Uplink

May/June
2001

TRANSPORTATION

Firestone and Explorers

By Josh Barbanel
The New York Times

It was late in the day. We were up to our ears in other projects when a business editor interrupted with one of those vexing requests. It was a nuisance - hardly worth the attention of one, not to mention two, highly significant reporters in the new Investigative Cluster of *The New York Times.*

"Can you help us open up this spreadsheet," the editor asked.

Sigh.

The spreadsheet turned out to be a messy table with the first release of government complaints about deaths attributed to Firestone tires on Ford Explorers. We clicked on the table icon in the email, counted the cases and were done. Back to our very important work.

But the next day, as my colleague Ford Fessenden likes to say, serendipity happened. We were discussing what we would have said, if the editors had seen us as something more than keyboard jockeys – and actually asked for help in thinking about and dealing with the rapidly developing story of the recall of more than 14 million Firestone ATX and Wilderness tires.

Well, we both had worked with government accident data files. We knew that Ford Motor Company was using warranty claims data to monitor the problem (and make its case that the problem was all the fault of Bridgestone/Firestone). We knew the federal government was relying on complaints called in by survivors, lawyers and the auto companies.

If these complaints involved real people killed in real fatal accidents, shouldn't we be able to track at least some of the same cases in the rich federal database of fatal accidents?

Hitting on Explorers

The editors never actually assigned us to look into the Ford Explorers. But we poked around, and they were fascinated when we told them we were able to clearly trace a significant surge in tire-related fatal accidents in Ford Explorers, not found in other sport utility vehicles.

This was remarkable (some may say debatable) because reports of tire-related deaths were rare events. At the time, the federal government had complaints of 88 deaths in tire-related accidents in the Ford Explorer, including some going back many years.

Continued on page fourteen

CENSUS 2000

Heavy lifting ahead

By Richard O'Reilly
Los Angeles Times

The first wave of data from Census 2000 has washed ashore, providing some with an exhilarating ride surfing the leading edge of computer-assisted reporting. It may have been more terrifying than thrilling for others, however.

Whichever situation you identify with, upcoming data releases will be much, much more challenging.

The data dictionary for PL94-171's four tables of race and race and Latino/Hispanic ethnicity was 10 pages long. For Summary File 1 coming this summer, the preliminary description runs 110 pages, and the final detailed listing will be many times that length.

PL94-171 had a total of 288 population count variables, plus the 80 geography variables.

Summary File 1 will have the same 80 geography variables plus 8,132 population and housing count variables. That includes 3,078 variables down to the block level plus another

Continued on page twenty-six

Inside Uplink

This issue of *Uplink* covers a lot of ground on the transportation front. Tony Manolatos of *FLORIDA TODAY* details the paper's investigation of fatal accidents on Interstate 95 (see page two). Andy Lehren of *Datline NBC* describes the effort to spotlight trucks avoiding inspections (see page four). In a First Venture, Dan Feldstein of the *Houston Chronicle* talks about his story on city-bus drivers and their overtime habits (see page six).

Health Care

An inside look at *The Seattle Times* series on conflicts of interest in medical research and deadly results.

See page twenty

Census

In addition to the look-ahead piece by Richard O'Reilly of the *Los Angeles Times* (see this page), Robert Gebeloff of *The Star-Ledger* (Newark, N.J.) talks about how preparation helped with the initial release.

See page twenty-three

Uplink

May/June 2001

Volume 13, Number 3
A newsletter of the National Institute for Computer-Assisted Reporting

EDITOR
Brant Houston

MANAGING EDITOR
Mary Jo Sylwester

ASSOCIATE EDITORS
Amy Sherrill
Mike Sherry

ART DIRECTOR
Yan Jin

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

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

DIRECTOR OF PUBLICATIONS
Len Bruzzese

SUBSCRIPTION ADMINISTRATOR
John Green

Uplink is published bimonthly 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

TRANSPORTATION

Fatal highway

By Tony Manolatos
FLORIDA TODAY

Fatal-crash stories along Interstate 95 were almost becoming routine. During a six-month period last year, 12 people were killed in 11 separate accidents on Brevard County's 73-mile stretch of Interstate 95.

One New York family, who lost their parents in a March crash, wrote FLORIDA TODAY and pleaded with reporters and editors to take the coverage beyond spot-news stories. They wanted to know the reasons why their parents and many others were dying on I-95. They wanted the newspaper to illustrate the human toll.

We, too, saw the potential of the story. But first we had to determine just how dangerous the interstate, one of two main arteries in Florida, really is.

A team of reporters and editors began culling and analyzing state data from nearly 30,000 crashes over five years along Florida's 382 miles of I-95. In an initial 10-page report on Dec. 3, 2000, we delivered the news that the number of people killed on the Florida highway had doubled over five years.

Gathering data for I-95 required working with four state transportation districts and compiling data from each district into one database.

The data were startling enough, but the process also was revealing and newsworthy. We learned that money was the biggest culprit impeding enhancements to the road. But we also learned that while state transportation officials collect crash data, an aging computer system does not allow those officials to thoroughly analyze the information to identify the causes of crashes.

In the 10-page report (called "I-95, More Dangerous Than Ever") and the dozens of stories, editorials, reader letters and a public forum that followed, enough urgency was raised to compel state and local officials to take action. It was impossible to ignore the findings of the report, including:

- In 1996, there was a sudden and dramatic increase in the number of fatal crashes on I-95. That was the year the speed limit was raised to 70 mph from 65 mph.
- The death rate per I-95 mile driven in Brevard County, home to FLORIDA TODAY, is more than twice the national average for interstate highways.
- Almost a third of the fatal crashes in Brevard County occur when vehicles veer out of control across grass medians and into oncoming traffic. Those "crossovers" lead to more multiple-death crashes than any other type of accident. They could be prevented by erecting median barriers in high-accident locations – a fairly simple, low-cost solution.

Comparisons

One thing the project did not do was compare I-95 to another road. We did gather fatality rates for U.S. interstates and Florida interstates and we compared those to I-95. But we did not, for example, look specifically at Interstate 75 in Florida.

We chose to compare I-95 to itself because it is Florida's eastern gateway and the most heavily traveled road in the state. However, taking the time to compare it to I-75, Florida's other major highway, would have added depth and perspective to our report.

Gathering data for I-95 required working with four state transportation districts and compiling data from each district into one database. Because each district is independent of the other, no one at the state level can effortlessly answer basic questions – such as the number of people killed in any given year – about I-95, I-75 or any other road that runs through multiple transportation districts.

State transportation officials said a new central computer system, expected to be in place by July, would give them a better and quicker understanding of state-maintained roads.

Continued on page three

Continued from page two:

Fatal highway

While the National Highway Traffic Safety Administration tracks all fatal wrecks in its Fatal Accident Reporting System database, the agency essentially relies on police reports for the information. Those reports often list multiple names of one road and NHTSA's database does not take that into account. In other words, you would have to know every possible name that law enforcement might use for a road in order to properly crunch the numbers.

Analysis

Anthony DeBarros, a database editor for *USA TODAY*, cleaned up the data, compiled it into one database and assisted with the number crunching. We used Microsoft Access. The database was 13 megabytes. We also used Microsoft Excel for some import work. The data was supplied to us at no cost.

State transportation officials said a new central computer system, expected to be in place by July, would give them a better and quicker understanding of state-maintained roads.

"The data was in fixed-width text files, but we discovered that one of the state zones was missing a couple of years of data, so we had to get that data and add it to our database" DeBarros said.

"In terms of cleaning, all the data was presented in code, and I spent considerable hours going through and converting those codes into meaningful text that the reporter could understand. Also, we had to figure out how to piece together (separate data from four districts) so we could get a picture from end to end."

The CAR work took about two months to complete as a side job. The project took about four months to complete.

DeBarros mentioned that while we did compile traffic data for I-95, using it to determine fatality rates, reporters who do similar projects should keep in mind that traffic volume must be figured into the equation before a stretch of road can be labeled the "deadliest."

Reaction

Reaction to the coverage has been swift and substantive. Here's what has resulted so far:

- Brevard County legislators pledged that I-95 safety would be their top priority during the 2001 legislative session. They backed that pledge by setting a meeting in January with the state secretary of transportation to push for median barriers along I-95 in Brevard County.
- The transportation secretary hired a traffic engineering company to analyze Brevard County's 73-mile stretch of I-95. The company is expected to examine 10 years of crash records.
- The Metropolitan Planning Organization, the chief transportation planning agency in Brevard, said it would use our report to build a public awareness campaign to improve safety on Brevard's stretch of I-95.
- In South Florida, Broward County officials initiated a public awareness campaign to alert drivers to the dangers of I-95 in that county. Transportation officials in Broward are continuing to study I-95 crash records in an area the newspaper identified as the stretch with the most fatal wrecks and the most traffic.
- And state highway patrol officials are using the newspaper's finding to lobby vigorously for more high-visibility enforcement along the entire interstate, including more marked and unmarked cars, increased DUI enforcement and aircraft speed-enforcement details.

Tony Manolatos can be reached by e-mail at tmanolat@flatoday.net

THE FLORIDA TODAY story, "I-95: More Dangerous Than Ever," is available from the IRE Resource Center, story #17007. To order, call (573) 882-3364.

