January/February 2000

Uplink

ON THE BEAT

Religion with CAR

By David Crumm
Detroit Free Press

Searching for God at the turn of the millennium requires more than prayer and meditation. It takes a computer.

The subject is vast and is growing more diverse by the decade. In the United States, 9 out of 10 Americans tell pollsters that they pray and are part of some religious group. While the vast majority of Americans claim to be Christian, most Christians feel increasingly free to switch back and forth among dozens of different denominations – or, if they're feeling restless, to break away entirely and form their own independent sects.

The influences of the world's other major faiths are surfacing more frequently in contemporary music, television, the Internet, movies and magazines.

Outside the United States, religion also remains a major force for good or ill. It can be a powerful catalyst in ethnic strife, a fuel for political aspirations and, in many cases, a helpful inspiration toward resolving conflicts and building communities.

To keep up with this amazingly complex landscape, I use my computer in four basic ways.

Internet

As a journalist, I don't idly surf the Internet, because there are thousands of religious Web sites and most of them are not factually reliable.

Over the past five years, I've developed my own list of reliable religious Web sites, most of them sponsored by well-established groups. The list changes constantly as Web sites appear, evolve and disappear. A good place to start sorting out your own list of useful sites is the Yahoo index to religion, which you'll find linked off Yahoo's main page (as a sub-category of "Society & Culture").

A few of the sites I use most frequently are the Vatican's information-rich Web site at *www.vatican.va*, a large evangelical site at *www.gospelcom.net* that links to nearly 150 religious groups and a helpful site at *www.wfn.org* where many Protestant denominations post news releases. A superb place to background yourself on religious history and traditions is the recently upgraded *www.britannica.com*, sponsored by the encyclopedia.

By carefully sifting sites, the Internet has played a major role in my reporting on ev-

Continued on page twenty-two

CENSUS 2000

Race tabulations

By Steve Doig
Arizona State University

No problem in American society has been more chronic and vexing than that of race relations. Add to that the strains of demographic change caused by the immigration to the United States of millions leaving their countries of birth to seek a better life here. The resulting frictions and adjustments are the source of many important stories for journalists, and Census data is one of the best ways of measuring those changes.

However, the actual tallying of persons by race and ethnic origin has changed greatly over time, and the differences in definitions and methodology from one Census to the next can make it difficult to do accurate comparisons. One of those big changes in methodology is occurring with the 2000 Census, and it will require careful attention from those of us who

Continued on page seventeen

Inside Uplink

CENSUS 2000

This issue includes two stories to get you prepared for Census 2000 data, including one on this page by Steve Doig of Arizona State University about changes in race classifications, and another by Doig and Griff Palmer, of *The Daily Oklahoman*, on a formula they developed for reapportionment.

See page two

MILITARY

Russell Carollo of *The Dayton Daily News* tells how he tackled previously unexamined military aviation databases for his series "Falling from the sky."

See page four

HEALTH CARE

Richard Dalton, Jr., and other reporters at *Newsday* used a variety of data to show how HMOs failed to tell customers about doctors who had been disciplined.

See page seven

CAR IN SCHOOLS

David Kesmodel of IRE and NICAR examines whether journalism students around the country are getting hands-on training in computer-assisted reporting.

See page thirteen

Uplink

January/February 2000

Volume 12, Number 1
A newsletter of the National Institute for Computer-Assisted Reporting

EDITOR
Brant Houston

DIRECTOR OF PUBLICATIONS
Len Bruzzese

MANAGING EDITOR
Mary Jo Sylwester

ASSOCIATE EDITOR
Jessica Larson
Cindy Eberting

ART DIRECTOR
Kerrie Kirtland

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

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.

CENSUS 2000

Reapportionment

By Griffin Palmer
The Daily Oklahoman
and Steve Doig
Arizona State University

The 2000 Census isn't just a good idea – it's the law of the land. The U.S. Constitution mandates a national headcount every ten years for one purpose, and the specified reason isn't to help fast-food franchisers decide where to build their next burger joint. Rather, the census is required for fairly apportioning the 435 seats in the House of Representatives among the 50 states. All other census applications, whether drawing political district lines, allocating federal funds or siting taco stands, are just constitutional gravy.

The decennial reapportionment of Congressional seats means great stories, and they're the very first ones that will come out of actual 2000 Census data. By law, the Secretary of Commerce (who oversees the Bureau of the Census) will tell the President in late December of 2000 the official population of each state, and the President will then notify Congress. At that point, the bloodletting will officially begin.

However, it's no trivial task to translate the states' populations into the number of seats each will get, as the two of us have learned.

Because the number of Congressional seats is capped at 435, some fast-growing states will gain new seats at the expense of those with slower growth. Reapportionment means political musical chairs in the losing states, and a big scramble for the newly-open seats in gaining states.

However, it's no trivial task to translate the states' populations into the number of seats each will get, as the two of us have learned. A simple proportional allocation (in which, for instance, a state with five percent of the nation's population gets five percent of the 435 seats) doesn't work for two reasons. For one, there are no fractional seats in Congress (although some of the occupants certainly are fractious), and so a proper rounding scheme would have to be devised. And more important, the Constitution guarantees that even the smallest state will get at least one House seat even if its share of the nation's population is "worth" less than half a seat.

Equal proportions

Numerous apportionment schemes have been devised, and five different ones actually have been used at various times in the nation's history. (See *www.census.gov/population/censusdata/apportn.pdf* for a brief history of those methods, and *www.census.gov/srd/papers/pdf/rr92-6.pdf* for a more exhaustive discussion of the history and mathematics of apportionment methods.) We'll only concern ourselves here with the so-called "method of equal proportions" which has been used since 1940 and was upheld unanimously in 1992 by the U.S. Supreme Court.

Simulation

We won't try to explain the mathematical reasoning that underlies the method of equal proportions, because then we'd have to pretend that we actually understand it ourselves. But we have devised an Excel spreadsheet that accurately simulates the method and thus will let you figure out where your state stands in the reapportionment shuffle. To satisfy yourself that it works, first we'll build the spreadsheet using the official 1990 populations. Here's how:

- Open a new spreadsheet.
- Put these labels in the first row on the specified columns: A: "State", B: "Population", C: "Seat #", D: "Multiplier", E: "Priority value".
- Fill range A2:A51 with the names of the states.
- Fill range B2:B51 with the 1990 apportionment populations of the states (which you can find at *www.census.gov/population/censusdata/table-a.pdf*)
- Fill range C2:C51 with the number 2.

Continued on page three

Census 2000

Continued from page two:

(We start this with seat 2 because every state is guaranteed at least one seat.)

- Put this formula in Cell D2: =1/SQRT(C2*(C2-1)). The formal name for this multiplier is the "reciprocal of the geometric mean," in case you get asked on "Who Wants to be a Millionaire?"
- Put this formula in Cell E2: =D2*B2
- Put this formula in Cell A52: =A2
- Put this formula in Cell B52: =B2
- Put this formula in Cell C52: =C2+1
- Copy range A52:C52 all the way down to Row 2951. You'll know you did it right if "Wyoming", "455975" and "60" are on that row.
- Now select range D2:E2 and copy it all the way down to Row 2951, too. It worked if you see "0.0168073" and "7663.716" down there.

(Okay, okay, here's what's happening. The method of equal proportions requires that you calculate a "priority value" for at least as many seats as each state could possibly get. As it turns out, the easiest thing to do is calculate such values for seats 2-60 for every state, on the grounds that not even California will get as many as 60 seats. And 50 states times 59 priority values means 2,950 rows.)

The decennial reapportionment of Congressional seats means great stories, and they're the very first ones that will come out of actual 2000 Census data.

Now open a new sheet and follow these steps:

- Go back to your first table and copy the entire thing, all 2,951 rows and five columns.
- Use "Paste Special...Values" to paste it onto the new sheet.
- Sort that second sheet by descending order of Priority Value
- Insert a blank row at Row 387. This cuts off the table at 385 seats.
- Insert 50 blank rows into Range A2:A51.
- Copy the names of the 50 states into Range A2:A51.
- Put a one (1) in Cell D2, and copy it down to Cell D51

Whew. You now have a table with 435 records, one for each seat in Congress. To find out how many seats each state gets, simply do a pivot table by dropping the State variable into both the Row and the Data boxes.

The results

The main table also is worth a detailed look because it shows the order in which the seats were parceled out. Notice, for instance, that California got 42 of its 52 seats before Idaho and Rhode Island got their second seats. Notice also the first few states in the remainder below your main table. Those are the ones that just missed getting another seat (at someone else's expense, of course) and where much of the wailing was heard after the 1990 Census.

Now that you've done one for 1990, repeat the process using projected populations for 2000, which you can find at *www.census.gov/population/projections/state/stpjpop.txt*. You'll see plenty of winners and losers in the data. And if your state is near the bottom of the main table, or near the top of the leftovers, you'll know how important a good census count will be to your readers – and your politicians.

(A final note for FoxPro mavens: Griffin first developed this apportionment simulation using FoxPro. If you want a copy of his method, contact him by e-mail.)

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

Griffin Palmer can be reached by e-mail at gpalmer@oklahoman.com

The U.S. Census Bureau's 2000 Census page: *www.census.gov/dmd/www/2khome.htm*
There you can get information about data releases and download 1990 population figures.

Recent tipsheets on the census, available from the IRE Resource Center, include:

- "Using Census 2000 to Investigate Changes," (#1096) by Richard O'Reilly of the Los Angeles Times, for the 1999 IRE Los Angeles Regional Conference. This tipsheet gives reporters an idea of how they can start planning for the 2000 Census.

- "Preparing for the 2000 Census," (#1097) by Steve Doig, Arizona State University, for the 1999 IRE Los Angeles Regional Conference. Doig lists a series of tips on how to begin investigating census data "during the relative calm before the data storm hits."

