinking, reasonable estimates based on a solid foundation can be made.

Barnet D. Wolf can be reached by e-mail at bwolf@dispatch.com.
Doug Haddix (dhaddix@dispatch.com) contributed to this article.

Stories

For more on gambling, call the IRE Resource Center, (573) 882-3364.

Story # 17010. In this article, the Associated Press uses computer-assisted reporting to find "the American Indian gambling industry has boomed, with annual revenues increasing from $100 million in 1988 to $8.26 billion in 1998. Poverty and unemployment rates changed little during that period."

Story # 13853. *The Journal-Bulletin* (Providence, R.I.) investigation into the state lottery leads to the dismissal of director of the lottery, investigations by the attorney general, state auditor and state Ethics Commission and the cancellation of a contract that would have put the state gambling agency in the television bingo business.

GAMBLING PART II

Tribes and casinos

By David Pace
The Associated Press

A steady drumbeat of news stories throughout the 1990s documented the rapid growth of casino gambling on American Indian reservations. But lingering questions remained.

How had the billions of dollars in gambling revenues changed the lives of Indians on reservations, historically among the nation's poorest areas? Was gambling solving the unemployment, poverty and welfare problems that have plagued reservations? The Associated Press set out last spring to answer those questions by gathering data on a broad range of quality of life variables during the 1990s and then using it to compare reservations that have casinos with those that don't.

The immediate problem in designing such a research project was the scarcity of data collected at the reservation level. The 1990 Census provides a detailed snapshot of reservation life at the beginning of the casino boom, from education levels to poverty rates to housing conditions.

But comparative information from the 2000 Census won't be available for several more years, and the Census Bureau collects no reservation-level data in the years between the decennial census. We found only two other government sources of reservation-level data: The biennial workforce report compiled by the Bureau of Indian Affairs from unemployment data provided by tribes, and the Agriculture Department's Food Distribution Program on Indian Reservation.

Improvising

We used the BIA workforce reports for 1991 and 1997, and the Agriculture Department's FDPIR data for 1990 and 1998 as two comparative variables. We also obtained from the Department of Health and Human Services a breakdown by race of welfare caseloads in each state. That enabled us to compare American Indian welfare participation during the 1990s in states that permit casino gambling with those that don't.

Finally, we asked the Census Bureau's geography division to create a database of all the counties that are part of Indian reservations. That enabled us to use the bureau's 1989 and 1995 county income and poverty estimates, and its 1990 and 1997 county business pattern records, for comparative analysis.

The next major hurdle was dividing tribes into gaming and non-gaming categories. To operate a Las Vegas-style casino, the National Indian Gaming Regulatory Act requires that a tribe first negotiate a compact with the state. From the National Indian Gaming Association, we obtained a current list of Indian gaming operations and used it to divide tribes into gaming and non-gaming categories.

We divided them based on the legal definition, designating as gaming tribes those that operate Class III casinos under compacts with states. Tribes with small-stakes bingo games were combined, with those that have no gambling operations, into the non-gaming category.

The analysis found that welfare participation on Indian reservations with casinos grew far less during the 1990s than on other reservations. But the historically high unemployment and poverty levels on reservations changed very little during the 1990s, despite the influx of gambling money.

The analysis found that welfare participation on Indian reservations with casinos grew far less during the 1990s than on other reservations.

Between 1991 and 1997, when the U.S. unemployment rate dropped from 6.9 percent to 4.9 percent, the unemployment rate for 146 tribes with casinos declined from 55.9 percent to 52.2 percent, and it actually increased slightly among the 55 tribes with casinos that opened before 1992. Among the 144 tribes without casinos, the unemployment rate increased from 43.6 percent in 1991 to 48.3 percent six years later. Similarly, the average poverty rate in counties where there are gaming tribes declined only

Continued on page twenty-two

The Associated Press story, "Tribal Casinos," is available in the IRE Resource Center, story #17010.

For more information on this story check out the January/February issue of the *IRE Journal.*

Maps to go

By John Maines
South Florida Sun-Sentinel

There are two truths about buying good computerized mapping software:

1. Single-user versions are expensive.

2. Internet or Intranet versions are very expensive.

Three years ago, I decided to see if there was a way around the high cost of commercial mapping programs. The idea seemed to make sense: Build an inexpensive, simple mapping system and pack it with census stats and anything else we could get our hands on, especially free public domain data from federal, state and local government agencies.

We've done it. So can you.

For less than $300, it's possible to build a fairly powerful program that can be installed on a PC, a laptop, or launched on the Internet or company Intranet. At the *South Florida Sun-Sentinel* we have test versions running now. We should have the bugs in time for the March release of Census 2000 data on race and Hispanic origin.

To do this, we combined our own software know-how with some off-the-shelf mapping software designed for people who want to build their own Geographic Information Systems (GIS) software. We created a program that can be tailored any way we want it and launched on the Internet or re-deployed as a stand-alone application for installation a PC or laptop.

The acid test

The 2000 presidential election provided a great acid test.

One of the map prototypes was a precinct map of South Florida, which displays a color-coded map and spreadsheet-like data grids. When a user clicks on a precinct, the map pulls up demographics and voter registration data for election precincts in the *Sun-Sentinel's* circulation area: Broward, Miami-Dade and Palm Beach counties.

The precinct borders were drawn from files that we downloaded – for free – from a Web page run by a state research agency.

So it was pure luck that, when the election went to hell, we had a fully-functional computerized precinct map armed and ready. We changed the map constantly in the days following the election, each time reloading an updated version into our company Intranet, Sunspot, for use by reporters and editors.