Data: Accident reports from four Florida transportation districts

Software: Microsoft Access and Excel

Data used for "Truck Stop":

- U.S. DOT Truck Census (available from NICAR)
- U.S. DOT Truck Accidents (available from NICAR)
- U.S. DOT Fatality Analysis Reporting System (available from NICAR)
- SafeStat, U.S. Office of Motor Carriers
- State-level transportation data

More information on the datasets available from NICAR is available at *www.ire.org/datalibrary/databases* or by calling (573) 884-7711.

Software:

Visual FoxPro 5.0 (process and analyze databases)

ArcView 3.0a (analyze geographic patterns)

SPSS 8.0 (statistical analysis)

Excel 97 (snapshot summaries)

TRANSPORTATION

Truck safety

By Andy Lehren
Dateline NBC

During a lull at a truck inspection station in rural Florida, I took a drive a few miles up the road. It wouldn't take much guesswork to figure out what I saw. Trucks. Lots of trucks. Some pulled over the side of the road. Others getting off at exits that had little discernible lure for commercial traffic.

They were doing whatever they could to avoid inspections. They were driving down tiny main streets, past scattered homes. They trundled by schools and farms. You could hear them speaking on CB radio about the best way to navigate the back roads so they could pop back on the interstate, past the inspectors. Some would play possum by pulling into a parking lot, waiting for the inspection station to close.

The reason we undertook such a thorough approach was to overcome some of the incompleteness and unreliability that happen if you relied solely on one of these databases.

Welcome to one of the things you won't see when you look at truck inspection data. When trucks are inspected, one of five is pulled out of service. A lot of the time it's for brakes that won't work right when there's an emergency; stuff that leads to accidents. When you see trucks pulling off the road to avoid inspections, you know there are a lot of inspections that never happen. That one-out-of-five figure is only a baseline.

Dateline NBC obtained more than a dozen databases to examine the problems with trucking safety. These included federal and state databases on accidents, inspections, truck carrier profiles, safety ratings and enforcement actions. The sources included the U.S. Office of Motor Carriers (now called the Federal Motor Carrier Safety Administration), the National Transportation Safety Board, the U.S. Bureau of Transportation Statistics, and transportation and public safety agencies in several states, including Florida, Missouri, Texas and California.

The reason we undertook such a thorough approach was to overcome some of the incompleteness and unreliability that happen if you relied solely on one of these databases. The result allowed the news magazine to more fully document the problems in trucking safety, in a story produced by Lynn Keller with correspondent Steve Daniels. It documented how many trucking companies systematically ignored safety rules and put bad trucks on the road, and how regulators fall short monitoring them.

Studies and reports

A starting point for our work was collecting various reports and studies that had been done on trucking safety, to establish sound methodologies. It also helped us gauge how well different states handle truck safety. California, for instance, is much better than Arkansas.

The simplest analysis was tracking which companies had unusually high out of service rates when their trucks were inspected. This is where having state data becomes important. Federal truck inspection data is only required to include interstate carriers. One in ten trucks on the road are intrastate truckers. They aren't licensed to go into other states. So their inspection records are not included in federal databases – even if they are part of a subsidiary that's an interstate carrier. These trucks tend to face less scrutiny. It's also one of the things you won't see when you check a company's history off the U.S. Department of Transportation's web site (*www.safersys.org*).

Some states offer far more details in their inspection and accident data. Those additional columns of data may tell you more about the driver or more easily identify a particular truck. The federal Fatality Analysis Reporting System is also good for double-checking truck accident data.

Continued on page five

Continued from page four:

Truck safety

Safety scores

Several of these databases were obtained under the Freedom of Information Act and state open records laws. The most significant victory was a FOIA request that *Dateline NBC* won, on appeal, with the U.S. Office of Motor Carriers. We were the first news organization to obtain a recently developed database known as SafeStat, short for the Motor Carrer Safety Status Measurement System. It's worth getting, and should be easy to obtain for any news organization seeking it these days.

SafeStat includes numerical grades on the safety of trucking companies. The scores are on a sliding scale, and allow you to compare different companies with scores for four areas: driver, vehicle, accident and safety management.

Dateline NBC used the FOIA to also obtain the safety classifications kept by federal regulators in a separate, older system, as part of the federal motor carrier census. These are three letter grades – satisfactory, conditional and unsatisfactory.

Just like the statistics books tell you, there's a lot of subtlety that's lost when data is categorical. Out of service rates, accident rates and SafeStat scores give you a much better sense of how much safer one trucking company is versus another. That sense of proportion is lost when information is telescoped into three grades.

But the older three-grade system is still worth a look. The grades are one more way to cross-check the different rates. It also gives you a sense of a company's history. Very few are unsatisfactory. But experts say that in this somewhat lenient system, keep your eye on trends. If a company is unable to maintain a satisfactory grade consistently over the years – say, it goes from unsatisfactory to satisfactory to conditional and then again conditional – it's worth checking out.

One way we used SPSS was to run a linear regression comparing the size of truck companies versus their serious accident rate, then comparing companies against like competitors. Reporters should look for outliers - which companies have higher than expected accident rates. (There is much in the literature on using numbers of trucks versus miles traveled, generally a less reliable number. Reporters should be aware of these factors, run the data both ways, and compare results).

Mapping

We used mapping software several ways. Perhaps the most interesting was plotting accident and inspection data to figure out which inspection stations we wanted to visit, to see how regulators handled their workload and dealt with companies that had a pattern of problems. That meant we were there when a military munitions truck was pulled out of service for not being safe to drive.

Mapping also helped us analyze whether a trucking company was inspected more in certain states, and whether those states had lax or stringent inspections.

Mapping also helped us analyze whether a trucking company was inspected more in certain states, and whether those states had lax or stringent inspections. One large trucking carrier, for instance, scored well in its home state, and usually scored much worse elsewhere in the country.

Having all this data allowed us to challenge trucking industry claims about safety. We knew which companies' paper records for on-site inspections we wanted to obtain. Enforcement data helped underscore the contrast between the problems on the roadways and the declining actions taken by regulators.

The data also helped us make sure we had a strong reason for highlighting the different companies and accidents that were profiled in the story.

Andy Lehren can be reached by e-mail at alehren@nbc.com

The *Dateline NBC* story, "Truck Stop," is available on videocassette, story #16152.

Other truck safety stories using CAR:

Story #12140. "Fatal Crashes with Big Trucks are on the rise," *The Morning Call* (Allentown, Pa.), 1995.

Story #16003. "Unsafe Passage," *The Times-Picayune* (New Orleans), 1999. Addresses oversight in both truck and bus industries.

Story #12584. "Asleep at the wheel," *Press Herald* (Portland, Maine), 1995. Examines car accidents **involving commercial trucks.**

The Houston Chronicle story will be available soon in the IRE Resource Center, (573)882-3364.

Data: Time, pay and accident records. The data included nearly a million pay records for the drivers.

From: The Houston Metropolitan Transit Authority.

Software: Excel, Access

The U.S. Department of Transportation's Fatality Analysis Reporting System and its Motor Carrier Management Information System include data on buses. Both databases are available from the NICAR Database Library. To order, call (573) 884-7711.

FIRST VENTURE

Bus driver OT

By Dan Feldstein
Houston Chronicle

William "Bones" Pierson is a legend among bus drivers at Houston's Metropolitan Transit Authority.

He routinely drives 19-hour days and seven-day weeks. He didn't take a day off between Aug. 21 and Nov. 4.

He earned $109,000 last year, more than tripling his base salary through overtime. Dozens of other drivers were close behind.

I hadn't known about Bones or the others until the computer tipped me off.

Fresh out of NICAR's January boot camp, I wanted something simple and comfortable for my first venture.

I ranked them in descending order and found my first real problem: The top guy drove 26 hours in one day. Good trick.

I covered transportation before I went to the *Chronicle*'s projects desk, so I picked on Metro – the Metropolitan Transit Authority. I'd heard that some of its drivers were racking up big bucks driving overtime.

I had four questions: How much were they earning? How much were they driving? Was this legal? Was this safe?

I requested the record layout of any database that contained time records, pay records or accident records. Neatly, they had all three. Each was separate, but each had a field for driver ID and a field for date. I requested them all for the calendar year 2000, plus a week on either end.

They came in Excel worksheets and I carefully put them into Access. The pay was trickier than I had expected. There were nearly a million records for only about 1,500 drivers because each type of pay was listed in a different record. On any one day, a driver could have one record for regular pay, one for overtime pay, one for a monthly bonus, etc.

Some simple SQL gave me a list of the top earners.

List of adjustments

The time records were next. They came by driver, by day. These were "platform minutes," the actual time the driver was on the street, behind the wheel. I ranked them in descending order and found my first real problem: The top guy drove 26 hours in one day. Good trick.

Oh yeah, said the Metro computer guy. You need the "list of adjustments." Every bus route has a set time. But if you get stuck in traffic, time can be tacked on. And if you get ill and leave early, time can be deducted. Plus other errors are corrected.

It turned out there were roughly 50,000 adjustments for about 350,000 records, which is when I learned about "outer joins." Using driver ID and date, I tacked the adjustments onto my time table in a new field. Then I summed that field with the original time field in a query to create a new master table.

It turns out the top shift was 21 hours behind the wheel, not 26 – so much more reasonable.

Checking the law

Was all this driving legal? The feds allow interstate commercial drivers no more than 10 hours behind the wheel each day, or 60 hours in seven days. Texas allows intrastate drivers 12 hours a day or 70 in seven days. But both governments exempt government drivers.

Still, 16 percent of all shifts driven by Metro last year would have violated the federal safety standard, and 5 percent would have violated the state standard.