Russell Carollo's story "Falling from the sky" is available online at: *www.activedayton.com/news/1999/10/24/falling.html.*

The series is also available from the IRE Resource Center (story #15844). Contact the Resource Center at (573) 882-3364 to order.

• Some tipsheets on covering the military, available from the Resource Center, include: "Checking out someone's military record," (#1107) by Don Ray, a freelance reporter, for the 1999 Los Angeles IRE Regional Conference. Lists sources and contact information for the National Archives and the National Personnel Records Center.

• "Fast Track: Military records," (#1106) by Vince Gonzales of CBS News, for the 1999 Los Angeles IRE Regional Conference. Provides pointers to locate sources from the Vietnam War era and modern military.

MILITARY AVIATION

Uncovering secrets

By Russell Carollo
Dayton Daily News

The *Dayton Daily News* began its 18-month examination of military aviation safety with a fishing expedition. I was fishing for computer databases that hadn't been worked over by 100 reporters when I learned that no one had ever asked for any of the databases on military aviation accidents.

So I filed Freedom of Information Act requests with the safety centers for the Air Force, Army and Navy, which includes the Marine Corps. (I gave up on the Coast Guard after numerous telephone calls and requests went unanswered.) Each service has its own database, and, though there are some similarities, the databases differ significantly from one another, even though some aircraft are flown by more than one service.

Pilots who survived accidents and the relatives of pilots who were killed frequently had no idea of the findings we had discovered on the databases.

Major civilian aviation accidents are investigated by an independent body, the National Transportation Safety Board, and the results of NTSB investigations are made public. But major military aviation accidents usually result in two separate internal investigations: One done by one or more people who produce a public report, and the other, a more through investigation, done directly through one of the military safety centers.

Interviews, findings, recommendations and conclusions of the second investigation - the one done through a safety center - are secret, and the courts have upheld this secrecy for years under the so-called Machan Privilege. Factual information, however, is not protected under the Machin Privilege.

So from the beginning, getting information from the safety centers was not easy. I negotiated for months over what was factual and what was protected under the Machin Privilege.

In the end, I got much more than expected. Though the databases differed greatly from one another, each service provided similar types of information: Dates and times of accidents, locations, types of aircraft, cost of accidents, types of injuries, names of crewmembers and even the names of parts believed to have failed. With the identity or parts suspected of failing, I had indications of what the secret safety investigation believed caused accidents - information usually kept from the public under the Machin Privilege.

Under the surface

But reading the databases wasn't easy.

Anything military usually is filled with acronyms and coded language, and military aviation is padded with additional layers of technical jargon. Because of this, stories on military aviation require additional fact-checking, which should include a review by a military aviation expert.

Dave Gulliver, who assisted with some of the data analysis, and I discovered that what appeared to be true on the surface often wasn't necessarily the case. For example, one part that appeared to have been linked to dozens of Army helicopter accidents turned out to be landing skids, something that frequently breaks during crash landings but is rarely the cause.

Using the parts tables, we were able to link a single part or a single type of problem to dozens of accidents, concentrating on parts that eventually caused major accidents. In one case, we found that hydraulic problems in the Sea Knight helicopters had caused more than 71 emergencies and accidents from 1988 through July 1998. Nearly all of the cases involved the utility hydraulic pump, and in at least three cases, helicopters were destroyed and crews barely escaped with their lives.

The Navy database even had a "reoccurring problem narrative," tracking problems

Continued on page six

ECONOMIC DATA

Cost of living

By Jennifer LaFleur
And Michelle Quinn
San Jose Mercury News

When most people think of Silicon Valley, they think of two things: technology and money. If a few national reports are to be believed, nearly everyone in Metro San Jose is a millionaire. Earlier this year, a team of *Mercury News* reporters set out to find out just how much money is being made in Silicon Valley, how it was distributed and how it has affected people's lives.

What our analysis of several private and public databases found was that the valley is out of whack economically compared with the rest of the nation, riding a boom filled with paradoxes.

According to one of the most aggressive estimates, 65,000 Santa Clara County households – roughly 1 in 9 – are worth more than $1 million, not including the value of their homes.

And yet, only 29 percent of Santa Clara County's households can afford the median-priced home, far below the national rate of 55 percent. And two out of five county households would need to stretch beyond advisable limits to rent the average two-bedroom apartment.

The bottom line of using any of these types of data sources is to compare the results to other information.

Perhaps most startling: although many people here earn exceptionally high pay compared with the rest of the country, most aren't particularly wealthy. Look at the net wealth of Santa Clara County households – their assets minus their debts – and the distribution closely mirrors the nation's as a whole.

Data sources

This series used several sources of data. Our biggest concern was that the market data models truly reflected Silicon Valley. Because many market surveys cap home values and incomes low values for our area, we compared data from several sources.

We ended up using wealth measures from Claritas, an Arlington, Va.-based marketing information firm, which estimates income and wealth distributions for the United States, including all ZIP codes and small areas. Estimates are based on a combination of factors: extrapolation of census data, Claritas' Market Audit Survey and other area-specific sources of demographic data. The Market Audit Survey annually collects information on household assets, liabilities, demographics and financial behavior for approximately 120,000 U.S. households. We used Claritas figures for some of the income data.

A note of caution: If you plan to use any market data, make sure you fully understand the methodology used in their model and that you can verify it with other data sources. Also, these data are good for large metropolitan areas, but not as good for smaller areas.

Additional sources of data we used to look at income included: Census Current Population Survey data (downloadable off *www.census.gov* using the Ferret extraction tool), for which we had to combine years to get a valid sample; and California state franchise tax board tax filing data. Using these data, we were able to show that the mean income was increasing faster than the median income, showing that the higher end was pulling up the average.

Housing data

Although many things may seem expensive in Silicon Valley, the one driving factor behind most people's struggles is housing. We used California Association of Realtors data and data from the Real Estate Research Board of Northern California to look at trends in housing costs. Depending on how assessments are done in a particular area, property tax data might be useful for this as well. Because Proposition 13 froze California assessment data at sales prices, it's not a useful tool for looking at housing costs.

Rent data was harder to find. A few private companies do rent analysis, but they vary depending on how the data are gathered.

Continued on page six

The *San Jose Mercury News'* story, "The Cost of Living in Silicon Valley," is available online at *www.sjmercury.com/svtech/news/special/wealth/*

Copies are also available in the IRE Resource Center (story #15718). Contact the Resource Center at (573) 882-3364 to order.

Some military Web sites:

www.militarycity.com: Catch up on the latest military news. If you subscribe to the sponsors (Defense News, Army Times and affiliates), you can search for military personnel in a 4-million record database and get details on military installations. Sponsored by the publisher of Space News, Federal Times, Defense News and Army Times.

http://lcweb2.loc.gov/pow/powhome.html: The Library of Congress' database includes synopses of all of the 131,000 publicly-available documents concerning military personnel listed as missing in action and prisoners of war from the Vietnam War. The site also includes instructions on how to order the full documents.

www.reporter.org/beat/military.html: Find other links to military sites on Shawn McIntosh's Beat Page for covering the military.

Continued from page four:

Secrets

such as those in the Sea Knight that occurred more than once. One field actually totaled the numbers of times a problem had occurred, and the reoccurring problem narrative added comments, sometimes from commanders concerned that the problem had not been fixed.

The Air Force database had an engine table, complete with serial numbers of engines involved in accidents. With this table, were able to track engines that had been in more than once accident. We found one engine that had been on three F-16s in less then four months. Twice the engine was removed after pilots complained of smoke and fumes in the cockpit, and finally the engine was put on a third jet, which crashed into a South Carolina neighborhood, killing one man and injuring seven others.

Beyond numbers

Using the information we gathered from the database, we then sought the lengthy reports on the accidents, filing more than 150 Freedom of Information Act requests with the various services. We also spent months tracking down crewmembers, relatives of victims, attorneys and other with information about the accidents.

One of the most interesting aspects of the reporting came when we shared what we had learned from the computer records and accident reports with pilots, accident investigators and other people who had investigated the accidents. Their explanations greatly enhanced our ability to search the databases for the more significant problems and disregard other problems that - although they appeared significant - actually were minor.

Pilots who survived accidents and the relatives of pilots who were killed frequently had no idea of the findings we had discovered on the databases.

Several people explored legal action as the results of telephone calls from the *Dayton Daily News* concerning our findings, and at least one attorney changed his lawsuit, adding to it a company our records showed produced parts the Army had questioned.

Russell Carollo can be reached by e-mail at rcarollo@coxohio.com

Continued from page five:

Cost of living

HUD Fair Market Rent measures are a good start, although they estimate rents lower than the median for an area. Another source for rent data is RealFacts.

The RealFacts data inventory has been assembled over a period of years, and provides complete and current information on more than 4,000 apartment complexes in 10 Metropolitan Statistical Areas (MSAs), including California, Oregon, Washington, Nevada, Utah and New Mexico. These tend to be slightly higher because they are for apartment complexes.

Income data

To get a handle on how much money insiders in Silicon Valley made, we used a database of SEC corporate insider trading files from Thomson Financial Wealth Identification (formerly CDA). We used this to get an estimate of the number of people with large amounts of wealth.

To figure out the value of stock options held by employees of select Silicon Valley companies, Institutional Shareholder Services did an analysis based on 10-K filings and several other company factors such as turnover.

After analyzing many economic measures for Silicon Valley, we had a starting point. As with any data-based project, the data are the beginning. The bottom line of using any of these types of data sources is to compare the results to other information.

The large amounts of data that we gathered were not all included in the stories, but we were able to make many of them available online so that individuals could look up wealth, income and other measures for their neighborhoods.