The first maps we created were simple – showing who won each precinct. Red for George W. Bush, blue for Al Gore. When reports came in that voters in Palm Beach County had accidentally voted for Pat Buchanan instead of Al Gore because they misunderstood the "butterfly ballot," we mapped out where Buchanan got the most votes. When we found out that most of the disqualified votes were in black precincts, we mapped that out as well, adding new map layers on top of the old.

Our art department polished up our graphics for use in the newspaper. Len De Groot, a senior editorial artist at the *Sun-Sentinel*, made a terrific map that showed undervote rates in all 1,754 election precincts in the three South Florida Counties. It ran the entire length of a newspaper page, and covered about one-third the width. (See for yourself at *www.electricalfire.com/sfprecincts.htm*).

Pros and Cons

I'm not saying that newsrooms should toss the programs that they've already bought and build a Frankenstein. Arcview, MapInfo, Maptitude or some other powerful program will still be the best way to analyze Census 2000. But a simple custom-made system provides a terrific secondary mapping system for the newsroom, because it does something that the big bruisers don't do very well: Create dynamic maps on the Intranet that can be used by even the most non-technical people.

Consider two popular brands of high-quality Geographic Information System (GIS) software.

Arcview 3.2a ($1,195) and MapInfo Professional ($1,495) are built for use by one person per copy of software. That person builds the maps and distributes them to others. Both companies have viewers that can be used on any PC and are available for free. But the viewers can't do much more than zoom in or out of a map and identify what's on it.

Continued on page twelve

A downloadable version of the *Sun-Sentinel's* election precinct map is at *www.electricalfire.com*

To check out the *Sun-Sentinel's* map on undercounted votes, go to *www.electricalfire.com/sfprecincts.htm*

Continued from page eleven:

Maps

Mapping software:

SylvanMaps:
www.sylvanmaps.com

ESRI mapping products:
www.esri.com

Maptitude:
www.caliper.com

MapInfo: ***www.mapinfo.com***

The viewers also require some effort by the end user. They need to read the instructions. That's minimal, but still enough to send many people running. A good general-use mapping program needs to be "Web page simple," allowing the user to learn by pushing buttons and generally messing around.

Internet mapping software can be incredibly expensive. One of the best deals on Internet-based mapping software is $2,995 for Caliper's Maptitude for the Web license. MapInfo's MapX product line starts at about $5,000 and ESRI's ArcIMS comes in at $7,500. Yikes.

Those costs are probably far beyond what most newspapers are willing to spend. So very low cost was one of five requirements in developing our software. Here are the other four:

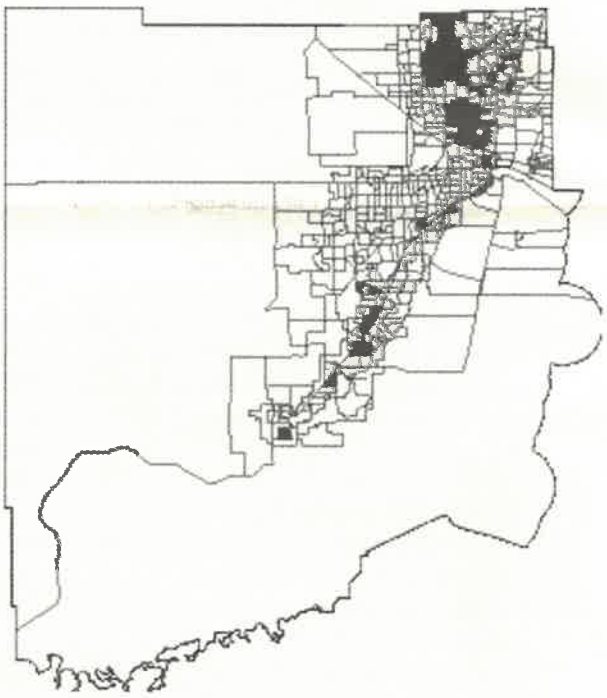
- It had to be extremely easy to use. The maps would come with no little or no instructions, allowing users to wing it.
- It had to read many kinds of commercial software, such as Arcview shapefiles or MapInfo's MIF files. It also had to be able to read data from programs used in most newsrooms, such as an Excel spreadsheet or an Access database.
- It had to be available in a desktop version, or on the Internet.
- It had to be of "rapid-deployment" nature, producing maps for deadline stories.

Creating a hybrid

Early on, it became clear that the system would be some sort of hybrid – parts of the software would be designed by us, and part would be purchased off-the-shelf. I thought it would be best for me to build the control panel that surrounded the map; the components that would allow reporters to push buttons and insert all the parameters for data analysis.

But I really didn't want to build the part of the program that would actually draw the map's lines, points and polygons. For that I needed some sort of "engine."

Continued on page thirteen



THESE TWO MAPS ILLUSTRATE HOW THE PRESIDENTIAL ELECTION "UNDERVOTE" WAS CONCENTRATED IN BLACK AREAS OF MIAMI-DADE COUNTY IN FLORIDA. THE FIRST MAP, RIGHT, SHOWS VOTING PRECINCTS WHERE 8 PERCENT OR MORE OF THE BALLOTS REGISTERED NO VOTE FOR PRESIDENT. THE SECOND MAP, PAGE THIRTEEN, SHOWS PRECINCTS WHERE THE POPULATION IS AT LEAST 50 PERCENT BLACK. VOTERS LIVING IN THOSE PRECINCTS REPORTED THAT THEY COULD NOT PUSH THE STYLUS THROUGH THE PUNCH CARD. ONE THEORY HELD THAT BECAUSE BLACKS VOTED OVERWHELMINGLY FOR VICE PRESIDENT AL GORE, A CHAD BUILD-UP WAS CREATED BENEATH HIS NAME, MAKING IT DIFFICULT OR IMPOSSIBLE TO PENETRATE THE PUNCH CARD WITH THE STYLUS.