For the single-day standard, it was easy to calculate. How many records were over 10 hours, over 12 hours?

The seven-day rolling total was wicked, however. It didn't exist, and I would have to create it. On every given day, how many hours had the driver been driving in the previous seven?

Continued on page seven

Continued from page six:

Bus driver OT

Help from NICAR-L

I turned to the NICAR listserv. Three people offered suggestions. Tim Henderson of *The Miami Herald* offered this:

```
SELECT a.driver_id, a.work_date, sum(b.minutes) as rollingminutes
FROM Platform as a, Platform as b
WHERE ((([b].[driver_ID])=[a].[driver_ID]) and ((b.work_date) between ([a].[work_date]-7) and ([a].[work_date]-1)))
GROUP BY a.driver_id, a.work_date
```

It worked! I also tweaked it slightly to get the last six days plus the current day.

Finally came the question of whether the excessive driving was more dangerous than regular shifts. Experts said yes, but was there statistical evidence?

Actually, it didn't show up in my crude analysis. I joined the time and accident tables, using driver ID and date. I found that, generally, the percentage of accidents attributed to drivers who had exceeded drive-time standards was the same as the percentage of shifts that were driven in excess of those standards.

Jumping into a project immediately after boot camp really solidified SQL in my mind. It would have escaped otherwise.

And frankly, it took me as long to refine that statement into a single sentence as it did to make the calculations.

Lessons learned

This story could be replicated in a lot of cities. Transit systems have a tough time making new hires because of drug tests and a good economy. They are as happy with overtime as the drivers are.

I learned several lessons in this project. While the data looked nice and neat, there were certainly some individual entries that had errors. Some of the accidents simply didn't join with the time table, for instance. The buses weren't being driven by phantom drivers -- people had entered wrong dates, wrong driver IDs, etc., into the database.

I kept my analysis broad because of that. And when I generated a list of the top 10 longest shifts that I was going to print, I gave it to Metro to check. Sure enough, they found a couple that were in error – a guy got sick, someone else hopped on his bus and their times were merged, etc. But most were correct.

SQL solidified

Other important things: Jumping into a project immediately after boot camp really solidified SQL in my mind. It would have escaped otherwise.

Also, what little I know about databases puts me one step ahead of most flaks. I made my best progress by suggesting a conference call among the flak, Metro's computer guy and me. It helped enormously for the flak to hear the computer guy say, "Oh that's easy." It took two or three weeks to get the data, free of charge.

"Bones" owns a nightclub and a Lexus that he doesn't get to drive much. Sadly, his wife told him to choose between marriage and driving so much overtime.

The new bachelor still has his pedal to the metal. And his picture was on the front page as the reigning overtime king of Metro

The story was the talk of bus riders for few days, but not much else has happened. The transit authority has changed no policies, saying it doesn't see a safety problem.

The authority did embark on a major advertising campaign to attract more drivers.

Endnote: My second CAR project, just completed, was with racial-profiling data at the Houston Police Department. They announced grandly in August 1999 that they were going to track all officer-initiated stops whether a ticket was issued or not. The story ended up being that the cops weren't bothering to fill out their profiling form.

Dan Feldstein can be reached by e-mail at dan.feldstein@chron.com

Other stories related to municipal buses are available from the IRE Resource Center, (573) 882-3364.

Story # 17402. WMAQ-TV (Chicago) conducted a yearlong investigation into the practices of Chicago Transit Authority (CTA) bus drivers. It uncovered 127 CTA bus drivers with drunk driving convictions and 56 drivers operating city buses with suspended or revoked licenses. WMAQ also learned that the CTA repeatedly failed to conduct background checks on its drivers, despite having a direct link to the Illinois state drivers license records.

Internet sites dealing with transportation data:

The National Transit Library Web: *www.fta.dot.gov/ntl/index.html*

The Intelligent Transportation Systems (ITS): *www.itsa.org*

U.S. General Accounting Office: *www.gao.gov*
The annual indexes contain a section for all transportation reports done in that year.

TRANSPORTATION

Water, roads and more

By Devon Hammes
IRE and NICAR

If you are seeking out a topic for an in-depth investigative piece, or if you just need that nugget of information to follow up on the most recent plane crash, transportation data has you covered.

Story ideas abound in data such as hazardous materials, boating accidents, truck accidents or any of numerous aviation datasets.

Readers of the *Fort Worth Star-Telegram* learned where some of the most dangerous routes in Texas were thanks to one reporter's use of data.

Jeff Claassen used Microsoft FoxPro and Excel to examine the dangers of driving on Texas highways and interstates. The paper specifically wanted to find where drivers in north Texas and statewide were most likely to encounter a bad accident.

With help from transportation experts, Claassen devised a method to use average daily traffic figures to calculate a rate of serious wrecks per volume of traffic.

One thing to keep in mind when working on any story that relies on statistics is that it is necessary to normalize the data. To normalize his data, Claassen took into consideration the number of vehicles travelling on a specific roadway.

"Other newspapers had analyzed Texas accident data, but focused on the total number of fatalities or fatal wrecks for each segment of a freeway or highway, called a 'control section,'" Claassen said. "That seemed incomplete because the fact that one roadway had more fatalities or fatal wrecks than another could just be the result of having more traffic. Also, the control sections varied in length from as short as a mile to more than 30 miles, making them difficult to compare to each other."

With help from transportation experts, Claasen devised a method to use average daily traffic figures to calculate a rate of serious wrecks per volume of traffic. He also broadened the story by including serious injury accidents.

The paper used accident report data from the Department of Public Safety, which consisted of about 1.9 million records from 1993 to October 1999 (the most recent available). The *Fort Worth Star-Telegram* also obtained the Road Inventory file from the Texas Department of Transportation, which provided average daily traffic figures for the roadway "control sections."

But this is where some problems began.

"Our first roadblock was incomplete data. The DPS data notes the location of some accidents using 'control sections,' and others using city accident-location codes. That omitted important places like Houston and San Antonio from the mix, so we requested more data to fill in the holes," he said.

Claassen then focused the investigation by analyzing wrecks from January 1997 to October 1997 to give the study more accuracy. Transportation analysts told him that time frame was an accurate look at the average number of accidents. State officials warned Claassen that the distance of control sections sometimes change. So, with help from the transportation institute, he was able to develop a method to calculate the rate of serious injury and fatal wrecks.

Cable cars

Even if a story doesn't start from the data, transportation stories can be found anywhere in the city. For example, Dan Noyes of KGO-TV in San Francisco took an in-depth look at cable car safety. The story found that the cable cars are not as safe as many people might think.

"I decided to look into the cable car story after seeing several accidents over the months," Noyes said. "I wondered how this piece of San Francisco history compared to

Continued on page nine

Continued from page eight:

Water, roads

other forms of transportation."

Using accident reports from the California Highway Patrol, Noyes was able to put together a spreadsheet in Excel. With that spreadsheet, he analyzed how many accidents occur per million passenger miles traveled for the different modes of transportation -- car, light rail, bus and cable car. The biggest problem KGO-TV had was getting authorities to realize that his station had a right to the information.

Boating accidents

Karl Ebert of *The Northwestern* in Oshkosh, Wis., used the federal Recreational Boating Accidents Database to look at the number of boating accidents on Wisconsin waters. He found several instances of accidents that were attributed to alcohol and the boater's inexperience in maneuvering the vessel.

"People up here love three things: booze, boats and speed. Each summer we report stories about boat accidents, most of them brought on by bad behavior of one sort or another, and decided that we needed to look into the issue to see just how big a problem there actually is out on the water," he said.

The data did help give the story more accuracy, but Ebert ran into a few problems along the way.

"The biggest problems were the inability and in some cases unwillingness of county officials and local wardens to provide information about tickets that had been written on the water. Because of time and publication scheduling constraints we decided to go with more anecdotal information. That was frustrating, but I don't think it hurt the main thrust of the story," Ebert said.

There were also instances in which he found some problems in the data itself. There was a change in the data's format from several years ago, and because of that, not all of the fields matched up. He also found that a field indicating alcohol involvement was unreliable in historical data because it hadn't been consistently reported.

Devon Hammes can be reached by e-mail at devon@ire.org

The following stories are available from the IRE Resource Center, (573) 882-3364:

The Fort Worth Star-Telegram's "Deadly Driving," story # 16809.

KGO-TV's cable car broadcast, story # 15295.

The Northwestern's "In Search of Safer Waters," story # 16952.

TRANSPORTATION

Data galore

NICAR offers numerous helpful transportation datasets to generate story ideas, add depth and authority to daily stories, and to develop enterprise stories.

Prices are based on the size of your news organization. Journalism professors and students fall into the smallest category.

More information about these datasets, including sample slices and record layouts, is available at *www.ire.org/datalibrary/datasets*

Air

• FAA Enforcements: Actions against airlines, pilots, mechanics and their designees, including all types of planes. Data are available from 1962 through August 2000 for $25, $50 or $75, depending on circulation or market size of your news organization.

• FAA Service Difficulty Reports: Contains information on mechanical difficulties collected by the FAA for the purpose of tracking repair problems with commercial, private and military aircraft and aircraft components. Reporting is voluntary, so the database is considered incomplete. Data is available from 1974 through December 2000 for $65, $130 or $155.