Jennifer LaFleur, who now works for the St. Louis Post-Dispatch, can be reached by e-mail at jlafleur@postnet.com

HEALTH CARE

Disciplined doctors

Richard J. Dalton, Jr.
Newsday

A landmark lawsuit, a monstrous database and rigorous searching using healthcare Web site search engines lie behind a recent Newsday series on disciplined doctors working in managed care.

Newsday's series, reported over a 10-month period, revealed that doctors who have been punished for serious or even fatal wrongdoing often continue to work for managed-care companies despite the healthcare industry's promise to screen their doctors and offer the best care to patients.

While many papers have examined HMOs' denial of treatment, this series broke ground by demonstrating that managed-care companies employ disciplined doctors, effectively breaking their promise of quality care.

Staff writer Thomas Maier reported and wrote the series, with computer analysis by this reporter and Dave Ewalt and assistance by Bob Fresco.

The Health Department regularly releases information from the database to researchers and reporters but removes data identifying doctors to preserve their privacy.

Before starting on the series, *Newsday* and *The New York Times* put its lawyers to work, suing the New York State Department of Health to obtain a database of information about the performance of doctors in treating patients in New York hospitals.

The Health Department regularly releases information from the database to researchers and reporters but removes data identifying doctors to preserve their privacy. That meant key information remained secret, including the license numbers of attending and operating physicians working on the patient, the admission date, payor identification number and the hospital name.

The omission of such information made it impossible to determine the quality of care by individual doctors and hospitals.

The newspapers argued that physicians had no such right to privacy but agreed that the data about individual patients should remain private. In June 1998, a state appeals court agreed with a lower-court ruling siding with the media, allowing access to the database of millions of hospital admissions.

Web sites

When the legal battles subsided, the scene of action turned to the World Wide Web for a Health Department Web site listing disciplined doctors, managed-care company sites listing doctors in the networks and the American Board of Medical Specialties' site revealing doctors with board certification.

The healthcare companies and agencies probably never envisioned their Web sites being scrutinized in the manner Maier examined them. From the board-certification site, he found doctors who erroneously claimed board certification in managed-care listings.

From the HMO sites, he didn't find a new doctor, he uncovered disciplined doctors – 132 of them disciplined in New York state in the 1990s.

Using brief descriptions of the disciplinary actions from the Health Department Web site, we created a database of disciplined doctors, including their names, license numbers, addresses, discipline dates and whether the doctors were board certified. We also noted the doctors' offenses, such as patient mistreatment, sexual misconduct or billing errors, and the disciplinary actions, such as suspension, probation or license revocation.

Next, we began to quantify how many procedures these doctors oversaw after they were disciplined, and how much money patients or insurers paid for care overseen by these disciplined doctors.

Data work

The hospital admission data contained millions of procedures performed from 1990

Continued on page eight

The *Newsday* series, "Managed Care and Doctors: The Broken Promise," is available online at www.newsday.com/news/doctors/edocs14.htm

The story, and other investigative stories, can also be accessed through the IRE and NICAR Web site list of online investigative projects at *www.ire.org/datalibrary/online.html*

Continued from page seven:

Doctors

through 1998. Using SPSS, we extracted the procedures overseen by doctors disciplined anytime in the 1990s.

That database contained 130,000 records, but only some procedures were performed after the attending or operating physicians were disciplined. We then matched the database with the list of disciplined doctors to uncover procedures performed after the discipline date.

The discipline date field was straightforward. But, as is common with most government data, the database of hospital procedures from the Statewide Planning and Research Cooperative System, which contained the hospital admission date, was quite quirky.

The admission date in the database was divided into three fields: a year, a month, and a day – not a day of the month, but a day of the week.

We wanted to ensure that we only counted patient visits and hospital charges that occurred after the doctor was punished to demonstrate that taxpayer money was being spent on medical care overseen by disciplined doctors. But the unusual format of the admission date presented a problem when a patient visited a doctor in the same month the doctor was disciplined.

So we decided to turn every admission date into the first of the month. If a patient was admitted in November 1998, we turned that into Nov. 1, 1998, using a module written in Microsoft Access and an update query.

Suppose the patient had been treated in November 1998 and the attending physician was disciplined on Nov. 20 of that month. Using Nov. 1, the visit would have preceded the discipline date and wouldn't be counted. We gave the doctor the benefit of the doubt because it was impossible to be more specific about the admission date.

Once we cleared up the date dilemma, we began to calculate the amount of money that the state had spent on managed-care doctors through Medicare and Medicaid.

At first, the task seemed easy – a field indicated the charges. Then came the caveats. The field represented the total cost of the hospital visit, including the room, surgery and payments to the attending and operating physicians and others.

Furthermore, Maier discovered the actual charges to Medicare and Medicaid turned out to be about 60 percent of the total cost.

So we discounted the figures by 40 percent, and could only say that the disciplined doctors "oversaw" the hospital visit, explaining why the data prevented us from determining how much each doctor actually received.

License problem

Each procedure record also contained three licenses of doctors associated with the treatment, listing the licenses for the attending, operating and "other" physician.

We chose to use only the attending and operating physicians. Still, multiple licenses for each record presented a quandary.

It's impossible to link the two licenses – one from the attending physician and one from the operating physician – to the one license field in the doctor table.

To explain why, here's an analogy. Suppose I have a lookup table named States that translates U.S. Postal state abbreviations (e.g., State.code contains "NY," "NJ," etc. into state names, and State.name contains "New York," "New Jersey," etc.).

Suppose I have an address book that lists someone's work address and home address, and I want to link that to the lookup table of state names. Suppose further that a friend lives in New Jersey and works in New York.

That state code for the work address, "NY," and home address, "NJ," cannot simultaneously point to the state code in the lookup table for state names. If it did, then what would the field State.name represent: New York or New Jersey?

To resolve that dilemma, add the doctor table to the query twice by clicking twice on the "add" option in the query/show table window. That adds another table called State_2.

To use it, link the home address to the state table and the work address to the State_2 table. Now State.name is "New York" and State_2.name is "New Jersey."

By the way, Access doesn't really create another table. It merely creates another pointer to the same table.

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

Following is the Microsoft Access module.
To use it, run an update query, replacing the admitdate field with the following: firstdate([admitmo], [admityr]), where admitmo and admityr are the fields representing the month and year of the admission.

```
Option Compare
Database
Option Explicit

Public Function
FirstDate(mCnt As
Long, yCnt As Long)
As Date
Dim dCnt As Byte
Dim y2Cnt As Integer
Dim NewDate As Date
y2Cnt = yCnt - 1900
NewDate = #1/1/
1900#
NewDate =
DateAdd("yyyy",
(y2Cnt), NewDate)
FirstDate =
DateAdd("m", (mCnt -
1), NewDate)
End Function
```

Where to look for data

By MaryJo Sylwester
IRE and NICAR

The *Newsday* series, "Managed Care and Doctors: The Broken Promise," tackled a topic that is becoming increasingly important for journalists to analyze as the health care industry continues to grow larger every year.

Whether you're at a large or small news organization, looking at health care in your community is possible.

There were three primary databases that *Newsday* used for its series:

• Hospital procedures performed at local hospitals. Use this to calculate the number of hospital visits and charges related to disciplined doctors. (*Newsday* and *The New York Times* sued to gain access to this database.)

• Doctors who have been disciplined. Each state has a regulatory agency for licensed professionals, which would be able to provide this information. (A Web site in New York lists the doctors, but it is not in database format. *Newsday* received this database electronically through a Freedom of Information Act request.)

There are plenty of industry sources where you can get averages or typical numbers for hospital finances, such as how much debt they have or how much money they make from outpatient versus inpatient procedures.

A list of board-certified doctors. One place to check is the American Board of Medical Specialties, an umbrella group that represents 24 specialty boards. (*www.certifieddoctor.com/verify.html*)

Finances

Another area of the medical industry that could be scrutinized, without requiring a great deal of resources, is finances of local hospitals or clinics. Most health care organizations have some sort of government regulation, and therefore financial reporting requirements. Each state should have a government agency overseeing health care, where you could obtain a great deal of information.

Some hospitals are non-profit organizations, which are required to file Form 990s. The Internal Revenue Service Web site *(www.irs.gov)* lists organizations that have filed the tax-exempt forms, including some basic information for each one. However, the forms are not available there.

Recently, some IRS-990s have been posted at the Guidestar site (*www.guidestar.org/news/features/990status.html*) or on the National Center for Charitable Statistics site (*http://nccs.urban.org/990/*).

Other sources

There are plenty of industry sources where you can get averages or typical numbers for hospital finances, such as how much debt they have or how much money they make from outpatient versus inpatient procedures.

The Health Care Financing Administration may be able to provide some good information. Their Web site (*www.hcfa.gov*) has some statistical data, such as payments to hospitals and national statistics you can use for comparison purposes.

On its Web site, the Department of Health and Human Services (*www.dhhs.gov/progorg/oig/cumsan/index.htm*) has a list of people who can't take Medicare or various other federal payments for services, and other useful information.

Before launching into a story of this type, you might also want to check the IRE Resource Center. They have a large number of not only stories that have been done by other news organizations on health care topics, but also some tipsheets that could be useful. You can search the Resource Center database at *www.ire.org/resourcecenter.*

MaryJo Sylwester can be reached by e-mail at maryjo@nicar.org

Web sites recommended by Richard Dalton:

The list of doctors disciplined in New York State is available at *www.health.state.ny.us/nysdoh/opmc/main.htm*

The Healthcare Association of New York State lists top state insurers on its Web site, *www.hanys.org/resource/mancare/hmotable.htm*

The American Board of Medical Specialties Web site lists the board certification of doctors at *www.certifieddoctor.com/verify.html*

The Money Trail: Tracking Campaign Cash

These IRE and NICAR workshops on using campaign finance data will be held over the next few months in various cities. More information is available at the Campaign Finance Information Center Web site, *www.campaignfinance.org/seminar.html*

Here are the dates and locations:

Jan. 22-23: Lansing, Mich.