THESE MAPS WERE CREATED BY THE *SOUTH FLORIDA SUN-SENTINEL* USING DATA FROM THE MIAMI-DADE SUPERVISOR OF ELECTIONS.

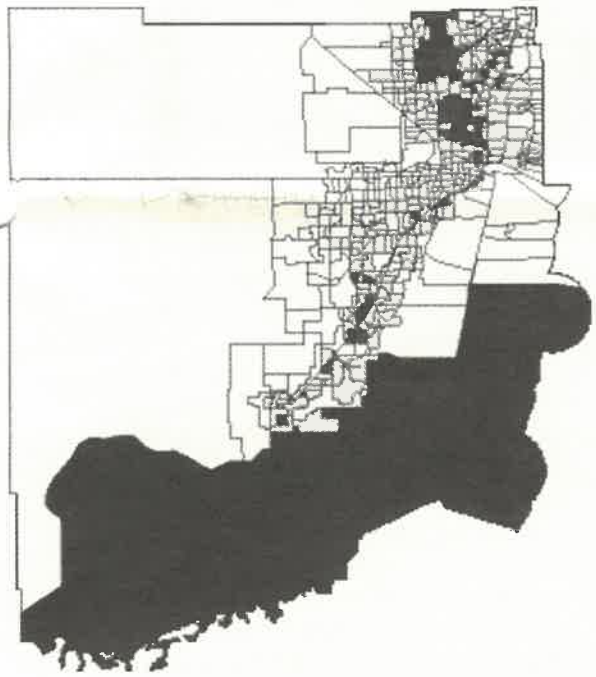
Continued from page twelve:

Maps

Several companies make such software. There's ESRI's MapObjects, MapInfo's MapX. But these cost thousands of dollars to buy and maintain and are really built for someone who builds high-quality maps for a living. Newsrooms don't need that degree of complexity.

One product that looked very good – at first — was ESRI's MapObjects LT, a trimmed-down version of MapObjects. Three years ago, the software only cost $295 and produced some great-looking maps.

But it lacked one very important feature. There was no easy way to put the names on roads in your map. It was a critical flaw in an otherwise good, inexpensive product. (The newest version of MapObjects LT has nice labeling capabilities. But it now costs $1,000).



After abandoning MapObjects LT, we tried about a half-dozen inexpensive GIS programs found on the Web. They all bombed for one reason or another. Some were too slow. Others couldn't handle geocoding. Some were just plain junk. Finally, in a late-night search of the Internet last February, I discovered a great little program called SylvanMaps.

A solution at last

SylvanMaps is sold by Sylvan Accent (*www.sylvanmaps.com*), a small company based in Taos, New Mexico. If you call Sylvan, there's a pretty good chance that the guy answering the telephone is Roger Bedell, the company president.

The software is easy to use and comes with clearly-written user manuals. There are three versions.

The first format, SylvanMaps.Net, is a $249 design-time control for Active Server Pages (ASP), the popular Web page program. Users build the map in Microsoft Frontpage or Visual Interdev and launch the finished product on the Internet. For another $496 you can buy the second version, which has more features than the first.

The third version, SylvanMaps OCX, creates maps in Visual Basic, Microsoft Access, C++ or Delphi, and you can install the program on the Internet or build a desktop program. Version three is the most expensive, $795. But that price allows you to build unlimited applications, which can be resold or distributed to co-workers and sister companies without fooling around with annoying royalties or redistribution fees All three versions are easy to use.

A beginner can put together a simple map, and work upward from there. The better your programming skills, the more sophisticated the maps you'll make. A very nice feature is that an entire map – including the roads, census tracts, and other features — is built from data stored in any standard relational database, including Microsoft Access, SQL Server, or Oracle. There's no "hidden" proprietary software.

SylvanMaps can also read or translate

Continued on page fourteen

The series, "Childhood Lost: Problems at DYFS," can be found at *www.app.com/dyfs* It will also soon be available in the IRE Resource Center.

Software: SPSS

Data: The Adoption and Foster Care Analysis and Reporting System
Source: state child protection agencies.

Data for fiscal years 1995 through 1998 is available through the National Data Archive on Child Abuse and Neglect at Cornell University. (*www.ndacan.cornell.edu*).

Elliot Smith, project manager at the National Data Archive, can be reached at 607-255-8104. The archive staff is helpful about answering questions on the use of AFCARS data.

Continued from page thirteen:

Maps

Arcview shapefiles, MapInfo MIF files, or the Census Bureau's TIGER line files. If translated, the map data can easily be installed in a database on your company's web server.

Such versatility is great, because it means you can use maps already created by other people — the Census Bureau, the state department of transportation, your local planning department, or whatever. The street maps we use at the *Sun-Sentinel* are Broward County's 911 maps, used for dispatching fire trucks and ambulances to emergencies. The maps are constantly updated, and much more up-to-date than anything that you can buy. And they're free. Mapping software is changing rapidly. Chances are that everything on the market now will be considered junk in a few years. We're going to continue to refine SylvanMaps for now, but at the same time testing other software that comes along as we search for the best product around.

Meantime, the data that goes into the maps keeps getting better and cheaper. For $20, we recently bought a CD-Rom from the University of Florida with all kinds of interesting public domain maps on it. These include hiking trails, bike routes, the paths of Florida hurricanes over the last 75 years and the locations of dozens of shipwrecks in the Atlantic Ocean and the Gulf of Mexico.

We plan to put those on our Intranet. Shipwrecks aren't as important as elections or Census 2000. But they're fun. And hey, that's Florida.

John Maines can be reached by e-mail at jmaines@sun-sentinel.com

Continued from page one:

Foster care

obviously does not detract from its usefulness. The data quickly answers questions like: what percentage of black children are in the system? What percentage of black children are in group homes? How many children were removed because of sexual abuse?