• FAA Accidents and Incidents: Contains information mainly about U.S. flights where there was an accident or an incident, such as crashes, collisions, deaths, injuries, major mechanical problems or costly damage. Data is available from 1973 through about April 2000. Cost for the most recent data, 1990 through 2000 is $40, $80 or $105.

• FAA Aircraft Registry: Contains a listing of all aircraft and aircraft owners registered in the United States. This dataset also includes tables on registered aircraft dealers and individuals or companies that reserved

Continued on page nineteen

Outside the gates

By Brian D. Crecente
Rocky Mountain News

Editor's Note: Crecente was a reporter with the Palm Beach Post when he wrote the Atlantis story.

A year of investigation, database entry and query building boiled down to just two numbers: 1,067 and 99.3.

But those numbers made the story.

The Palm Beach Post found that in 1997 and 1998, the tiny gated city of Atlantis quietly took control of three major roads framing its boundaries and began churning out tickets to nearly everyone but its own residents.

In one year's time, 99.3 percent of all tickets written by city police went to out-of-town drivers.

In the past five years, Atlantis fine collections jumped from about $5,400 to more than $63,000 – a 1,067 percent increase.

I knew we were on to something, but I had no idea how dramatic the numbers were going to be until I ran my first query.

Despite Florida's wonderful sunshine laws, getting to those numbers wasn't as easy as it could have been. The story began as we followed a tip from a disgruntled ex-Atlantis cop who alleged tickets were being issued on racial lines. We later found that not to be the case.

My colleague, Matt Mossman, made dozens of requests for traffic ticket data from the city, but officials said tickets weren't tracked and that there weren't any electronic records. Finally we went to the source, the Atlantis police chief, and requested copies of all traffic citations and warnings issued by the department from July 1998 to June 1999. The result of the request came in four large boxes, 1,506 tickets sorted by date and several hundred warnings.

Construction

Database editor Christine Stapleton constructed a database in Access which included fields for everything from the officer's badge number to the specifics of the citation and when and where it was issued.

Mossman spent the next few months entering the data and then as his data entry drew to a close, left the paper.

I was left with the enviable task of sorting through a well-organized database, filled with the sort of numbers investigative reporters dream about. There were about 2,000 records between Access and Excel tables.

I knew we were on to something, but I had no idea how dramatic the numbers were going to be until I ran my first query.

Digging in

After a little fiddling around, I typed a query that would show how many of the tickets issued were to people whose driver's license showed them living anywhere but Atlantis.

Of the 1,506 tickets analyzed, only 11 went to residents. The other 99.3 percent went to non-residents.

We had our story.

The database also showed that the city's police seemed to spend a majority of their time outside the walls of their city, with all but 60 of the tickets written being for traffic stops on the three main roads framing the city's walls.

Atlantis residents were also more likely to receive warnings than tickets when pulled over. About 78 percent of the Atlantis residents who where pulled over received warnings, while only 16 percent of the out-of-towners pulled over received warnings.

It was important not to stop with just the data we collected on speeding tickets. I also needed to understand how the city was able to annex the three roads that framed the city's walls.

After nearly a month of talking to city, county and state officials, I discovered that the county and the state either didn't offi-

Continued on page eleven

The Palm Beach Post story, "Outside the gates," is available through the IRE Resource Center, at (573) 882-3364. Story # 17699.

Other traffic-ticketing stories in the Resource Center include:

Story # 12963. A three-part series in *The Intelligencer Record* (Doylestown, Pa.) reveals a subjective ticketing system, including judges and police officers.

Story # 15458. *The Atlanta Journal-Constitution* investigates a community that collects about 57 percent of its revenue from traffic fines.

Story # 13685. *The Columbus Dispatch* (Columbus, Ohio) finds **that Ohio financially depends on the misfortunes of traffic violators. In the name of safety, traffic tickets have become a billion-dollar business resulting in selective enforcement, punitive fines and jailing of the poor.**

Continued from page ten:

Gates

cially allow the annexations or didn't fully understand what the annexations meant.

According to city records, Atlantis didn't have to do anything on the three newly-annexed streets, but write tickets.

The county commissioner, after hearing this, quickly launched an investigation into the annexation and the city's policy of writing tickets outside of their city but not working crashes on those same streets.

Crashes

Next I went to the Palm Beach County traffic division to collect data on the number of crashes that occurred on the streets annexed by Atlantis.

The spreadsheet showed that while Atlantis police spent a lot of time patrolling the streets for speeders, they hardly ever worked the crashes that occurred there.

Again, while I found the officials all too willing to comply with the sunshine law, the data was not in an electronic form. I found myself thumbing through oversized notebooks copying the data into my notepad, which I later transferred to Excel.

The spreadsheet showed that while Atlantis police spent a lot of time patrolling the streets for speeders, they hardly ever worked crashes that occurred there. Of the 733 crashes that occurred on those three roadways from 1996 to March 2000, Atlantis police worked three, but none of the five fatal ones.

In an interview, the police chief didn't admit to the over-enforcement, but simply said his residents didn't speed. He added that the department's written policy stated officers are supposed to spend only 30 minutes per eight-hour shift parked outside the city walls. He also said that the reason for the city's enforcement was because they wanted to reduce crashes and that cars passing his city were not only speeding they were going well over the limit.

After getting back to the office, I ran a few quick queries to check out the chief's answers.

Chief's answers

I found that despite the increased traffic enforcement in the area, crashes along the three roads went up about 15 percent while the amount of traffic in that area went up by only 9 percent.

I also found at least four officers who broke the 30-minute rule, showing up in the database for writing tickets over as much as a five-hour period.

I set up another interview with the chief.

When I told him what I found he said that his officers had gotten so over-zealous in their pursuit of traffic tickets he had to issue a moratorium on writing tickets.

He also said that inside the city's limits police follow a policy that allows officers to ticket someone going only 5 to 9 mph over the limit, if they have received a prior warning.

Luckily, we had created a separate Access database of all of the warnings issued over the year, so when I got back to the office I ran a query for all tickets issued inside the city to people traveling less then 10 mph over the limit and then ran that list against our database of warnings.

I found 14 occasions where officers ticketed drivers who were traveling 9 mph over the speed limit but had no prior warning.

Again I met with the chief. He said that 11 of the 14 tickets were written by two officers who are no longer with the department and that the problems were anomalies.

The end result of the data entry, queries, spreadsheets, and back-and-forth interviews was a two-story package that outraged readers and lead to the county commissioner re-evaluating the county's annexation procedures.

Brian D. Crecente can be reached by email at crecenteb@rockymountainnews.com

Upcoming CAR Boot Camps:

July 15-20
August 5-10

Missouri School of Journalism
Columbia, Mo.

These unique seminars train journalists to acquire electronic information, use spreadsheets and databases to analyze the information, and to translate that information into high-impact stories. In addition, the institute then provides follow-up help when participants return to their news organizations.

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

The following transportation tip sheets are available from the IRE Resource Center, (573) 882-3364.

Tipsheet # 1171 outlines several databases that are helpful for investigating transportation safety.

More on this story appears in the Jan/Feb issue of *The IRE Journal*. A tape of the story is also available through the IRE Resource Center, (573) 882-3364, story #17404

TRANSPORTATION

Tread dangers

By David Raziq, Anna Werner, and Chris Henao
KHOU-TV, Houston

The story began with a simple conversation: a lawyer told us of a fatal accident involving something called "tread separation." In this case, he claimed a woman was driving her Ford Explorer down a Texas highway when suddenly one of the car's tires, a Firestone Radial ATX, simply lost its tread. At that point, it was just a tip, but we decided to look into it further. Additional contacts with attorneys and others revealed more cases involving the same tire and car combination: Firestone ATX tires and Ford Explorers.

Obviously, our curiosity was piqued: How big was the problem? We found that some traditional resources didn't provide the answer. For example, the government's Fatality Analysis Reporting System database did not list tire brands or makes. And the tire manufacturers, who collected tire return and failure claim data considered that information to be "proprietary." (Although such information was often turned over to plaintiffs' attorneys during lawsuits, it might be protected by court order.)

Paper trail

So we gathered our own data by using, among other documents, court records and accident reports. We also interviewed a number of experts who were in a hands-on position to know what the industry's data on these tires might be like.

We knew the limits of our approach and our data wouldn't allow for any kind of statistical analysis, but we decided that that was not our goal in any event. In the first place, any real "answers" about the allegations would only be obtained through an objective and scientific analysis of both a full body of the industry's data along with a mechanical investigation of both the tire types and perhaps even the SUV model.

Therefore, the point of our story was that there clearly seemed to be enough allegations to warrant an in-depth investigative broadcast asking whether or not there was a problem and to bring it to the attention of Firestone, Ford, the appropriate government officials, and the public itself. If that happened, then perhaps it would lead to an open and scientific look at the issue.

We went to work and soon sat down to discuss the cases we had accumulated – at the time, more than two dozen deaths that victims' families claimed allegedly resulted from tread separations on Firestone ATX tires mounted on Ford Explorers.

A spokesperson for The National Highway Traffic Safety Administration told us agency officials were unaware of any pattern of accidents involving the tire and SUV, despite the dozens of complaints from Ford Explorer owners we had found on the agency's own Web site. (Although many of those complaints did not identify a specific model or even make of tire, by this time we knew that the Radial ATX had been original equipment on thousands of Explorers and many of the complaints appeared to be about tread separations.)

Our findings were further supported in the first story, Feb. 7, 2000, by a former employee of Firestone's North Carolina tire plant, who told us about what he felt were quality control problems at the plant.