Feb. 5-6: Columbus, Ohio

Feb. 26-27: Springfield, Ill.

March 4-5: Madison, Wis.

March 25-26: Indianapolis, Ind.

CAMPAIGN FINANCE

CEOs' giving habits

By Richard Dunham
Business Week

In the hyper-competitive world of presidential campaign reporting, it's a constant challenge to stay ahead of the pack – particularly if you work for a magazine that is published weekly.

But the best way to get noticed, I've discovered as White House and national political correspondent for *Business Week*, is to create a specialized niche in the overcrowded marketplace of political stories. For me, that's been the political donation habits of America's top CEOs.

Early in the year, before the presidential candidates filed their first disclosure forms with the Federal Election Commission, I wrote a story about the early giving habits of prominent business figures. This was done the old-fashioned way: a lot of phone calls to a lot of executives and to all of the campaigns.

As the campaign heated up after Labor Day, I devised a plan to take the next step: to comprehensively study the donation patterns of big businesses' top corporate executives. To beat the competition, we needed to do the analysis as soon as the 1999 third-quarter campaign spending reports – due on Oct. 15 – were made public.

Project goals

At the start of the project, I composed a list of reporting goals: A simple scoreboard of the candidates' total number of CEO donors, an analysis of the extent to which CEOs "hedged their bets" among different candidates, a breakdown of corporate givers by economic sectors (such as technology, telecommunications, entertainment, finance and energy) and a review of the 2000 donations made by CEO backers of the 1996 Clinton-Gore team.

As a reporter with good (but not state-of-the-art) computer skills, I knew that I'd need a little bit of help to complete my analysis within days of the data becoming available. I got that assistance from Campaign for America, a nonpartisan watchdog group that enlisted two of the most knowledgeable money-in-politics experts, Kent Cooper and Tony Raymond.

We compiled a list of the top 1,000 corporations by market capitalization, with their addresses and chief executive officers, and arranged it in Excel.

Cooper and Raymond checked the list against the new FEC reports and came up with more than 5,000 possible hits.

We had to pare down the list by hand, eliminating people with identical names living in different states or people with different middle initials.

Just under 600 potential donors made the cut. We "scrubbed" the data one more time to make sure we didn't confuse fathers with sons, such as "Smith Jr." with "Smith IV."

Our final list had 538 contributions by CEOs. Some of these executives had donated to more than one candidate (some to as many as four White House wannabes) and some had sent multiple contributions to a single candidate. We even found four individuals who were CEOs of two different companies, so their contributions were listed twice in our database. Again, we adjusted our totals.

In the end, we found 389 contributors.

Economic sector

To analyze the contributions by economic sector, *Business Week's* Fred Jespersen sent us the magazine's computerized listing of corporations coded by economic sector. This is proprietary data, so it can't be replicated by outsiders, but a reporter can create a reasonable facsimile by looking up each corporation in a business directory like Standard & Poor's publication.

With the sectoral data now merged in our Excel files, we rearranged our data by sector and tallied the results.

As expected, George W. Bush had a huge lead among energy company CEOs. But he had an unexpectedly large lead over Al Gore among the Vice-President's beloved technology companies. In fact, Gore placed third behind Bill Bradley in techland. What's more, Bradley matched Gore in contributors among the big Hollywood CEOs.

There are computerized lists available of anything from lawyers to bankers to professional athletes. A regional reporter could pare down a list of CEOs (or bankers or lawyers) to include only those from a certain state or region.

Richard Dunham can be reached by e-mail at richard_dunham@businessweek.com

Online tax calculator

By Genevieve Anton
The Colorado Springs Gazette

Tax cuts are extremely popular among most voters, but difficult to write about in a way that hits home with your readers. These are classic pocketbook issues, yet our stories are usually heavy on numbers, bureaucratic language and political spin.

Colorado had a $704 million surplus this year, a state record that prompted historic tax cuts. The question wasn't whether to cut or spend – constitutional spending limits mean surpluses must be returned to taxpayers – but who would get the tax breaks.

When the proposed tax package emerged from the Colorado General Assembly, my statehouse colleague Michele Ames and I scrambled to make sense of it to readers. Who would be the real winners and losers if this passed?

With only three days to write the first analysis for the Sunday edition, we decided to find several real people in our area, each representing a different socio-economic situation, and figure out how the specific tax proposals would affect them.

What we needed was a way to let individual taxpayers see how the proposed tax cuts would affect them directly.

Based on information they provided about their age, marital status, taxable income and investment exclusions, we calculated their individual tax bill under the new plan. The same type of analysis was done for a small business and a large manufacturer in our area.

We bolstered these findings by creating imaginary taxpayers and plugging their financial data into algebraic formulas for each tax cut proposal to get a larger sense of how it would impact various taxpayers.

It was a crude and tedious process, one that we desperately wanted to expand and refine.

But there's a tight lid on information about taxpayers. The Colorado Department of Revenue said it had no reason to track certain information – even on an aggregate level – such as how many people currently file for certain exemptions, exclusions or deductions.

What we needed was a way to let individual taxpayers see how the proposed tax cuts would affect them directly.

That's when I heard about a creative approach to tax coverage being tried by New Hampshire Public Radio: a Web site devoted entirely to the issue in that state. Although their Web site was a massive project that had taken months to pull together, there was one element we could copy: an online calculator that lets people see how the tax cuts affect them directly.

Since the New Hampshire Web site hid the coding and formulas for the calculator, we had to start from scratch with little to go on.

Fortunately, our editors loved the idea. Unfortunately, they had no clue how difficult it would be and gave us only one week to pull it all together.

Calculator creation

In creating the calculator, we had no experience or examples to follow, and not even the state's top budget or tax experts had attempted anything like this – although some did offer to look over our work when done.

The idea was to give people a rough estimate of what they would save under each proposed tax reduction and a comparison of their final tax bill under the entire tax package versus the alternative – a straight refund of the surplus.

So we took each proposed tax cut and broke it down into an algebraic equation based on revenue formulas that ranged from one line to two pages in length.

For example, a proposed reduction of the state income tax rate from five to 4.7 percent looked like this: (your taxable income * .05) – (your taxable income * .0475) = tax savings.

No, we didn't make the public walk through these tortuous calculations.

We simply figured out what information we needed to know and narrowed it down to

Continued on page sixteen

The *Colorado Springs Gazette* Web site on tax cut proposals, "It's Your Money," – including the online calculator – is available at: *www.gazette.com/yourmoney/taxcal.html*

Through the NICAR-L listserv you can ask questions about computer-assisted reporting, offer advice to others or simply see what's being talked about on the list.

To join, send a message to:

listproc@lists.missouri.edu

In the message area type:

SUBSCRIBE NICAR-L [your name]

More information about the listserv is available on the NICAR Web site, *www.nicar.org*

Weighting data

By Griff Palmer
The Daily Oklahoman

In pursuing a recent education project, *The Daily Oklahoman* staff writer Bobby Ross persuaded Oklahoma City schools administrators to give us an electronic file containing all students' addresses and school assignments.

We wanted to come up with demographic indicators for the students. We explored linking student addresses to county assessor data, using home values as an indicator of economic status. We gave up that idea, though, after realizing that dealing with students living in apartments and rental properties would be impossible, given the limitations of the Oklahoma County assessor's data.

We settled on using 1999 block group-level median income estimates and 1990 educational attainment data from Claritas as indicators for students. The question of how to properly weight the data generated lots of interesting discussion on NICAR-L, NICAR's electronic mailing list for CAR practitioners.

How we did it

To link students up to census block groups, we first geocoded students' addresses with MapMarker software. We used Atlas GIS to assign each geocoded address to its appropriate block group. It was then a simple matter with FoxPro to write the median income and educational attainment data for the corresponding block group to each student's record.

Once we had assigned a block group level median income to each student, we wanted to calculate a median for each school in the district, and here, the question of weighting came into play. At most schools, most students live in one or two block groups. Those block groups should carry more weight when calculating a school-level median than do block groups in which just a few students live.

Using a program like SPSS, it's simple to apply a weighting factor to a variable.

As the statisticians in *The Oklahoman's* marketing department explained to us, though, it's not necessary to apply a weighting factor to determine a weighted median if you have a dataset with every case in it and not grouped summaries. The weighting occurs naturally as medians are calculated. When ranking a set of values to determine the central value, a category (in this instance, a block group) encompassing a large number of cases (students, in this instance) automatically carry more weight in the rankings.

To give a simple example, say you have three cases with block group level medians assigned, and you want to calculate a median from those three cases:

1	24,586
2	29,483
3	32,771

The median of the three values is 29,483. But say you have five students, three of whom live in the same block group and therefore have been assigned the same representative median income. If you rank the cases, the value that occurs more than once weights the median:

1	24,586
2	24,586
3	24,586
4	29,483
5	32,771

To double-check, I threw all 40,200 cases into SPSS and ran case summaries, calculating the median of median incomes, grouped by schools. Then, I used FoxPro to group the 40,200 cases by school and block group. The query created a table listing, for each school, all block groups from which the school was drawing students and the number of students it was drawing. I opened the resulting table into SPSS, weighted the data by number of students in each block group, and then calculated the grouped median for each school.

The results were the same both ways.

SPSS's ability to weight grouped data by frequency is one of the many features of this program that make it a powerful tool for data analysis. Calculating weighted medians of grouped data in Excel can be done, but that's another article.

Commercial vendors

As we worked with our schools project, and discussed the nuances of weighting data, we also found ourselves grappling with the question of whether and how to use market-

Continued on page twenty-one

IN THE CLASSROOM

The future of CAR

By David Kesmodel
IRE and NICAR

When student journalists Katie Crecente and Summer Parrish got a tip last year that prices for the same drug prescriptions varied widely among pharmacies in their community, they thought they might have a story. An analysis of an Excel spreadsheet confirmed their hunch.