Child welfare experts say that the longer children are in foster care without a permanent placement, the more they are harmed emotionally. By subtracting dates in AFCARS, CAR journalists can calculate median and average time for children in care. Times in care can then be calculated for racial groups and sex.

Experts also say that moving children from foster home to foster home, or repeatedly removing them from their natural home, also injures their psyche. The AFCARS database will show, for example, what percentage of children have been moved in the system two or more times, or removed from their natural home more than once.

The AFCARS data is required to be kept and reported by federal law, and states that do not report the information accurately can be subject to fines. Data for fiscal years 1995 through 1998 is available through the National Data Archive on Child Abuse and Neglect at Cornell University (*www.ndacan.cornell.edu*).

More recent data is not yet available from federal officials, but because the data is required under federal law, I requested 1999 and 2000 data from New Jersey DYS and received it in 48 hours. Our analysis in SPSS, which led off a five-day series, showed the state has made little progress on its reform effort, and it demonstrated worse care for black children in the system.

Although the dataset at times seems like it can leap tall buildings in a single bound, there are limitations. The 1995 and 1996 datasets are not considered as accurate or complete as later versions.

Some states do report some categories differently, so it can be difficult to compare states to each other in those instances. Although it will show if a foster child has died, it provides no information on the deaths. It does not report maltreatment in foster care, and it provides no information on prior foster care episodes.

The 1999 AFCARS data is expected to be ready for release in January, according to Elliot Smith, project manager at the National Data Archive. Smith can be reached at 607-255-8104. The archive staff is helpful about answering questions on the use of AFCARS data.

Jason Method can be reached by e-mail at jmethod@app.com

Parking tickets

By Jason Callicoat
South Bend Tribune

One of the first stories I wanted to do after returning from a NICAR Boot Camp last spring involved linking a database of local parking tickets with a database of weather and climate information. I wanted to see whether there was a correlation between bad-weather days and the number of parking tickets handed out.

This was a story idea that was mentioned as an example during one of the boot camp presentations.

I decided to look at data from Jan. 1, 1998, to as close to the present as I could get.

Weather data

The weather database was fairly easy to get. The National Weather Service already had a table of each month's weather data saved as a text file. They maintained it this way so they could print it each month in a weather newsletter.

The weather service had about 18 months' worth of weather data saved on a disk, which they let me borrow – for free – if I promised to mail it back to them.

There was some prohibition on them giving away data (they were supposed to charge for it), but the person I was dealing with was very accommodating. I couldn't copy the disk, or have the disk, but I could "borrow" it if I mailed him back the original.

For the months before the weather service began storing the data electronically, he gave me the newsletters in which they published the final data. Again, I couldn't have copies of these, but I could take them all back to the office with me if I mailed him back the originals.

I copied the text files for each month that they had saved on disk, transferring them into an Access database. Then I typed in about 12-months' worth of daily data from the weather service's paper newsletters. I only had about 10 fields, so it didn't take that long.

Parking tickets

The parking tickets database was a little harder to get, but I used some of the tricks I learned at boot camp.

The clerks in the city's Ordinance Violations Department said their software was 15 years old and not compatible with Microsoft.

They offered to make me a printout of all the parking tickets from the past two and a half years (roughly 50,000 of them), but it would take several days of print time for their dot-matrix printer to finish with that job.

I asked who was in charge of backing up their data every night, and they gave me the name of the information technology director.

I got in to see him fairly easily, and he looked through the backup files and found one for parking tickets. He had no problem with burning me a CD of the file for free. It was stored in DBase format, which is easily converted into Access.

But when I got it converted and reorganized the data, I found it was a couple years old and didn't have the tickets that matched up with my weather data. When I went back that afternoon, the IT director found the right backup file and burned another CD. It had about three years' worth of tickets, up through the tickets that were given the previous day.

On to analysis

I linked the tables by date and found that the 10 percent of the days with the most tickets were generally nice-weather days, and the 10 percent of the days with the fewest tickets were generally bad-weather days.

The story ended up taking about six or seven weeks to complete, although I was pulled away from it several times to do daily stories. The CAR portion of it took about two-thirds of that time.

The parking tickets database also proved useful, beyond the original story, because it listed license-plate numbers and registered owners of all the cars (tens of thousands of them) that had gotten a parking ticket in South Bend in the past several years. I ended up doing a few searches for another reporter who had a group of license plates he wanted to trace.

Also, while working with the parking tickets database, I found quite a few government cars listed. I ran a query to pull them all out, and I found that almost none of the tickets on these government cars had been paid.

The cars belonged to a group of armed forces recruiters who had offices down the street from our building.

Continued on page sixteen

Software: Microsoft Access

Data: Monthly weather data for local area
Source: National Weather Service

Data: Parking tickets
Source: South Bend Ordinance Violation Department.

The *South Bend Tribune's* stories will soon be available in the IRE Resource Center. Search the story database at *www.ire.org/resourcecenter* to find the order number.

Continued from page fifteen:

Parking tickets

Together, they owed more than $4,000 in tickets and late fees. I did some checking, and it turned out that the city had issued a legal opinion that was preventing Ordinance Violations from collecting on these tickets.

The weather service had about 18 months' worth of weather data saved on a disk, which they let me borrow - for free - if I promised to mail it back to them.

The opinion said that armed forces recruiters could not be taken to small claims court for unpaid parking tickets the way anyone else would be. That's because it would involve one branch of the government suing a larger branch of the government, which is prohibited by the Supremacy Clause of the U.S. Constitution.