Results

Three nights later, as we were about to air a follow-up report, the president of our corporation received a multi-page letter from Firestone disputing our story. The letter called our report "misleading."

The letter also called our story unbalanced and stated "the broadcast may be grounds for finding of actual malice." We included the company's response in our February 10 story, reporting their most damaging allegations in the interests of fairness. We also posted the full letter on our Web site for viewers to read.

However, we continued to investigate, eventually learning that NHTSA was doing its own internal investigation of complaints reported by our viewers to the agency. By May, the agency announced it was beginning a defect investigation into the Firestone ATX, ATX2, and Wilderness AT tires.

As we monitored the investigation, there was another break in the story. On July 30,

Continued on page thirteen

Continued from page twelve:

Tread dangers

a source called to let us know that since late May, Ford had been recalling ATX and Wilderness tires from Ford Explorers in Venezuela because of accidents similar to ones in the U.S. We even found out that Ford was planning to include similar recalls in Ecuador and Columbia. Within days, the national media jumped on board and coverage of the issue became widespread. Finally, just weeks later Firestone made the announcement it would voluntarily recall more than six million ATX, ATX II, and Wilderness AT tires.

Obviously the story didn't stop there. We also broke the story that Venezuela was considering criminal charges against both Ford and Firestone. And, of course, events here eventually led to congressional hearings, and the recent passage of a new law that among other things sets criminal penalties for manufacturers who don't address defects and forces them to tell the federal government when they recall products in other countries.

IRE Awards

CAR winners

Several winners and finalists in this year's IRE Awards used computer-assisted reporting as a key part of their investigations. Two of those stories that won certificates, "The Body Brokers" by *The Orange County Register*, and "Treading on Danger?" by KHOU-TV in Houston, are featured in this issue [see pages twelve and eighteen].

Other certificate winners:

MEDIUM NEWSPAPERS: "Detroit Fire Department: Out of service," *The Detroit News*, by Melvin Claxton and Charles Hurt. The reporters documented irresponsibility and incompetence in the Detroit Fire Department, outlining myriad flaws – malfunctioning fire trucks, broken hydrants, closed stations – that contributed to deaths. They built databases off paper records that included dispatch tickets, run books, maintenance sheets, out-of-service reports and daily attendance records from each firehouse. IRE Resource Center story #16983.

NETWORK TELEVISION: "The Paper Chase," Dateline NBC, by John Larson, Lynne Dale, Allan Maraynes, Neal Shapiro, Andy Lehren and Mable Chan. Dateline's investigation documented how insurance companies used people with no medical training to write up reports that routinely denied patient claims after accidents. Data work included creating databases of internal company documents and medical reports, obtained from confidential sources. They used ArcView to analyze maps of State Farm district and regional offices. A free-text database was used to look for patterns in paper documents. IRE Story #17394.

ONLINE: "Our Private Legislatures – Public Service, Personal Gain," Center for Public Integrity, by Diane Renzulli, Meleah Rush, John Dunbar, Alex Knott, Robert Moore and Ken Vogel. The Center for Public Integrity investigated conflicts of interest involving legislatures in all 50 states. It found more than 20 percent of state legislators sat on committees charged with regulating the legislators' professional or business interests. They built a database from financial disclosure statements for more than 7,000 lawmakers, and another database of employers of state-registered lobbyists. IRE story #17325.

STUDENT: "Cycle of Influence: How Campaign Contributions, Lobbyist Spending and Personal Financial Interests Affect the WV Legislature," *Charleston Gazette*/ University of Missouri, by Scott M. Finn. The report showed through many different windows how personal and business conflicts affect legislation in West Virginia. For the data work, he used databases of personal financial interests of state legislators, West Virginia lobbyist spending, state lobbyists and their employers, and campaign contributions to state legislators. IRE story #17208.

The full list of winners and videostreamed excerpts of television category winners are available at *www.ire.org/history/pr/2000winners.html*

The following transportation tip sheets are available from the IRE Resource Center (573) **882-3364:**

Tipsheet # 776 provides tips on using CAR to cover transportation. Included is a list of relevant databases available in most cities and states and some story examples.

Tipsheet # 816 includes tips for aviation reporting and information on databases of transportation accidents and transportation Web sites.

These other stories on the Firestone/Ford Explorer controversy are available from the IRE Resource Center, (573) 882-3364.

Story #17124. *USA Today* alerted the nation to a problem first uncovered by KHOU-TV (Houston, TX) months earlier (see page 12): The treads on Firestone tires had peeled apart on 90 Ford Explorers killing four people. *USA Today*'s national coverage of the problem contributed to Firestone's eventual recall of the defective tires.

Story #17282. In a series of stories, *The Los Angeles Times* findings included that: "State Farm Insurance Company had notified federal regulators about problems with Firestone tires as far back as 1998, but got no response" and that "Ford Motor Co. was aware of instability problems with its Explorer SUV...but twice had declined to make design changes..."

Continued from page one:

Firestone

The analysis was based on the Fatality Analysis Reporting System of the National Highway Traffic Safety Administration. This is the main source for much of what we know about fatal accidents and has documented an extraordinary increase in auto safety and decrease in fatal accidents over the last decades.

The database is based on accident investigation data compiled from police accident investigators across the country, and includes information on various factors, from the weather, to road conditions, speed, alcohol use, and mechanical and driver problems that the investigators concluded might have contributed to the accident.

This was remarkable (some may say debatable) because reports of tire-related deaths were rare events.

This was remarkable (some may say debatable) because reports of tire-related deaths were rare events.

I opened up the online documentation in Microsoft Word and searched for the word "tire" and then "rollover" and I was on my way. The FARS dataset is one of those Internet wonders. Getting it once required weeks of work on the mainframe. You can now download the manual and the data in minutes. And it was available as both a text file and as a SAS file that I was able to load into SPSS with little effort. [FARS is also available from NICAR] There is even an online query system you can try out on your own at *http://www.fars.nhtsa.dot.gov/www/query.html.*

Trouble spots

But as I spent more time with the data, I realized that the results were not quite right and wouldn't be until we solved a series of technical and reporting problems. We at first used the coding for vehicle type we found in the federal database and immediately discovered a problem. Since the Ford Explorer was the successor to the Bronco II (which became notorious because of rollover problems in the 1980's), the FARs people used the same coding for both vehicles.

That's where the extraordinary resources at *The Times* began to make a difference. Keith Bradsher, our guy in Detroit, said, no, no, no, you can't combine the Explorers, which began to appear in 1991, with the older Bronco. Older vehicles tend to have more fatal accidents (especially the Bronco II) and Ford's own safety data compares the Explorers to other vehicles in use since 1991.

So we followed Ford's own methodology and our first problem was solved.

The next problem was how to account for the fact that accidents are related to how many vehicles are on the road. The more cars on the road, the more accidents. We needed to standardize the information by computing some kind of rate.

It turned out that all the information about how many vehicles were on the road were owned by a private data company, The Polk Company. The people at Polk were very helpful in providing some dollops of data to us, but when we first asked for lots of year-by-year data, we had a problem. (You can probably get use of their entire dataset for under $100,000).

Instead, we borrowed some techniques used in medical research and did what the epidemiologists call a retrospective study. Instead of asking what percent of cars on the road had tire accidents, we looked backwards and figured out what proportion of fatal accidents were caused by tires, and then we compared the experience of Ford Explorers to all sport utility vehicles.

Finding the story

The results: Explorers involved in fatal accidents were nearly three times as likely to have tire problems as other vehicles. The numbers were small, though were statistically significant, and the trend was clear.

Though the database did not identify tire brand, we used the vehicle coding in the database to identify the two plants used to produce Ford Explorers in the mid-1990s. We showed that Explorers produced at plants that

Continued on page fifteen

Continued from page fourteen:

Firestone

installed Firestone tires had far more accidents than Explorers produced at the other plant.

But what did it mean. We already knew that Explorers had a problem. So what was new? Stephen Engelberg, the new Investigations editor at *The Times*, was relentless and convincing in arguing that the government and the auto industry should have, or at least could have, been monitoring the data and been aware that something was amiss at least by 1998.

Working with Matt Wald in Washington, we put together the story on deadline. We were refining definitions and rerunning tables and graphics up until the evening.

The next morning, I went back to the FARS Web site and was stunned to discover that in the few days I had spent working on Ford Explorers, the agency had released another year of data. I reran my programming in SPSS against the new data and found that there was a surge in tire-related accidents in 1999, more than in all previous years combined.

This was a clear indicator to me that the number of Ford Explorer-Firestone tire complaints was certain to rise in the coming months. But my editors were skeptical, so we made an extraordinary effort to verify that the Explorer accidents actually involved Firestone tires.

We matched the list against the list of fatal tire complaints, and in other cases tracked down accident reports, and in some cases the state troopers or local police who filed them to determine whether the accidents involved Firestone tires. In some cases, officers pulled out accident photos and looked for the name of the tire with a magnifying glass.

We reported that the complaint death toll was likely to rise, and the death count has in fact risen several times since then. So never duck when someone asks you to help open a spreadsheet.

Josh Barbanel can be reached by e-mail at barbanel@nytimes.com

Data: Fatality Analysis Reporting System

Source: National Highway Traffic Safety Administration

Software: SPSS (statistical software)

The following tipsheets on how to use statistics in reporting are available from the IRE Resource Center, (573) 882-3664:

Tipsheet #1306. Neill Borowski of *The Philadelphia Inquirer* explains 10 statistical tools that are useful to know.