The University of South Carolina students asked the pharmacies for prices on four types of drugs and built their spreadsheet in a computer-assisted reporting class. Crecente reported the story on the local cable TV system and won first place in the student competition for the Southern Regional Emmy Awards.

Crecente and Parrish are not alone. Students in dozens of U.S. journalism and mass communication programs are learning to analyze information in spreadsheets and database managers, and to find the story behind the numbers.

Many of the schools now teaching CAR based their approach on the program at the Missouri School of Journalism, in Columbia, Mo., which just completed its 10th year. The program has also led the way by training dozens of professors at boot camps offered by NICAR, a joint venture between the school and Investigative Reporters and Editors.

The issue many journalism educators say they are now wrestling with is not whether to teach CAR skills, but how and to what degree. Increasingly, they are trying to decide whether to teach the skills in a stand-alone class, weave them into other reporting and writing classes, or do both.

According to an informal e-mail survey of 100 accredited journalism and mass communication programs for this story, educators appear to be evenly split in their approaches to teaching CAR. One half of the 64 respondents said they offered a stand-alone CAR class, although some do not offer the class every semester.

Of the 32 programs that don't teach a stand-alone CAR class, 25 said the skills are taught in other courses. And most of the schools that offer a freestanding class say they also try to integrate the skills throughout the curriculum.

Steve Doig, the Knight Chair at Arizona State University's Cronkite School of Journalism and Telecommunication, said the goal of trying to teach all journalism students basic CAR skills is difficult to achieve, particularly in large schools like his. "I have 15 students each semester in my (CAR) class, 30 in a year," he said. "There are 1,400 students in the Cronkite school. I can't teach them all."

Doig said the school is tinkering with ways to circulate CAR throughout the curriculum. His current tactics include introducing students to online resources in his basic news writing class and presenting guest lectures on CAR in his colleagues' classes.

The classes

The names that are given to stand-alone CAR classes vary, including precision journalism or computer-assisted reporting. Some schools spend a lot of time teaching online research skills, while others concentrate on spreadsheets and database managers.

At the Missouri School of Journalism, CAR is taught in a stand-alone class and incorporated into an investigative reporting class taught by Brant Houston, executive director of IRE and NICAR.

"Part of our mission here is to make sure this tool is blended into good reporting," said Houston, author of the textbook, "Computer-Assisted Reporting: A Practical Guide."

One of his students, Mark Greenblatt, used National Bridge Inventory data to report on failing bridges in Missouri. (see Greenblatt's story on page fifteen.)

For the stand-alone class, Jo Craven trains students to use Microsoft Excel and Access and to negotiate for data from government agencies. In addition to weekly lessons, students work on two semester-long projects.

Craven wants students to take away from her class the ability to logically go about their reporting, whether the story involves a dataset or not. "I want them to learn to respond to those bells in their head that say, 'Something here is not right.' The worst thing they can do as reporters is have a funny feeling and not follow up on it," she said.

One of her students, Charles Choi, recently built a spreadsheet of data on check fraud in Columbia and wrote a front-page

Continued on page fourteen

Tipsheets on CAR in the classroom, available from the IRE Resource Center:

- **Tipsheet #885, "Checklist of spreadsheet skills," by Jeff South of Virginia Commonwealth University, for the 1999 National Computer-Assisted Reporting Conference. Tipsheet provides a checklist of CAR skills to help students gauge what they've learned (or have to learn) and to make sure instructors cover the basics with each class. Other lists can be found at: *http://saturn.vcu.edu/~jcsouth/carskills/skillindx.html***

- **Tipsheet #886, "Graduate seminar in research syllabus," by Teresa Allen, Boston University, for the 1999 National CAR Conference. Syllabus is provided for a class designed to help students locate and obtain public documentary information, how to read for newsworthy facts and how to cull facts into powerful news stories.**

"Computer-Assisted Reporting: A Practical Guide," by Brant Houston. It can be ordered from IRE and NICAR for $25 for IRE members or $30 for non-members plus shipping. Call (573) 882-2042 to order.

IRE and NICAR can bring specialized training into newsrooms, including the necessary equipment and useful handouts.

Costs vary depending on the number of people trained and the difficulty in developing training materials. Contact Trainer Tom McGinty at tmcginty@nicar.org or (573) 882-3320 or Executive Director Brant Houston at brant@ire.org or (573) 882-2042.

Continued from page thirteen:

CAR classroom

story for the Columbia Missourian.

Choi said journalism schools with stand-alone CAR classes are giving aspiring journalists more reasons to get a degree in journalism, as opposed to a liberal arts degree.

"This makes journalism school more important," he said. "Not everybody needs a V-8 engine, but it's really nice."

Factors

A lot of factors affect a school's approach to teaching CAR, including faculty members' philosophical views about the subject, whether adequate resources are available, and whether faculty members are qualified to teach CAR.

The bulk of the accredited programs offering a stand-alone CAR class teach it as an elective, because under the accreditation rules, students must take about three-fourths of their classes outside their major, and the schools can offer only so many core journalism classes.

"Teaching CAR as a separate class perpetuates the illusion that computer competence is an isolated and esoteric specialty accessible only to a few," Philip Meyer said.

Accreditation has influenced the decision of the journalism department at Washington and Lee University not to offer a freestanding CAR class, said Ham Smith, department head.

"In the highly restricted accredited curriculum, it does not seem wise to require a CAR course," he said.

Meanwhile, he argued, teaching CAR as an elective is problematic.

"Isolating CAR in a non-required course unfairly limits the number of students knowledgeable" in CAR, he said, "so we dribble it out everywhere as a standard reporting activity."

Philip Meyer, a professor of journalism at the University of North Carolina-Chapel Hill and a pioneer in using computers to enhance reporting, feels strongly that his school should not offer a stand-alone CAR class. At UNC, the skills are taught in other classes.

"Teaching CAR as a separate class perpetuates the illusion that computer competence is an isolated and esoteric specialty accessible only to a few," Meyer said.

Virginia Commonwealth University also has subscribed to the theory that CAR should not be taught as a separate class, said Jeff South, an associate professor in the School of Mass Communications.

But what South and some other educators say they have discovered is that when you don't teach a specialized class in CAR and depend on faculty members to integrate the skills in other courses, the job doesn't always get done.

"When everyone is responsible for CAR, it can turn out that no one is responsible for CAR," South said. At VCU, he said, "some faculty members did not have adequate CAR training and did not feel comfortable teaching CAR."

Now, the school is looking into requiring a course in information gathering, which would introduce students to online resources and perhaps to spreadsheets and database managers.

With time, say many educators, stand-alone classes will be phased out as CAR becomes a more routine reporting tool. Schools may teach advanced classes in CAR, but more and more students won't blink when handed data in a spreadsheet.

"Hopefully the concept of using a spreadsheet and going online will become so ingrained that treating it as a separate weird specialty will no longer be important," Doig said.

David Kesmodel can be reached by e-mail at david@nicar.org

FIRST VENTURE

Troubled bridges

By Mark Greenblatt
University of Missouri

I wanted to find an easy database to work with for my first computer-assisted story. I had taken a CAR course at the University of Missouri-Columbia earlier, but had put off applying what I learned. My only goals were to use the CAR skills I learned and also to play around a little bit with my brand new computer.

Since bridges have always fascinated me, I thought I'd take a look at the conditions of the ones in my area for a project in an investigative reporting class, taught by IRE and NICAR Executive Director Brant Houston.

I wanted to play it safe the first time and avoid getting a database with dirty data, so I obtained my copy of the bridges inspection database for Missouri directly from NICAR. I had no idea what I was about to get myself into.

It turned out we could only use the database as a tipsheet to finding bridges of interest, and also to find the bigger picture.

Three months later, the story I uncovered would have Missouri lawmakers calling for quick reform. The state of the bridges in our area turned out to be in dismal shape.

Starting out

KOMU TV's news director Stacey Woelfel and I originally thought we could do a quick story to see if any bridges open to traffic in Columbia and around our viewing area were in bad shape. We had heard rumors of bad bridges but had nothing nailed down. It would be easy, I thought, to simply run a few queries on the database, come up with a list of the worst bridges, and broadcast the story after confirming a few facts with local bridge inspectors. I couldn't have been more wrong, and that turned out to be good.

Just learning about the database took me a couple weeks of playing around on my spare time. The record layout of the National Bridge Inventory was more than 100 pages long, and it included a lot very technical terms like sufficiency rating formulas, structure numbers and deck geometry.

The sufficiency rating is a computer-generated formula that is supposed to represent each bridge's overall condition. I ran different queries showing the bridges in mid-Missouri with the lowest sufficiency ratings. With that list in hand, I contacted a local bridge inspector with the Missouri Department of Transportation to learn more.

He immediately told me that using sufficiency ratings was the wrong way to do this study. The only way to look at bridges is to look directly at the safety ratings for their substructures, superstructures and decks, he said. He took me out on bridge inspections for a day to teach me about his job.

By the end of the day, I had determined there was no accurate way we could use the numbers in the National Bridge Inventory to point us towards any list of the 10 worst bridges in mid-Missouri. It turned out we could only use the database as a tipsheet to finding bridges of interest, and also to find the bigger picture.

New direction

I went back to my computer and started running more queries. I ranked the mid-Missouri bridges with the worst substructures and the worst decks, and for the first time grew concerned about the sheer numbers of bridges in bad shape in our area. Far too many bridges were rated in poor condition, and many of them carried thousands of cars a day on major highways.

After examining the written records of countless bridges to match them with the computer data for accuracy, I noticed a common trait. All the bridges the computer data pointed to as in bad shape were all determined to be structurally deficient. That means they can't carry the legal loads they were designed to hold.