Eventually, these recruiters had figured out nothing ever happened if they failed to pay their parking tickets. (They preferred to park on the street outside their offices, even though the government rented parking spaces for them in a garage right across the street. But the spaces were on the 6th floor, and the recruiters didn't like driving all the way up there.)

After a couple years of unsuccessful attempts, the city stopped trying to collect the money for the recruiters' tickets.

I wrote a story about it, and the recruiters were all very contrite when I talked to them. The Marines sent a colonel up from Indianapolis to pay $1,000 worth of their tickets a day or so after I interviewed the recruiters. Some of the other recruiters paid smaller amounts of their tickets, as well.

Tips for others

I'd say the parking tickets/weather story is a good one to do as a "first venture." The National Weather Service was not at all possessive with its data, and the people I dealt with there actually seemed pleased that someone was taking interest in their work. They went out of their way to make their data available.

In South Bend, the records of our parking tickets are maintained by the Ordinance Violations Department. It's not a city department we report on very regularly, which seemed to work to my advantage.

They hadn't been burned and didn't seem suspicious at all, just a little curious. They didn't react badly when I gave somewhat vague answers to their questions of what I wanted the data for. The only hurdles with this department were technical: working around their 15-year-old software and identifying the correct back-up file.

After returning from NICAR's Boot Camp, I found these stories were a good way to get my feet wet with CAR. I'd recommend similar stories to anyone as a "first venture."

Jason Callicoat can be reached by e-mail at jcallicoat@sbtinfo.com

For similar parking-related stories and tip sheets, check out the following files at the IRE Resource Center, (573) 882-3364, or on the Web at *www.ire.org/resourcecenter/*

Story #: 14251. KOMU-TV (Columbia, Mo.) Reporter Jack Dolan found that the odds of getting a parking ticket in Columbia on a day with even the slightest amount of precipitation (.01 inches or more) are about half what they are on a perfectly sunny day, and the odds plummet as the weather worsens. Nov. 5, 1997.

Story #: 10450. *Roanoke (Va.) Times & World-News* describes the battle between downtown employees and the local police department over illegal parking; describes the lengths to which violators go to avoid tickets and includes a computer-assisted analysis of the biggest offenders and the spots most often ticketed, March 27, 1994.

Continued from page two:

CAR

porters, tech-support staff and librarians. Even so, she publishes just once a quarter because she "found it impossible to find the time to produce it more often than that," indicating just how time-consuming a newsletter can be.

See how long it takes to produce your first newsletter to gauge how often you'd like to issue your newsletter.

Let's face it. Writing about Excel can be boring. But if your newsletter is fun to read, more people will get through it. Even fun graphics you can pull from clip art liven up the page.

Online versions

Provide a paper edition and online edition, if possible. Remember to include links for data and Web sites related to the story.

Richard Dalton can be reached by e-mail at rdalton@newsday.com

Continued from page one:

Race problems

categories. And if we want to do stories on segregation or diversity that talk about changes since 1990, we need a way to properly tabulate those who count themselves as multiracial – a so-called "bridge" between the two datasets.

Before I suggest a solution, I should hasten to add that we won't always want to turn the 2000 racial/Hispanic categories back into the 1990 version. The new categories are a much richer trove of information about the state of racial interaction in this country, and deserve plenty of attention in their own right. (Look, for instance, at doing stories about which racial/ethnic combinations are most and least common. Or focus on age groups, where I suspect you'll see that increasing numbers of multi-racial youth can be considered direct evidence of the fading racial and cultural prejudices of their parents and grandparents.)

Some of the most important uses of census racial and ethnic data will be to monitor progress towards racial fairness in such matters as hiring, voting, lending, access to education and health care, etc.

With that said, I'll further note that some places won't need to worry much about using a bridge. For instance, the 1998 dress rehearsal census in 11 South Carolina counties found that only about one percent of the population considered themselves multiracial. No matter how those in the multiracial column are parceled out in places like that, they won't change the overall mix significantly.

But lots of other places – like Sacramento, where fully 5 percent of the population chose multiracial status in the dress rehearsal census — will need to make some decisions about how best to treat those folks when comparing the 2000 numbers to 1990. For what it's worth, here are some suggestions, in increasing order of complexity:

Ignore them

If your local multiracial population is a tiny part of the whole, perhaps a fraction of a percent, you could pretty safely just drop them from any calculations. Simply calculate your percentages using the "Race X alone" columns, divided by the total of those "Race X alone" columns. You will need to add 2000's "Hawaiian and Other Pacific Islander" and "Asian" categories to be equivalent to 1990's "Asian and Pacific Islander" category, but that's easy. The results you get will be within a few tenths of a percent of what you would get with the more troublesome methods suggested below.

Parcel out

Okay, you're queasy at the idea of just ignoring them, even though their numbers are very small. Instead, figure the percentage each single-race category is of the total population, then allocate the multiracial population using the same proportions. For example, if your population is 100,000 Whites, 10,000 Blacks, and 1,000 folks of two or more racial backgrounds, then treat your population as 100,900 Whites and 10,100 Blacks. Like the "ignore them" strategy, this only works comfortably in a place where the multiracial population is quite small.

Count races, not people

Paul Overberg of USA Today suggests this method of looking at the racial mix of an area. You won't get this in the March data release, but later census products will have tables counting the number of people who claim each race at least in part, labeled "Race X alone or in combination with other races." The total will add up to more than the actual population, thanks to the multiple counting of people with multiple racial backgrounds. But calculating the percentage each race is of the total count by race would be easy and informative. I'm less convinced, though, that this will produce a good way to compare 2000 to 1990.

Continue on page twenty

Census seminar:

The next IRE and NICAR Census seminar will be held Feb. 24-25 in Boston, Mass.