Tipsheet #628. This handout gives tips on regression – an average that can be used to predict values and questions of relationships between values for statistical analysis.

The Washington Post story, "Explorer has higher rates of tire accidents," is available in the IRE Resource Center, (573) 882-3364. Story #16864.

TRANSPORTATION

More Resources

Dawn Fallik of the *St. Louis Post-Dispatch* compiled a tipsheet of useful transportation Web sites for the 2000 Lexington CAR Conference. Below are some of the links she included. For a complete version of this tipsheet, call the IRE Resource Center, (573) 882-3364 and ask for tipsheet #1290. Or, download it from *www.ire.org/resourcecenter/*.

General Crash/Explosion/Accident data

- National Transportation Data Archive (NTDA). *www.bts.gov/ntda/*
- University of Michigan Transportation Research Institute: The institute has a lot of projects, data and research. *www.research.umich.edu/research/proposals/proposal_dev/UM_Resources/UMTRL.html*
- Volpe Center: Part of the U.S. Department of Transportation's Research and Special Programs Administration. This Web site includes links to many government transportation Web sites. *www.volpe.dot.gov/resref/linksdir.html*

Trucking

- Motor Vehicle Carrier Association. Has statistics for commercial trucking. *www.safersys.org*
- Bureau of Transportation Statistics: Provides North American merchandise trade data by commodity type and surface mode of transportation with geographic detail for the U.S. *www.bts.gov/transborder/*
- Tanker advisory center (oil spills): 212-628-7686

Cars

- National Motorists Association: *www.motorists.com*
- First auto accident ever: An out-of-town driver ran into a New York cyclist *http://salwen.com/carcrash.html*

Hazardous Materials

- Miscellaneous hazmat links: *http://hazmat.dot.gov/otherinter.htm*

Crash scams

By Fred Schulte and Jenni Bergal
South Florida Sun-Sentinel

Just about everybody in the auto insurance business in Florida has a story about fraud, from rings that stage fake accidents to crash-care clinics that bill huge sums for needless treatments.

Unfortunately, most insurance companies are wary of talking to the media. They don't want to discuss fraud schemes for fear of drawing a slander suit from those involved, or out of concern that doing so would reveal how vulnerable they are to cheaters. But a number of public records exist to help reporters get an inside look at this multibillion dollar criminal enterprise that investigators say operates in some form in nearly every state.

Our seven-month investigation identified more than 4,000 Florida residents who played roles in at least 1,400 suspicious auto accidents statewide during the 1990s.

In "Crashing For Cash," a four-part series published in December, we exposed an underworld of traffic accident stagers, corrupt chiropractors and pain clinic owners, ambulance chasing lawyers, patient solicitors and medical testing brokers who band together to cheat insurance companies and consumers.

Under Florida law, a doctor or clinic can bill up to $10,000 for every crash patient treated. Lawyers can tap insurers for thousands of dollars more in personal injury settlements. This fast money drives the industry. Our seven-month investigation identified more than 4,000 Florida residents who played roles in at least 1,400 suspicious auto accidents statewide during the 1990s.

Hundreds of them were chiropractors or lay owners of crash clinics who had been reported to state fraud investigators for padding bills or running needless medical tests on uninjured accident victims.

Others belonged to crash rings, whose members often were related by blood, ethnicity or neighborhood ties. We entered each name we encountered into one of seven Microsoft Access databases that allowed us to keep a running total of crash-for-cash participants.

Data sources

Here are some of the primary sources we used:

• State Department of Insurance files. To our surprise, the Florida Division of Insurance Fraud allowed us to visit five of their offices around the state, where we inspected more than 500 closed cases, dating to 1995. Most had been closed without an arrest.

Patterns quickly emerged.

Dozens of chiropractors and clinics allegedly were paying "runners" to recruit patients, sometimes offering them $300 to $600 apiece, for instance.

In other cases, solicitors obtained the names of people in minor accidents by purchasing crash reports in bulk at local police stations and calling them at home with promises that they could make thousands of dollars by claiming injuries in a lawsuit.

In addition, some people had obtained drivers licenses under several aliases. We decided to check their driving backgrounds.

• Department of Highway Safety & Motor Vehicles. We accessed driving histories and accident records for people investigated for insurance fraud, accident staging or possessing a phony drivers license, the criminal charge that often gets filed against accident stagers.

We entered this information into a database, which included the name of the driver and passengers in the accident, the speed at impact, damage amount and age of the other driver.

The data showed us that the same drivers were involved in questionable crashes over

Continued on page seventeen

Some Web sites that Fred Schulte and Jenni Bergal found helpful in looking at the auto insurance crash industry:

Federation of Chiropractic Licensing Boards:
www.fclb.org
Links to state chiropractic licensing boards and other statistical information about chiropractors.

Fraud Defense Network:
www.frauddefense.com
An insurance industry-funded site that offers a daily news bulletin about insurance fraud schemes around the country.

Florida Bar Association:
www.flabar.org
This site has press releases about disciplinary actions, but is difficult to negotiate because it lacks an index of disciplined lawyers.

Florida Department of Health Physician Profile Information:
www.doh.state.fl.us/IRM00profiling/searchform.asp
Used to check arrest histories and disciplinary status for chiropractors.

Continued from page sixteen: # Crash scams

and over. Some were members of the same family. Some targeted elderly drivers as victims.

One man with six aliases was in three dozen crashes. Eight people living in Palm Beach County wracked up 139 crashes. One family-based ring showed up in 68 suspicious crashes.

We used standard industry criteria for determining whether a crash was suspicious. For instance, often the accident was little more than a fender bender and the estimated damage to the vehicles was under $500. Yet the accident stager and his passengers would immediately go to a chiropractor's office and begin a series of expensive treatments.

In reaction to the series, Florida lawmakers are expected to take up a number of legislative fixes, including one that would regulate auto accident clinics and limit their ownership to medical professionals

Many other crashes in our sample occurred in parking lots at about 5 miles per hour.

• Civil court files. Increasingly, auto insurers are refusing to pay bills they believe are inflated or bogus and are pressing their cases in court. Records of these civil lawsuits, especially depositions of doctors and clinic owners, are invaluable, both for locating patients and establishing questionable billing practices and medical testing abuses.

Other court cases can be helpful as well. Insurers in some states, for instance, have filed class action suits against chiropractors and clinics, claiming they have deliberately inflated costs or billed for services that weren't rendered. Some of the biggest cases have been filed in Chicago and New Jersey.

• Disciplinary and criminal case files. We obtained a list of all Florida Board of Chiropractic Medicine disciplinary cases in the last five years and requested hard copies of cases that involved insurance fraud or patient solicitation, including penalties imposed.

More work

We combined those named in disciplinary cases with other chiropractors, doctors and clinics whose names surfaced in state insurance fraud complaints, arrest records or civil court cases for suspected accident fraud.

This list grew to more than 60 medical clinics in Miami-Dade County alone and more than 400 chiropractors, many of them reported to the state repeatedly for accident fraud.

We also examined the applications of chiropractors with criminal histories and found that the board had granted licenses to dozens of chiropractors from other states with criminal convictions including insurance fraud and cocaine trafficking.

Checking the criminal histories of accident clinic operators, we also found a number of owners with felony convictions, including ex-drug dealers who had migrated to the crash treatment racket.

In reaction to the series, Florida lawmakers are expected to take up a number of legislative fixes, including one that would regulate auto accident clinics and limit their ownership to medical professionals.

Other proposals include increasing criminal penalties for insurance fraud and accident staging and cracking down on runners who exploit crash victims on behalf of doctors and lawyers.

Judging from the early reaction among state officials, we're glad we told this story.

Fred Schulte is the South Florida Sun-Sentinel's investigative editor. He can be reached by email at fscult@sun-sentinel.com. Jenni Bergal is a senior writer at the paper. She can be reached by email at jbergal@sun-sentinel.com.

This story is available on the IRE Resource Center. Story # 17374.

More about the *South Florida Sun-Sentinel* series, "Crashing For Cash," is included in the May/June issue of *The IRE Journal*.

Tipsheets to aid in investigating insurance scams are available at the IRE Resource Center:

"Doctors, Lawyers, Car Crashes and You," Arnold Diaz, Glenn Ruppel and Janet Klein, ABC News 20/20, #14585. The story investigates automobile insurance fraud in both small town and large cities around the country.

"No Exit," Patrick Weiland, John Larson, Steve Snug, Leslie Bergman, Patricia Chang and Marianne O'Donnell, NBC News Dateline, #16392. This investigation focuses on a form of auto insurance fraud, called "swoop and squat."

IRE Resource Center, (573) 882-3364.

Body parts industry

By Ronald Campbell
The Orange County Register

When a team of reporters began work in October 1999 on the story that became "The Body Brokers," we started with a vague idea: People were buying and selling body parts donated at death. CAR played an important role in uncovering the dimensions and practices of the $500 million body parts industry.

Among the findings of the five-part series: Nonprofit tissue banks work hand-in-glove with for-profit processors, often through exclusive contracts; the pressure to get more bodies is so strong that tissue bank employees and researchers occasionally take body parts without asking; and although the Food and Drug Administration and the industry claim a near-perfect safety record for tissue products, they aren't looking very hard.

Although we explored many issues in our story – legal, political, even moral and religious issues – The Body Brokers was at heart a business story. Like any business, tissue banks and processors had to find raw material, win financial backing, develop technology, deal with pesky bureaucrats and market their products. Those universal business needs created paper and online trails.