A query we ran showed that close to one-third of the 3,000 bridges in mid-Missouri

Continued on page sixteen

Other stories using the National Bridge Inventory database, available from the IRE Resource Center, include:

- "Tennessee Bridges," (story #15293) by WSMV-TV in Nashville, Tenn., 1998.
- "Bridges on the Brink," (story #12972) by the *Rochester Post-Bulletin*, Rochester, Minn., 1994.

Contact the Resource Center at (573) 882-3364 to order.

The National Bridge Inventory database, 1994-1998, is available from the NICAR data library either for the entire United States or for an individual state. Prices vary depending on news organization size.

To order, call the data library at (573) 884-7711.

More information is available on the NICAR Web site at *www.nicar.org*

Continued from page fifteen:

Bridges

were in that bad shape. The break in this story came when the computer data pointed us towards one bridge that seemed like it should be closed based on its ratings. The written report said the bridge was in "exceptional" condition.

It turned out the Missouri Department of Transportation keeps a list of 19 bridges in mid-Missouri that are "exceptionally" deficient, yet still open to traffic. One of a few inside sources I had developed anonymously passed that list on to KOMU.

The list was remarkably frightening. It showed the Missouri Department of Transportation knew the bridges in mid-Missouri were falling apart in some cases, but it had no immediate plans to fix them. One bridge we focused on was described as having "concrete periodically falling into traffic lanes" on Interstate 70.

We queried the database again, and found out it would take 49 years to fix all the bridges that needed repairs in mid-Missouri at the rate the state was going. The problem was, state officials told us the average lifespan of a bridge is only 50 years.

In the end, the computer database we used added depth to the story and made it possible to more quickly uncover larger, specific problems.

After the story aired, the Speaker of the House of the Missouri House of Representatives requested a copy of the story to use in the Legislature to push for reform. Other lawmakers also called, and an internal MoDOT e-mail was forwarded to me that showed MoDOT was "putting everything else on the backburner" for a while to examine what they can do about their bridges in mid-Missouri.

I think it's safe to say I'm hooked on CAR after doing this project. I can't wait to get ahold of another set of data.

Mark Greenblatt can be reached by e-mail at mpg680@mizzou.edu

Continued from page eleven:

Tax calculator

10 basic questions such as: What's your taxable income, are you married, and what's your investment income from interest, dividends or capital gains?

You then assign a value to each answer, which is used behind the scenes in your calculator's mathematical formulas. The result popped up in a separate, easy-to-understand format that basically said, "Here's how much you'll save on each of these proposals."

To see how this works, you can visit the real thing at *www.gazette.com/yourmoney/taxcal.html*

Testing it out

Once you've put together the formulas (by yourself or with help from someone who understands both mathematics and the policy involved,) you'll probably need help translating it into computer-speak for the Web site. Each server may recognize a different language, but we used Java Script. We found a computer programmer to help us for a small fee (once you write out the formulas it's pretty straightforward work for those who know how.)

Once it's on the site, you must run tests over and over using different scenarios and compare it to your own calculations on paper to make sure it works.

We found numbers with too many decimals or result boxes that turned up blank or simply outrageous results like millions in tax savings for a poor, single mother. If it doesn't work, you'll have to go over the formula step-by-step to find the problem and rewrite the calculation.

Believe it or not, we met our deadline, and more than 500 readers visited the site in the first week to look at proposed tax cuts and try out the on-line tax calculator. Feedback forms also helped generate great stories and new ways of thinking about how we could cover the tax debate in the future.

Genevieve Anton can be reached by e-mail at ganton@iex.net

Continued from page one:

Race tabulations

want to look at racial and ethnic changes since 1990.

In both 1980 and 1990, the Census (and the federal government at large) officially recognized five racial groupings: White, Black, American Indian, Asian or Pacific Islander and Other Race. Respondents could pick only one of those choices.

Notice that "Hispanic" is not one of the choices, because "Hispanic" is not a race. Instead, the Census had a separate question asking respondents if they were "of Spanish/Hispanic origin." The result of both the race and Hispanic-origin questions is that a person could be counted as one of a total of 10 possibilities (discounting the various subcategories).

OMB ruling

However, this relatively tidy state of affairs has changed dramatically for the 2000 Census. The reason is a 1997 ruling by the federal Office of Management and Budget (OMB) that the Census Bureau and other federal data collectors must start allowing respondents to pick more than one racial category if they wish. The ruling also split the 1990 "Asian or Pacific Islander" racial category into two new categories: "Asian" and "Native Hawaiian and other Pacific Islander."

The OMB ruling is a good thing in many ways. It certainly reflects the social reality of this increasingly diverse country better than the old one-race-only choice. It stops forcing people to declare only one part of their heritage. And it will make a much richer source of data for research and stories.

But this change also creates a potentially huge problem for those of us who want to compare race/ethnicity in 2000 with 1990 and earlier. The fact that respondents can pick more than one category means there are a total of 63 possible categories of races alone or in various combinations, ranging from "White alone" to "White, Black, American Indian, Asian, Pacific Islander and Other Race." (Anyone you can find from this category probably is worth a human-interest story in their own right!) And you need to double that number to account for whether respondents are of Hispanic origin or not.

The 2000 Census initial data release, the PL94-171 reapportionment data, will collapse these 126 possibilities into a somewhat less cumbersome array of summary tables. Looking just at 2000 data, these tables will give journalists and researchers a good picture of where each group stands in relation to the others. But there is no table that is truly comparable to the 1990 ten-category table because the 2000 tables include a brand-new category: "Two or more races."

Nor can the "two or more races category" simply be ignored, except perhaps in parts of the country where the population is overwhelmingly white. The 1998 dress rehearsal conducted in parts of California, South Carolina and Wisconsin found that up to five percent of the respondents (in Sacramento) chose two or more races.

This means it will be up to the data users like us to choose an appropriate algorithm to transform the publicly-available 2000 tabulations into something we can use to do comparisons with 1990 data.

One way the 2000 Census deals with the multiple race category is to create a series of tables giving the count of "Race X alone or in combination with one or more other races." Such tables are created for each of the major race categories.

But you can't simply build a 10-category table comparable to 1990 using these. The reason is double counting. A person who claims White and Black ancestry will be counted both in the White table and in the Black table.

I tested this method with the Dress Rehearsal data for Sacramento and found that

Continued on page eighteen

"United States Census 2000," (#938), by Steve Doig and Peter Bounpane, Arizona State University, for the 1999 National Computer-Assisted Reporting Conference. This tipsheet includes four handouts on potential shortcomings of the census, the U.S. Bureau of the Census' official plan, and data products that will be available.

"Data analysis and census data," (#945), by Jennifer LaFleur, formerly of the San Jose Mercury News, and Philip Meyer, of the University of North Carolina-Chapel Hill for the 1999 CAR Conference. The tipsheet contains analysis tools for working with the census data.

Census 2000

Census glossary (From handout #809, by Paul Overberg of USA Today):

CCC: Complete Count Committee. Local advisory groups assembled to get the word out and cue the Census Bureau to local circumstances.

ICM: Integrated Coverage Measurement. A national survey of 750,000 households that will be done just after and matched to the census, duplicating its core in miniature. It will be used to double-check the accuracy of the data and, Congress willing, to adjust totals.

NRFU: Non-response follow-up. If you don't mail back your census form, a live human comes to interview you. Or eventually, your neighbors, if you can't be found.

some tracts produced population totals as much as 16 percent greater than in reality.

Possible methods

The demographics experts at the Census Bureau are well aware of the implications of this change and are wrestling with ways to make the data the most useful. Some of the possible methods require access to individual respondent records, which because of the confidentiality requirement is something that only the Bureau itself has. For instance, the "equal fractions" method involves splitting multiple-race respondents into their components, so to speak. Thus, a respondent who lists himself as "White" and "Asian" would add one-half a person to the White tally and one-half a person to the Asian tally.

Such a retabulation using individual record access may be done by the Bureau at some point during the data release cycle. But it appears as if the first major data release, the PL94-171 data, won't include a retabulation. This means it will be up to the data users like us to choose an appropriate algorithm to transform the publicly-available 2000 tabulations into something we can use to do comparisons with 1990 data.

Luckily, prime time still is a year away – the first real data won't appear until early in 2001.

Some retabulation methods can be applied by non-Bureau researchers. One such method (called "proportional distribution") involves spreading the "two or more races" tally among the major single race/ethnic categories based on each of those categories' share of the total. For instance, assume that a census tract had 100 multiple-race respondents, and that the rest of the tract's population was 60 percent non-Latin White, 20 percent non-Latin Black, 10 percent non-Latin Asian and 10 percent Latin White. Under the proportional distribution plan, 60 of the multiple-race respondents would be added to the non-Latin White total, 20 to the non-Latin Black total, and so forth.

This approach is attractive for at least a couple of reasons. It uses mutually exclusive categories that produce a total population equal to the actual count. And it can be done with relative ease on a spreadsheet or with programming.

However, it raises the likelihood of creating "fictional" respondents. In the example above, it may be that all of the 100 multiple-race respondents actually listed themselves as non-Latin "Asian and White." In that case, this method would have created 20 Black respondents who don't actually exist.

Another objection to this method is that it arguably overweights the White category at the expense of the minority races. A primary reason for redoing the tabulations into a format compatible with 1990 is so that changes in racial segregation and employment opportunity can be measured. Under most anti-discrimination laws, a person would be counted as a minority eligible for protections and preferences if one parent was a minority and the other parent was White.

By that standard, all the multiple-race respondents should be distributed among the minority race categories, yet the proportional method will put most of them in the White category. Nor can you solve it by simply taking non-Latin White out of the proportionality formula; if so, then you typically wind up with too-high counts for one of the minority categories.

I've played with yet another approach, assigning different parts of the 2000 tabulations to each of the 1990 categories. My still-in-development method seems to allocate the multiple-race respondents pretty well. But it doesn't produce a total population equal to the actual count (although it usually comes quite close.) So my method, too, needs some more study before it's ready for prime time.