Instructors Paul Overberg, of USA Today, and Steve Doig, of Arizona State University, will cover topics such as census story ideas, the data rollout schedule, census definitions and geography, mapping, reapportionment, getting the data, making comparisons to 1990, the new race categories, and the controversy over using adjusted census counts.

More information, including registration forms, is available at *www.2000census.org/*

Census listserv

The Census-L listserv is a discussion list about census topics, and occasionally mapping. To subscribe, send an e-mail to: majordomo@nicar.org

In the body of the message type: subscribe census-l youremailaddress

You will receive an e-mail confirming your subscription, and start receiving e-mails shortly thereafter.

More information is available at *www.ire.org/membership/subscribe/census-l.html*

Continued from page nineteen:

Race problems

One race answer

The very best bridge method is to allocate them based on how they would answer if they could choose only one race. For instance, if 90 percent of people who identified themselves as "White and Black" would identify themselves as "Black" if only one choice could be made, then you simply apply those various percentages to the various multiracial categories.

Problem is, no one knows what those percentages are. Learning them will require either a massive survey far too expensive for any news organization to undertake, or - the real answer – access to individual census returns. The good news is that the Census Bureau says it will tackle the problem, by comparing a large sample of 2000 multicultural returns to what those same people answered to the race/Hispanic origin questions back in 1990. The bad news is that it may be another couple of years before they get around to it. So, on to another solution.

Cut them into pieces

Remember the story of King Solomon, who suggested a baby be divided in half so that both mothers who claimed the child could have a share? The bridge method I ultimately favor uses the same, though hopefully less messy, approach. You take each of the 126 possible race/Hispanic origin categories, cut them up into their single-race components, and then allocate those pieces into the correct 1990 categories. For example, 100 people of Black and Asian background would be counted as 50 Blacks and 50 Asians.

There is a further refinement, though, that I think is needed. Some of the most important uses of census racial and ethnic data will be to monitor progress towards racial fairness in such matters as hiring, voting, lending, access to education and health care, etc. Clearly, for those purposes someone who is of mixed parentage would be considered to be of minority background even if one parent was White. For that reason, I argue that 100 people of Black and White background should be counted as 100 Blacks. My reasoning falls in line with a federal directive that persons of mixed White-minority parentage should be counted as minorities in all instances where that label offers protections or benefits.

My method is pretty simple in conception, though damned ugly in execution. Here, for example, is how you can use the variable names from the PL94-171 files that will be released in March to calculate the 1990 equivalent of Latin Blacks:

```
P0010004) - (P0020006) + (P0010011)
- (P0020013) + (P0010016/2) -
(P0020018/2) + (P0010017/2) -
(P0020019/2) + (P0010018/2) -
(P0020020/2) + (P0010019/2) -
(P0020021/2) + (P0010027/2) -
(P0020029/2) + (P0010028/2) -
(P0020030/2) + (P0010029/2) -
(P0020031/2) + (P0010030/2) -
(P0020032/2) + (P0010037/3) -
(P0020039/3) + (P0010038/3) -
(P0020040/3) + (P0010039/3) -
(P0020041/3) + (P0010040/3) -
(P0020042/3) + (P0010041/3) -
(P0020043/3) + (P0010042/3) -
(P0020044/3) + (P0010048/3) -
(P0020050/3) + (P0010049/3) -
(P0020051/3) + (P0010050/3) -
(P0020052/3) + (P0010051/3) -
(P0020053/3) + (P0010052/3) -
(P0020054/3) + (P0010053/3) -
(P0020055/3) + (P0010058/4) -
(P0020060/4) + (P0010059/4) -
(P0020061/4) + (P0010060/4) -
(P0020062/4) + (P0010061/4) -
(P0020063/4) + (P0010064/4) -
(P0020066/4) + (P0010065/4) -
(P0020067/4) + (P0010066/4) -
(P0020068/4) + (P0010067/4) -
(P0020069/4) + (P0010069/5) -
(P0020071/5) + (P0010071/5) -
(P0020073/5)
```

No, don't start typing. On my "Reporting Census 2000" website (go to *http://cronkite.pp.asu.edu/census/race.htm)* I have placed a downloadable copy of the SAS code necessary to do the conversions for all ten 1990 race/Hispanic categories, as well as an Excel template that will do the same thing. Those who use other flavors of database programs, like SPSS or SQL, should be able to adapt this pretty easily to your favorite environment.

Steve Doig can be reached by e-mail at steve.doig@asu.edu

Restaurant inspections

By Rick Linsk
St. Paul Pioneer Press

It took a strong stomach to successfully pull off our "Restaurant Reservations" series (Oct. 23-27, 2000) — not to mention oodles of computer-assisted reporting resources and considerable tenacity.

First, a disclosure: We didn't invent this wheel. Restaurant inspections are a longtime staple of investigative reporting. Television stations in the Twin Cities and other cities, sometimes aided by hidden cameras, have reported filthy or unsafe practices. Newspapers in Toronto, San Jose, Charlotte and St. Louis, among others, have also examined restaurants. Our series built on the techniques of our predecessors, and in some ways went even farther.

We expected to be finished by the end of 1999. But slow-moving government bureaucracies, dirty data, and other factors would confound our timetable.

The editors and reporters central to the project – Projects Editor Jeff Kummer, Food Editor Kathie Jenkins, restaurant beat reporter Gita Sitaramiah and myself, a member of the investigative team – began in mid-summer 1999 with several fundamental questions. We wondered whether area restaurants were generally following or failing the rules; whether the government's system of overseeing restaurants was working; and whether the average consumer had any way to distinguish between the safe and unsafe, the clean and unclean.

We expected to be finished by the end of 1999. But slow-moving government bureaucracies, dirty data, and other factors would confound our timetable.