Nonprofits

We were dealing with the intersection of nonprofit groups and for-profit, often public corporations. As a result, we often could get information from corporate documents that was unavailable from the nonprofits, and vice-versa. For example, we learned early in our investigation that the American Red Cross was among the nation's largest tissue suppliers. We knew it had a relationship with Osteotech, a New Jersey-based public company. The Red Cross wouldn't give us a copy of its contract with Osteotech. But Osteotech had included the contract in a filing with the Securities and Exchange Commission. It took just a few minutes to find the contract on EDGAR, the SEC's irreplaceable online database of corporate documents (*www.sec.gov*).

(A hint for those seeking obscure corporate documents: Read the footnotes at the end of a company's 10-K or 10-Q. These will list not only documents attached to the current filing but citations to documents that were filed years ago, including the name and date of the relevant filing. A recent Osteotech filing pointed us to a 2- or 3-year-old filing that contained the American Red Cross contract.)

When we began our research, Guidestar (*www.guidestar.org*) – a wonderful source for reporters on the nonprofit beat – had just begun posting Form 990's for hundreds of thousands of nonprofits. We had to get our 990s the old fashioned way, through a massive request to the Internal Revenue Service.

CAR played an important role in uncovering the dimensions and practices of the $500 million body parts industry.

Once we got sufficient 990s and 10-K's, we assembled them into a series of spreadsheets. Although this information was a year to 18 months out of date, it gave us an excellent idea of the industries' finances and of the nonprofit/for-profit relationships that made everything work. This put us in a strong position when we interviewed industry executives; we knew too much to be given the standard industry spin on the "gift of life."

Federal regulation

The Food and Drug Administration was an indifferent regulator of the industry. But its Web site (*www.fda.gov*) still provided many valuable clues. I strongly recommend the agency's weekly online bulletin, which lists every FDA-sanctioned recall. Its online list of warning letters also is valuable. After discovering that a simple search on "human tissue" or "bone" missed many relevant recalls, I settled down to spend a day and a half reading every FDA bulletin from the beginning of 1994 (shortly after FDA began regulating tissue banks) through the end of 1999. This provided the starting point for our story on FDA's half-hearted enforcement

Continued on page twenty-five

More about *The Orange County Register*'s series, "The Body Brokers," is included in the May/June issue of *The IRE Journal*.

A copy of the series, "The Body Brokers," can be ordered from the IRE Resource Center, story #16536, (573) 882-3364.

It can also be viewed online at *www.ocregister.com/*
Look for the link to the series on the right side of the page.

Continued from page nine:

Data galore

the N-number for their plane. The registration table in this dataset can be used to identify the registered owner of an aircraft by searching on the registration number (commonly called the N-number). NICAR updates its data every three months. The most current version, February 2001, is available for a one-time purchase for $25, $50 or $70 or an annual subscription, which includes four updates, for $50, $100 or $125.

• FAA Airmen Directory: A listing of pilots and other airmen, including the type of certificate(s) they hold and their ratings. As a result of a new law that allows airmen to have their information witheld from the public, this listing is incomplete. This data, current as of October 2000, includes about 65 percent of the total. NICAR updates its data every six months. The most current, November 2000, is available for $30, $60 or $80.

• NASA Air Safety Reporting System: Anonymous reports of airplane safety submitted by pilots, flight attendants, air traffic controllers and passengers. Anyone is eligible to file an ASRA report, and pilots appear to be the most frequent filers. Reports might include month, year and day of the week, but it will not include the exact date. Data is available from 1988 through 1999 for $65, $130 or $155

• NTSB Investigations of Aircraft Accidents: Contains information on civil aviation accidents and incidents, including pilot, crew and airplane descriptions and environment conditions within the U.S. its territories and in international waters. Data is available from 1984 through October 2000 for $75, $150 and $175

Roads

• DOT Fatality Analysis Reporting System: A good tool to analyze the number of fatal accidents on a specific roadway, in your state or in the nation. FARS is a nation-wide database of fatal road-vehicle accidents, truck and trailer accidents are included. The data files cover 1975, when the DOT first began tracking fatal accidents, through 1999. It is available from NICAR for the entire U.S., most recent year at $40, $80 or $105. A file containing 1994-1999 data merged together by NICAR is available for $45, $90 or $115. Earlier years are also available separately.

• NHTSA Vehicle Recalls and Complaints: A database of vehicle complaints, recalls, service bulletins and inspections. This data is available for the entire U.S., all years and costs for $40, $80 or $105.

• DOT Truck Accidents: A database of accidents on U.S. roads involving a commercial vehicle weighing more than 10,000 pounds, including semi-tractor trailers, buses and hazardous material carriers. This data is available for the entire U.S., all years or by each state, all years and costs $90, $180 or $205 for the entire U.S. or $30, $60 or $85 for one state.

• Truck Census: This U.S. Department of Transportation database contains records on each company that has commercial interstate vehicles weighing more than 10,000 pounds. The data in available for the entire U.S. for $95, $190 or $215 or for one state for $30, $60 or $85.

• Hazardous Materials Incident Report Subsystem: The database contains the incident reports of unintentional releases of hazardous materials for all modes of transportation (air, highway, railway, water). Among other things, the data allows reporters to determine which hazardous material carriers are having the most accidents and in which states. Chemicals most often released and exact accident details are also recorded. The report number column has unique case numbers, making it possible to join different databases for more in-depth investigations. NICAR has this data from 1971 through May 2000. This data is available for the entire U.S., all years,for $50, $100 or $125.

Waterways

• Boating Accidents: The database contains information on recreational boating accidents in the United States. The data is available for the entire U.S., all years for $25, $50 or $75.

• Boat Registration: The database contains information on registered recreational and commercial boats. This data is available for the entire U.S., current year for $40, $80 or $105.

More Aviation Data:

The Department of Transportation compiles data on air carrier service quality.

This on-time performance database currently tracks the major airlines that account for more than 90 percent of domestic operating revenues.

The U.S. DOT collects this data monthly. More information is available on the DOT Web site, *http://www.bts.gov/ntda/*

This page also includes links to information on many of the other DOT datasets.

HEALTH CARE

Medical conflicts

By David Heath
The Seattle Times

Once officials at the Fred Hutchinson Cancer Research Center figured out that we were working on a story that might not be flattering, they became quite secretive.

We wanted to know more about financial conflicts of interest in experimental trials, yet they flatly refused to give us any information about the financial holdings of top researchers or even the center itself. The fact that some of these doctors had financial entanglements with the treatments they were researching – or that patients died from these treatments – didn't sway them at all.

The state open records act was useless because the Hutch is a private institution, even though nearly 70 percent of its funding comes from federal grants. And an intriguing amendment to the federal Freedom of Information Act – one intended to open records on projects funded with government grants – was rendered worthless as well when research lobbyists prevailed in getting new regulations that severely undercut the law.

We seemed to be stuck.

With a little bit of computer programming, reading through those documents could be automated.

Unfortunately, there are no off-the-shelf databases on physicians' financial holdings. We knew that there was a lot of information to be gleaned from Securities and Exchange Commission records, patent records and the like, but we faced the nightmare scenario of reading through thousands – perhaps hundreds of thousands – of pages looking for names.

Then it hit us. With a little bit of computer programming, reading through those documents could be automated.

It wouldn't be as easy as joining the sex-offender database to the day-care worker database. But with the wealth of data easily accessible on the Internet and the sophisticated tools we have available today, it was clearly doable.

The results appear as part of our five-part series, "Uninformed Consent: What patients at the Hutch weren't told about the experiments in which they died." The series, written by Duff Wilson and myself and edited by David Boardman, documents two failed clinical trials in which doctors had financial interests and in which patients died from risky treatments without being properly informed of the risks.

To get information about financial conflicts, first we needed researchers' names. Then we would search for those names in a slew of financial documents from the SEC, an insider-trading database and the U.S. patents full-text database. We'd finally add our results to the few financial disclosure forms that we were able to obtain.

How we did it:

First we headed to the library to gather up every biennial scientific report the Hutch has ever published. These included the names of their entire faculty, including post-doctoral fellows. We sent photocopies to a vendor to be keyed into a database. That was the easy part.

Next we tackled SEC documents. Those are free through the Edgar database at *www.sec.gov*. But we had two problems: How do we download all documents for companies involved in the health-care industry? And how do we search those documents for names?

The answer to the download problem was to write a program in Visual Basic. If I ever doubted the value of learning a bit of programming to do CAR work, this experience resolved those doubts. VB includes functions that allow you to go out onto the Web and grab files. Other languages do too.

It took me two solid days to write the

Continued on page twenty-one

The Seattle Times' story, "Uninformed Consent: What patients at the Hutch weren't told about the experiments in which they died," is available online at *http://seattletimes.nwsource.com/uninformed_consent/* Or it will be available soon **at the IRE Resource Center.**

More information about Texis software is available at *www.thunderstone.com*

Continued from page twenty:

Medical

program in one of those work binges where you forget if it's almost dawn. But the payoff was running the thing and actually seeing it work. I annoyed my wife and plenty of friends by showing off the finished product.

The list of all health-care companies came from Edgar's own documentation pages. Using my program, I grabbed every annual report and proxy statement for all years and put them into one massive text file.