Luckily, prime time still is a year away – the first real data won't appear until early in 2001. Until then, other census reporters and I – and the real experts at the Census Bureau – will continue chewing on this problem.

So the best advice I have now is, "Stay tuned!"

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

TECH TIP

Parsing nightmare

By MaryJo Sylwester
IRE and NICAR

Many times there is more than one solution to a problem – as Robert Gebeloff discovered in late November when faced with an unusual parsing dilemma.

His post to the NICAR-L thread elicited about eight different solutions.

Here's the problem: a state agency had grudgingly turned over some data in ASCII, delimited with semicolons, with each record ending in the person's name. The big problem, though, was that there wasn't a delimiter between records. Here's a sample:

```
GRADY,SELMA ;8M MIGLIORE MANOR
;ELIZABETH;NJ;07206-64; NB00010500;
Z;99;19340127;SELMA GRADY
```

Gebeloff wanted to know if there was a way to construct a search/replace that would recognize the name reversal at the end of the record and insert some kind of delimiter.

Here's a few of the suggestions, including the ultimate solution Gebeloff used:

Suggestion one:

Matthew Ericson, of *The Philadelphia Inquirer*, suggested using Perl to construct a search-and-replace expression that will put a line break after each record:

```
perl -p -e 's/((.+?),(.+?)\s*;.+?)\3
\2\s*/\1\n/g;' datafile.txt
```

More information about Perl is available at *www.perl.com*

Suggestion two:

Paul Skolnick, of Thunder & Lightning News Service in Belleair Bluffs, Fla., suggested this solution: run the text file through a macro in a (newer version) word-processing program to parse the name fields. Since the first and last words in the records are the same, you would insert a carriage return-line feed after the second appearance and set that off as one distinct record.

The macro recorder in Word 2000 is in the Tools menu. It has a Record function in it, but you would have to doctor the recorded macro some.

Suggestion three:

Offered by David Heath, of *The Seattle Times*, this solution preserves the original data, so commas before first and middle names are kept. This program takes bits of the file to work with at a time, which requires more code, but ultimately makes the program run faster.

```
Declare variables, including an array
for the nine fields.
Dim intReadFile, intWriteFile As
Integer
Dim strLine, strTempLine As String
Dim strField(1 To 9) As String
Dim x, ctr As Integer
Dim lngCheckName, lngLengthOfFields
As Long

Open the text file and the output file
intReadFile = FreeFile
Open "d:\data\parse\newhha1.txt"
For Input As #intReadFile
intWriteFile = FreeFile
Open "d:\data\parse\output.txt" For
Output As #intWriteFile

Print the field names in the first line
of the output file
Print #intWriteFile, "Name" + "," +
"Address" + "," + "City" + "," + _
   "State" + "," + "ZipCode" + "," +
"LicenseNo" + "," + _
   "Code" + "," + "Expires"
Start a loop to read the text file. In
actuality, it will loop only once.
Do While Not EOF(intReadFile)
Stuff the whole file into a variable
"strLine"
Line Input #intReadFile, strLine

Start a second loop to read smaller
chunks of the file
This program is made more
complicated by efforts to speed
up the processing. It obviates the need
to resave the whole file into memory
after parsing each field.
Do While Len(strLine) > 0
Initialize the counters
x = 1
ctr = ctr + 1
lngLengthOfFields = 0

Cut a chunk of 5000 bytes off the file
and manipulate it.
This makes the program run much
faster.
strTempLine = Mid(strLine, 1, 5000)

Feed the nine fields into an array,
using semi-colons as the delimiter
For x = 1 To 9
```

Continued on page twenty

A couple of interesting new sites to check out:

Guide to Statistical Resources: *www.mnsfld.edu/depts/lib/govstats.html*

The ten individual Web pages that make up the Guide to Statistical Resources list hundreds of print reference works, government documents, and other materials that provide numeric data. In many cases, online or CD-ROM versions are also available. Links to specific Web locations are given whenever possible.

StatBase (United Kingdom): *www.statistics.gov.uk/statbase/mainmenu.asp*

Many datasets may be downloaded. StatBase has been set up to provide access to a comprehensive set of key statistics drawn from the whole range of official United Kingdom government datasets.

NICAR data library: The Federal Aviation Administration's Enforcements database is now updated through Sept. 1999.

The database includes enforcement actions against airlines, pilots, cargo carriers, foreign airlines traveling in the U.S., mechanics and other aviation personnel.

More information, including sample slices of the data, is available on the NICAR Web site, *www.nicar.org/data/faae*

Cost is $50 for small news organizations, $70 for 50,000 to 100,000 circulation (25-50 market) and $90 for large news organizations.

To order, call the data library at (573) 884-7711.

Continued from page nineteen: **Parsing**

```
If PositionOfSemiColon(strTempLine) =
0 Then Exit Do
strField(x) = Mid(strTempLine, 1,
PositionOfSemiColon(strTempLine) - 1)
strTempLine =
TruncateLine(strTempLine)
lngLengthOfFields = lngLengthOfFields
+ Len(strField(x)) + 1
Next x

Cut off the start of the file already fed
into the array
strLine = LTrim(Mid(strLine,
lngLengthOfFields + 1))

Measure the length of the extraneous
data and cut it out of the next record
If ctr = 1 Then lngCheckName =
Len(strField(1))
If ctr > 1 Then
strField(1) = LTrim(Mid(strField(1),
lngCheckName))
lngCheckName = Len(strField(1))
End If

Write the data to a comma-delimited
file
Write #intWriteFile, LTrim(strField(1)),
LTrim(strField(2)), strField(3),
strField(4), strField(5), strField(6),
strField(7), strField(8)
Loop
Loop

Close the input and output files
Close
Notify the user when the program is
done
retval = MsgBox("All Done", vbOKOnly,
"Go Home")
End
End Sub

Create a function to locate semi-colons
Private Function
PositionOfSemiColon(String1 As String)
PositionOfSemiColon = InStr(1,
String1, ";", vbTextCompare)
End Function

Create a function to cut data from the
file once read
Private Function TruncateLine(String1
As String)
TruncateLine = Mid(String1,
PositionOfSemiColon(String1) + 1)
End Function
```

Solution:

A Visual Basic script, drafted by Tim Henderson, of *The Miami Herald*, and modified slightly by Gebeloff, creates a text file and nine variables.

It then creates a loop that goes through the data looking for field delimiters. The contents of the first nine fields are assigned to the variables, and then it stacks the next nine records underneath, and so on until it runs out of data. The result is theoretically a tab-delimited text file that has line breaks after each record.

Gebeloff discovered, however, that his software didn't recognize tabs, so he changed the script to insert a comma between fields in the new table.

He also had to swap in commas for semi-colons, which meant getting rid of extraneous commas in the original data.

The script (run from Word's V Editor):

```
Sub ParseDelimited()
variables for the two text files we need
Dim outfile As Integer
Dim infile As Integer

variables for the fields we create
Dim add As String, city As String, st As
String, zip As String, licno As
String, code As String, expires As
String, dob As String, name As String

Open the file to receive the parsed
fields
outfile = FreeFile
Open "c:\temp\hhc2.txt" For Output
As outfile
Open the original datafile
infile = FreeFile
Open "c:\temp\healthaides\hhc2.txt"
For Input As infile

Loop through the file inputting the
variables
Do While Not EOF(infile)
Input #infile, add, city, st, zip, licno,
code, expires, dob, name
Each time you get all 9 fields put them
in the new file separated by commas
Print #outfile, add & "," & city & "," &
st & "," & zip & "," & licno & ","
& code & "," & expires & "," & dob &
"," & name
Loop
Clean up
Close #infile
Close #outfile
End Sub
```

MaryJo Sylwester can be reached by e-mail at maryjo@nicar.org

Continued from page twenty: **Weighting**

ing data from a commercial vendor.

Experienced CAR practitioners approach commercial data with cautious skepticism.

After deciding to use census block group median household income as an indicator of students' economic status, we were faced with a choice: use the by-now-badly-outdated 1990 census data, or use updated estimates from a commercial data vendor.

After giving it a good bit of thought, and weighing comments offered up on NICAR-L, we opted to use 1999 estimates from Claritas.

Whenever the subject comes up on NICAR-L, the comments offered are thought-provoking.

The Miami Herald's Tim Henderson views commercial data with a jaundiced eye. Tim points out that annual estimates put out by commercial marketing data firms are just that: estimates. He further points out that the commercial firms, for competitive reasons, don't like to share a lot of information about how they arrive at their estimates.

Overall, we believe the data we're using is allowing us to measure what we set out to measure.

The fact that estimates are being drawn using vague methods should be enough to give sober journalists pause. But then, so should using 1990 census data to try to talk about the characteristics of our community in 1999.

On one point there is strong agreement among most who have weighed in on the topic on NICAR-L: it would be risky, indeed, to use annual estimates, whether from the U.S. Census Bureau's Current Population Survey or from commercial vendors, to try to measure changes over time at the local level – and more dangerous still to try to use such data to measure year-to-year changes at the tract level.

Using small-area data to compare different parts of your community within the same year is a different matter than using the data to try to measure changes over time (though the rationale for using 1999 Claritas estimates – the assumption that it is more up-to-date – presumes that the commercial source has made some sort of reasonably accurate assessment of changes over time).

Use with caution

A consensus emerges among many NICARians as to how to handle the use of commercial data (and these principles apply more broadly, to data other than commercial products): tell your readers/viewers as completely as possible what data you're using, how you're using it, and what its limitations are. Look for other sources of data to either corroborate or question the commercial data.

This is the approach we're following at *The Oklahoman* as we use this data. We've cross-referenced Claritas' block group household income data, aggregated to the school level, with school-level free and reduced-price lunch figures, and we've found a rough, if imperfect correlation. As the median income figure for a school goes down, the percentage of students receiving free and reduced-price lunches tends to go up.