Getting the records

From the outset, we faced an obstacle: Fragmentation of oversight. Across the nation, restaurants are inspected by a patchwork of state, county and local government health agencies. In Minnesota, ultimate responsibility rests with the state Department of Health, but the department delegates its powers to dozens of counties and cities. In the end, we studied restaurant inspections by MDH and five local agencies in the Twin Cities that were willing and able to give us computerized records.

Some government agencies resisted our request for inspection data and only relented after official requests, meetings, and follow-up phone calls and e-mails. Even then, some took weeks or months to provide the records.

The most notorious agency was the state Department of Health, which oversees food establishments in Dakota County and most of outstate Minnesota. MDH failed to provide complete inspection data for six months, and only after the *Pioneer Press* noticed that a computer file initially provided by the department was incomplete. The department also took two months to provide computer records about foodborne outbreaks, and only after "sanitizing" the records of information that could identify victims. (More on that below.)

In one county, officials gave us the computer data from inspectors' portable computers, but said their contract with a software vendor prevented them from providing any background on how to read the data.

Analyzing the data

As the old saying goes, be careful what you ask for, because you just might get it. That's how we felt after the various agencies' data arrived. The reason: Regulatory agencies use computers to help keep track of inspection schedules and compliance, not sift trends from the data as we wanted to do. The six databases had several different formats. Some were missing records. Others were littered with duplicate records. One county had only entered about half of all restaurant visits into the computer system - the "bad" inspections, leaving the better visits documented only in hundreds of paper files.

St. Paul's data came from an ancient computer system (finally replaced in the middle of our project) with some confusing twists -

Continued on page twenty-two

The full series is online at *www.pioneerplanet.com/restaurantreservations.* It will soon be available in the IRE Resource Center.

Tipsheets:

Tipsheet # 843: "Food Investigations: What happened to the chilling effect of Food Lion," (1998, New Orleans), provides tips for investigating restaurants, a list of useful databases, Web sites and government sources for investigating the food industry. Audio tape of this panel is available through Sound Images, Inc., (303) 649-1811.

Tipsheet # 298: Lists questions to ask when looking at payroll records, property records, political campaign contributions, voter registration, city budget, and specialty data bases that you can ask for or create yourself such as databases on restaurant inspections, handgun permits, ect. Audio tape of panel available through Sound Images, Inc., (303) 649-1811, for $10. Ask for tape # IRE 95-3. (1995, Miami)

Continued from page fourteen:

Restaurants

among them, taking violations by restaurants long since out of business and attributing them to the building's current occupant. St. Paul also was unable to give us a list of restaurants issued tickets, something other agencies were able to do.

Most of our tasks had to be multiplied by six – for instance, six sets of queries for the worst and best restaurants; six sets of phone calls to officials to request paper records, confirm our findings, ask about "good guys," and other needs; six different sets of preparation materials for interviews with officials, and so forth.

One of the dilemmas we wrestled with was how to analyze the inspection scores. There's a strong argument that scores can be misleading. A restaurant can get a mediocre score, say in the low 80s, for a bunch of minor violations that involve "floor, walls and ceilings" but not food safety. On the other hand, a restaurant can get a score of 90 while committing one or two really gross violations, like improper food temperatures and cross-contaminating cooked food with bacteria from raw food -- things that can make people sick.

So from early on, to avoid this controversy, we never contemplated actually publishing the averages, or scores from a given year, etc. Instead, we made plans to count up and rank restaurants based on how many "critical" violations they received.

In total, we found Twin Cities restaurants had violated food-safety rules more than 130,000 times from 1995 to mid-1999, including more than 20,000 violations that government agencies classify as "critical" because they can make diners sick. We also obtained a state database of foodborne illness outbreaks, many of which had never come to light before. As for enforcement, we pieced together agencies' records – sloppy in most cases – and found penalties against violators were rare, even when they were proven to have made people sick.

Tips for others

Advice for this or other health and safety related stories:

Carefully document, for instance in a spreadsheet, all your requests for information – when, to whom, how officials responded, and where things stand. It gets harder to keep track as the weeks and months go by, and multiple requests pile up.

To locate victims, consider such public records sources as lawsuits, foodborne illness complaints, death certificates, hospital discharge records, even letters to the editor.

Think early, and if possible decide early, how to document and illustrate your findings regarding specific restaurants. We began by looking at bests and worsts, but had concerns about whether the data would be accurate, timely and complete enough. Late in the game, we shifted gears and wound up publishing inspection results for the most popular restaurants (as measured by the Zagat Survey).

Rick Linsk can be reached by e-mail at rlinsk@pioneerpress.com.

Other restaurant inspection stories:

"Grade 'A' Bribery," by KCBS-TV (Los Angeles), Nov. 1998, investigated corruption within Los Angeles County's restaurant inspection program. Tape # 15223.

"Dangerous Dining," by Trebor Banstetter of the *News-Journal* (Daytona Beach, Fla.), Aug. 1997, examined the restaurant inspection system in two Florida counties. Story # 14302.

Copies of these stories can be ordered from the IRE Resource Center by calling (573) 882-3364. Search the story database at *www.ire.org/resourcecenter.*

Conitnued from page ten:

Casinos

gaming tribes, the poverty rate increased slightly, from 18.2 percent to 18.4 percent. In the U.S. as a whole, the poverty rate increased from 12.8 percent to 13.8 percent during that period.

To understand why gambling revenues weren't having more of an impact, we tried to determine where the money was going. We filed a Freedom of Information Act request with the National Indian Gaming Commission - the agency with regulatory oversight of the Indian gambling industry – asking for a revenue breakdown of the casinos.

The commission refused to release financial data on individual casinos, saying it was proprietary information. But it did classify the casinos into six different revenue categories, from those making more than $100 million a year to those making less than $3 million a year, and provided a cumulative revenue total for each category.