In instances where the techniques we're using show obvious weakness, we're pointing out those weaknesses to our readers.

Overall, we believe the data we're using is allowing us to measure what we set out to measure. Even though it's imperfect, it's allowed us to take a deeper, more meaningful look at the story than we could without the data.

As Steve Doig, Knight Chair in Journalism at Arizona State University points out, before too long we'll have a solid benchmark against which to compare the data we're now using: the 2000 census.

"I think a really interesting story to be done shortly after the 2000 data is released will be a comparison of how closely the various vendors were to reality," Doig says.

Griff Palmer can be reached by e-mail at gpalmer@oklahoman.com

The MAUDE dataset is available from the NICAR data library, with records updated to March 1999, for the entire United States.

Cost is $100 for small news organizations, $125 for the 25-50 market and 50,000 to 100,000 circulation and $150 for large news organizations.

To order call (573) 884-7711.

A sample slice of MAUDE is NICAR Web site at *www.nicar.org/data/maude*

Coming soon...
The next issue of Uplink will be published in March. It will include stories about journalists outside the United States using CAR for projects and everyday assignments.

Continued from page one: **Religion**

erything from the annual Muslim fast of Ramadan to the plight of the Chinese meditation movement Falun Gong.

I also subscribe to e-mail lists from many of the major religious groups I cover. If you look around the major religious sites on the Web, you'll find places where you can sign up to receive almost daily e-mails from places as diverse as the Evangelical Lutheran Church in America, the National Council of the Churches of Christ, the Council on American-Islamic Relations, informal neo-pagan networks and even the supporters of imprisoned Falun Gong activists. Many of these e-mails are quickly dispatched to my "trash" folder each morning – but a quick scan of the dozens of "subject lines" I find in my inbox often sparks a future story idea.

CD-ROMs

I use a variety of religion-related CD-ROMs I've collected over the past five years. My CD-ROM library includes a tour of the Vatican, a guide to holy places in the Middle East, a scholarly Bible dictionary, a history of the Holocaust, a text of the Dead Sea Scrolls and a collection of Jewish reference books. These titles tend to be released by publishers and then quickly go out of print, so watch for the release of helpful CDs and snap them up as soon as they go on sale.

I use a variety of databases to quickly organize and search the gargantuan amounts of information that flow from the world of religion.

One excellent CD-ROM to consider is "On Common Ground: World Religions in America" by Columbia University Press, an in-depth look at many faiths by Harvard University's Pluralism Project. And, because writing about religion often involves references to history and traditional culture, it's very useful to keep a simple copy of an Encarta encyclopedia handy.

The most basic software for religion writers is a searchable Bible to quickly find key references that may crop up in stories on deadline. But remember that English translations vary widely and you'll probably want at least a King James Version as well as a more modern English version. I have four sets of Bible software on my PC and use the "New Oxford Annotated Bible" most frequently.

Databases

I use a variety of databases to quickly organize and search the gargantuan amounts of information that flow from the world of religion.

MS Access is excellent for building customized databases. For example, the Access database I've created on members of the College of Cardinals – the group from which the Catholic church selects its popes – lets me track this influential group season after season. (see sidebar on page 23)

I also use askSam-brand SurfSaver software on the Internet, which allows me to save key Web pages in categorized folders on my PC. Later, I can search the entire text of these voluminous holdings.

I use MS Word to create a crude but amazingly effective database out of reams of electronic data I've received about religion. I save important wire stories, significant e-mail press releases and other similar information by copying/pasting these texts into 10 ongoing Word files I have created. They represent categories, such as, "Catholic," "Politics and Religion" and "Islam." As an individual file reaches about 1.5 MB, I archive it onto a Zip disk and create a new file. When I'm reporting and want to pinpoint a particular story or document, I simply open up the appropriate Word file and use the Edit/Find function to search for key words.

E-mail

Because reporting on religion almost always involves an attempt to connect news with the lives of ordinary people, I use e-mail as a lifeline to keep in touch with people I've encountered over the years.

I also maintain specialized lists of people

Continued on page twenty-three

Continued from page twenty-two: **Religion**

by category related to my beat. For example, when the Catholic hierarchy is considering a change in church rules that might affect local parishes, I will fire off a quick e-mail to my list of Catholics and ask whether they know of anyone especially concerned about such a change. People who e-mail me with an intelligent comment – readers, academics, clergy and others – are potential future sources. I save their e-mail addresses and may include them in the lists I'm building.

Ultimately, covering religion at the turn of the millennium is like covering presidential politics – if there were 1,000 parties fielding candidates. The subject can be overwhelming, unless journalists use all the available RAM at their disposal.

David Crumm can be reached by e-mail at crumm@freepress.com

ON THE BEAT:

Predicting next Pope

By David Crumm
Detroit Free Press

Since the current pontiff, Pope John Paul II, is 79 and in poor health, many Catholics are concerned about the identity of his possible successors. What kind of man will emerge as the next worldwide leader of the Catholic Church?

That's what the Free Press hoped to explore in an in-depth examination of leading papal candidates.

The problem is that, if John Paul died tomorrow, there would be 107 qualified candidates: all of the current cardinals under age 80.

Beyond interviewing Vatican-watchers, church scholars and our local cardinal, we wanted to offer our own analysis of the current College of Cardinals. The first question was the most basic: exactly how many cardinals were under age 80 and thus were eligible candidates?

Beyond that, we wanted to know more about the background of the cardinals. For instance, could another non-Italian be elected to follow John Paul? How heavily were various countries and continents represented in the current group?

To resolve these questions – and to be able to accurately track the exact makeup of the College of Cardinals in future stories – I created an MS Access 97 database of all current cardinals. To provide a comparison with the makeup of the college that elected John Paul II, I also built a database of all cardinals involved in the 1978 conclave.

I collected data on more than 200 cardinals: their names, birthdates, countries of origin, current assignments, etc. Building the database took several full days of painstaking work, because the information was available only in a printed format – and had to be extracted from a series of annual almanacs. In this case, I could not directly import electronic data.

I wound up with more than a dozen categories of data about each man – including a category to identify those cardinals in our database who had played a role in the church since the 1978 conclave, but who had died.

Immediately, we were able to identify 107 men as eligible for election – as of our publication date. Looking deeper into the data, we found that, since John Paul's election, he has thinned the ranks of Italian cardinals from 23 to 17 – and has spread a few of those votes to new cardinals from Africa and the United States.

We also determined that John Paul has personally appointed nearly 90 percent of the current College of Cardinals, an indication that most of the candidates tend to agree with the current pope on major issues.

The database was the key to writing our story with authority and specificity about the numbers – and, overall, to pushing our analysis of the candidates to a higher level.

The valuable bonus in this case is that, as we continue to follow the issue of papal transition, we can quickly check in on the accurate list of candidates at any point in the future.

David Crumm can be reached by e-mail at crumm@freepress.com

Here are just a few religion Web sites from Shawn McIntosh's beat page (*www.reporter.org/beat/religion.html*):

Christian Coalition: *www.cc.org/*

Covenant Baptist Church – Texas: *http://CovenantBaptist.org/*

Religion News Links (Zondervan Publishing House): *www.zondervan.com/newslink.htm*

Nation of Islam Online: *www.noi.org/*

Internet for Christians: *www.gospelcom.net/ifc/*

The Islamic Herald: *www.ais.org/~bsb/Herald/links.html*

Bits, Bytes and Barks

Uplink survey

Thank you to everyone who took the time to answer our survey on the NICAR Web site during the fall. We've compiled the results and will use them to provide content that better suits your needs.

Overall, we found *Uplink* has a wide range of readers in terms of skill levels. At least from the survey (which, of course, isn't statistically valid), our readership appears to consist of an equal mix of those who consider themselves beginner, intermediate and advanced.

The biggest finding of the survey was that readers of all skill levels want more technical stories – with specific "how-to" examples. This is something we're going to address immediately, and I hope you'll see gradual improvements in this area in the coming issues. And keep in mind that you can help by offering to share your solution to a tough technical problem. Eventually, somebody else may offer a solution that would help you as well.

We're also working to incorporate more information pertinent to small news organizations – since doing CAR with limited resources is very different.

As always, we encourage you to share the results of your work, no matter how large of a news organization you work at or whether it's print or broadcast. Your examples and tales (of both success and failure) can help others learn.

Contact me at maryjo@nicar.org or by phone at (573) 884-7711 if you have story ideas or questions/comments about the content of *Uplink*.

–MaryJo Sylwester

NICAR data library

The boat registration dataset has been updated through June 30, 1999. This database, the U.S. Coast Guard's Marine Safety Information System, contains information on commercial and recreational vessels.

It can be used mainly to find out information about a specific boat or ship, such as finding the owner. Reporters have used this dataset when a shipping accident occurs or to find out information about U.S. Coast Guard boats.

There is documentation for 265,584 vessels registered with the Coast Guard, including the company name and address. A second table includes information on boats with expired registration, foreign vessels and recreational boats.

The dataset is available for the entire United States at a cost of $60 for small news organizations, $80 for those between 50,000 and 100,000 circulation (or 25-50 market) and $100 for large news organizations.

More information is available on the NICAR Web site, *www.nicar.org/data*, or by calling the data library at (573) 884-7711.

Census Workshops

IRE and NICAR will be holding Census workshops both before and after the data is released. We're working with Steve Doig, of Arizona State University, and Paul Overberg, of *USA Today*, to provide hands-on training on working with the Census 2000 data. Dates and locations haven't been set yet.

Stay tuned here or at the IRE Web site, *www.ire.org*, for more information.

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

NON-PROFIT ORG.
U.S. POSTAGE
PAID
PERMIT NO.286
COLUMBIA, MO. 65211