We then used the tribal enrollment numbers in the workforce reports from the Bureau of Indian Affairs to calculate the actual number of Indians belonging to tribes in each of the six categories.

IRE STUDENT REPORTING CONFERENCE MARCH 31,2001

At the Missouri School of Journalism, Columbia, Mo.

SPEAKERS INCLUDE:

Brant Houston, IRE and NICAR
Dave Dietz, TheStreet.com
Stuart Watson, NBC 6 - Charlotte
Charles Davis, FOI Center
James Neff, Random House author
Mark Lagerkvist, News 12 - Long Island
Leon Dash, Univ. of Illinois, Champaign-Urbana
Rosemary Armao, Sarasota Herald-Tribune
Neil Reisner, Miami Daily Business Review
Mike McGraw, The Kansas City Star
Dave Raziq, KHOU-Houston
Jennifer Lafleur, St. Louis Post-Dispatch
Dianna Hunt, The Fort Worth Star-Telegram
Ken Wells, The Wall Street Journal
Drew Sullivan, IRE Board
Greg Reeves, The Kansas City Star
Joel Kaplan, Syracuse University
David Boardman, The Seattle Times
Shawn McIntosh, The Clarion-Ledger
Rose Ciotta, The Philadelphia Inquirer
Stephen Miller, The New York Times
Paul Adrian, Investigative reporter-producer
Jo Craven McGinty, Univ. of Missouri
John Ullmann, The World Press Institute
Diana Henriques, The New York Times
Victor Merina, Freedom Forum/Poynter Institute
Steve Doig, Arizona State University
Dawn Fallik, St. Louis Post-Dispatch

PANELS:

Investigating politicians and local government
Cultivating sources
Investigating nonprofits
Dissecting a school, city or state budget
Cops and mining court records
Writing the story for print and online pubs
Finding out about licensed professionals
Using property records and tracing land holdings
Using the FOIA and local sunshine laws
Using tax records
Producing the story for TV

HANDS-ON TRAINING:

Internet research – Using the Web for your beat
Spreadsheet basics – Taking apart a budget
Using databases – Hunting for scoops

INTENSIVE ONE-DAY CONFERENCE

Come and learn from some of the best in the business! Sponsored by Investigative Reporters and Editors – and hosted by the Missouri School of Journalism – the conference will expose young journalists to some of the most talented reporters and editors in investigative reporting.

IRE, along with the National Institute for Computer-Assisted Reporting, is a leader in journalism education and practical hands-on training. NICAR is a joint program of IRE and the Missouri School of Journalism.

Register by February 15 to get a registration discount and a free box lunch.

REGISTRATION

STUDENTS

IRE member	$25 ($20 if registered by Feb. 15)
Non-member	$50 ($40 if registered by Feb.15) (Includes a 1-year membership)

FACULTY

IRE member	$25 (free if registering with 5 or more students)
Non-member	$75 (Includes 1-year membership)

PROFESSIONALS

Open to professionals from papers under 100,000 circulation or TV stations in markets smaller than top 50.

IRE member	$25
Non-member	$75 (Includes 1-year membership)

IRE For more information visit our Web site at www.ire.org or call 573-882-2042

Bits, Bytes and Barks

Fresh data

The NICAR Database Library has recently updated the U.S. DOT Hazardous Materials Incidents, the NEA Grants and the AIDS Public Information databases.

The Hazardous Materials data contains the incident reports of unintentional releases of hazardous materials for all modes of transportation (air, highway, railway, water) from 1971 through May 2000. Cost is $50, $100, or $125 depending on size of your news organization.

The AIDS Public Information Data Set contains information from Acquired Immunodeficiency Syndrome cases reported to state and local health departments. The most recent data available from the Centers for Disease Control is for 1998. Cost is $25, $50 or $75.

The National Endowment for the Arts grants database contains all organizations or individuals who've received money from the NEA since the program began in fiscal year 1987. The most recent data available is for fiscal year 1999. Cost is $25, $50 or $75.

More information about these datasets, including sample data and record layouts, is available at *www.ire.org/datalibrary/datasets*. To order the data, call the NICAR Database Library at (573) 884-7711.

Recent CAR stories:

• "**Gunrunner's Paradise**," *Dayton Daily News*, Dec. 10-13, 2000 *www.activedayton.com/partners/ddn/local/projects/guns/1210gunmain2.html*

The paper's analysis found Ohio ranks behind only Florida, Georgia, Virginia and Texas for guns used in crimes out of state. The data, obtained from the Bureau of Alcohol, Tobacco and Firearms, show that in 1997 and 1998 more than 1,000 guns used in crimes nationwide came from Ohio.

• "**Crashing for Cash**," *Sun-Sentinel*, Dec. 17-20, 2000 *www.sun-sentinel.com/crashingforcash/*

To study the dimensions of auto insurance fraud, the *Sun-Sentinel* analyzed thousands of state Department of Motor Vehicles records and other government and insurance industry claims data.

The *Sun-Sentinel* built a database that captured the driving and accident histories of hundreds of people identified in arrest reports or other investigative files.

• "**Destined to Fail**," *The Beacon Journal*, Dec. 10-13, 2000

www.ohio.com/specials/buchtel_series/

A six-month *Akron Beacon Journal* examination revealed factors that have nothing to do with a child's family income or race help explain why poor minority children in Akron perform behind nearly every other group of students.

Global conference

IRE and NICAR are among the sponsors of the first global journalism conference on investigative reporting and computer-assisted reporting, April 26-29, 2001 in Copenhagen, Denmark.

The 4-day conference will focus on great stories, the latest techniques and the effect of media convergence on investigative journalism. More information is available at *www.ire.org/training/denmark/2000conf/*

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