The next problem was what to do with this massive file. I could have written another Visual Basic program to search for each researcher's name in the file, but that would have been deadly slow. What I really needed was a text-indexing tool to make the search speedy. My first thought was to use Lotus Notes but I was lucky to have a tool called Texis, a document retrieval database that we use at the Seattle Times for our own electronic clips library.

This project is a good example of the seamless blending of CAR work with investigative reporting.

I won't bore you with the details here. I'll just say that Texis is extremely powerful and fast. Best of all it's a relational SQL database that can do joins. So once I got the text file loaded, parsed by company name, document type, year and the full text of each document, I could run a fairly simple join with the researchers' names. The query would search each document's full text looking for name matches, like an automated clip search. We searched for 1,800 names.

I had a big assist on this from research editor Tom Boyer, who developed our electronic library. For more information on Texis, go to *www.thunderstone.com*.

The result of the query was a new file with a large chunk of text on either side of a name match. I loaded that dataset into SQL Server and then started the manual process of verifying each match. With a few exceptions, this was easier than it sounds because typically the Hutch would be mentioned somewhere in the text.

Patents

The next challenge was to get all patents by these researchers. The data are available for free at *www.uspto.gov/patft/index.html*. Again, a modification of the Visual Basic program allowed me to grab off the Web all patents in which Hutch researchers were listed as inventors. I limited this search to inventors from the state of Washington, which helped to produce more clean matches.

Again, I had to weed out false hits manually. If the person invented a new type of typewriter ribbon or garden tool, I automatically eliminated that patent. However, a patent for radiation delivery to lymphoid and marrow tissues was a good candidate for a true hit. Finding true matches wasn't terribly difficult.

We searched an insider-trading database with the help of a company called First Call, *www.firstcall.com*

One last piece came from investigative reporter Deborah Nelson, who unearthed these stories but left for *The Washington Post* before she could complete the project. She had requested all financial disclosure forms for medical faculty at the University of Washington who also had positions at the Hutch. That information – useful but incomplete – was then keyed into a database.

All of these data were loaded into SQL Server. I developed an interface in Access to pull all the data we had on each researcher together on one page. We created an 'A' list of the researchers that we wanted to focus on. As we learned of other financial connections, we were able to key these in manually through the Access interface.

Continued on page twenty-two

Some sources used in the "Uninformed Consent" story:

National Institutes of Health site called Medline, ***www.ncbi.nlm.nih.gov/pubmed/***

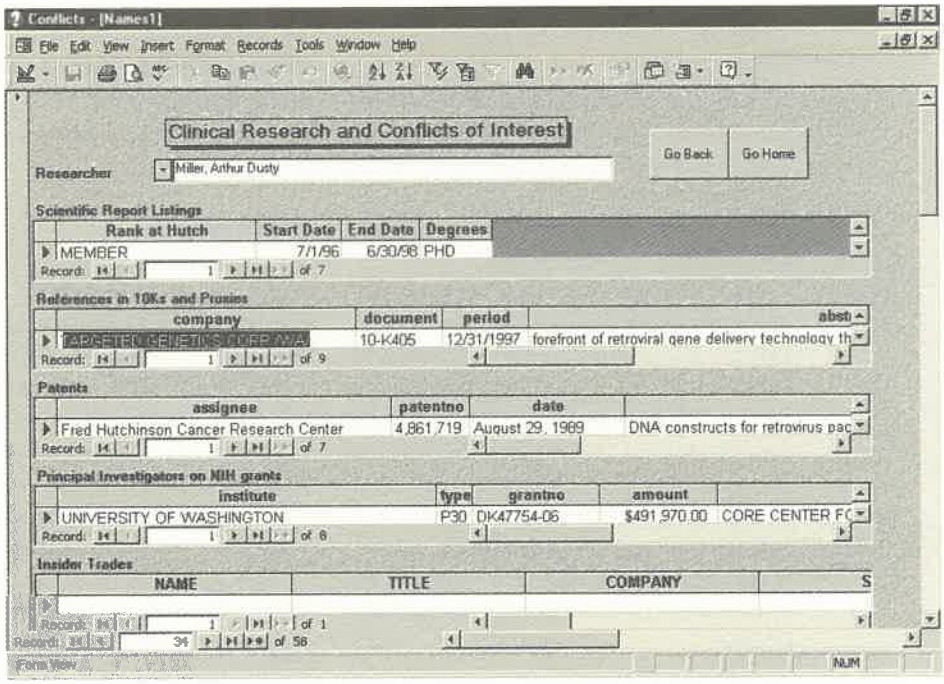
SEC Edgar database, ***www.sec.gov***

Insider-trading database through a company called First Call, ***www.firstcall.com***

Continued from page twenty-one:

Medical

A quick note: to add researchers to the 'A' list later, I used an incredible tool in SQL Server called triggers. FoxPro has the same capability. The trigger would automatically launch a stored procedure to create a unique ID for any name we added to the 'A' list and then automatically search all datasets and add that unique ID to other tables to allow for joins. In short, you type in a name and suddenly all the data about that person magically appears.



Triggers may be routine for database developers, but I'd never heard of them being used by CAR folks. I was awestruck by them and encourage you to investigate.

Medline

In the course of our reporting, we also downloaded abstracts of medical journal articles written by Hutch researchers in the past decade. These data are available through a National Institutes of Health site called Medline, *www.ncbi.nlm.nih.gov/pubmed/*

Downloading them involved a total of about 18,000 searches. My original program ran as fast as possible, and I hadn't really thought about the demand that this would put on NIH servers until the annoyed database administrator blocked our access. I called to let him know what we were doing and he gave us access again providing we spaced our searches five seconds apart and ran the program after 9 p.m.

We pulled full copies of articles for researchers with financial interests the old-fashioned way, photocopying them at a medical library. This effort produced a mere paragraph in our series about how researchers don't disclose their interests in medical journals, even when the articles are about their patents.

We also made heavy use of our death certificate database in tracking down patients. Other data included all grants from the NIH and all clinical trials ever advertised by the Hutch on the National Cancer Institutes' Web site.

After doing the CAR work, I assumed a more traditional role of a shoe-leather reporter on the series. This project is a good example of the seamless blending of CAR work with investigative reporting. Even though the series was aided by the most sophisticated data work I've ever done, the focus was not on data but on the compelling stories of the patients, their families, a whistleblower, a financier, a Nobel Prize winner and the Hutch's doctors.

David Heath can be reach by e-mail at dheath@seattletimes.com

Online sources for The Body Brokers:

- **Securities and Exchange Commission EDGAR database:** ***www.sec.gov***

- **Food and Drug Administration:** ***www.fda.gov***

- ***The Register*** **also found industry Web sites by using search engines, such as Alta Vista (*www.altavista.com*) or Google (*www.google.com*), to see how the industry marketed itself.**

- **They obtained non-profit 990's from the Internal Revenue Service, however most are now available at Guidestar, *www.guidestar.org***

Preparation pays off

By Robert Gebeloff
The Star-Ledger

Let's be frank – the 2000 Census is a four-alarm headache.

There's a mountain of data to climb, and at the summit sits some of the most nondescript field names in the history of data processing.

Quick – how many P0010005s did you have in your state? And what did they call P001005 in 1990 and 1980? Then there's the matter of mapping. Learning how to use the canned maps that come with ArcView is no easy feat. But learning how to convert Tiger maps and tie them to the Census data – that's an entirely new skill.

And then there's the none-too-small matter of journalism – You've conquered the data, now what's the story? Can you transform this mountain of information into news, news that would qualify as left(news,3), rather than a rehash of trends you've already reported to death?

Be prepared

All of this requires preparation – the Census is just too big to do well on deadline.

At *The Star-Ledger*, we took all the standard measures. We sent people to IRE Census seminars, we acquired and processed historical data, we did battle with the Tiger Maps, and we devised and revised our list of stories, graphics and maps throughout the weeks leading up to C-Day. But we also made a major investment in what I'll call logistical preparation. We cover the state and compete in eight local markets as well.

On Census release day, we would need to get data that was cleaned and analyzed to our bureaus in Washington and Trenton, not to mention Woodbridge, West Caldwell, and Morristown – on deadline.

What we came up with was a somewhat Rube Goldberg-esque system of Access tables, SQL Server Views and intranet Active Server Pages. The idea was that the data would arrive on Census day, a button would be pushed, and poof – the 2000 numbers would flow into place, fuel for thousands of preprogrammed calculations.

Smooth sailing

I expected, of course, that the system would cough and wheeze a few times along the way, but to my amazement, it worked without a hitch, and enabled us to execute our coverage plan pretty much to the letter. The New Jersey files arrived on my hard drive at about 3:15 p.m. on a Thursday afternoon. By the time the last edition was hitting the presses, we had managed to get seven stories in the paper, illustrated by four photos, five maps, and 15 charts. The stuff was flowing through production so fast that in one early edition, a map got printed twice, but otherwise, the paper came out looking like it was the result of a month-long project. And then we were well positioned to come back Sunday with more.

All of this requires preparation - the Census is just too big to do well on deadline.

I mention quantity, but having prepared was a matter of quality too. For example, we got the 1990 PL 94-171 data and went to the voting precinct level to generate demographics for the current state legislative districts. Then we coded each district based on the make-up of the three-member delegation. This enabled us to do precision reporting on political shifts.

While the competition noted generally that population growth appeared to be greater in Republican-controlled districts, we reported precisely that 422,000 of the state's 684,000 population increase occurred in districts with all-Republican representation.

Showing growth

Same idea with sprawl. It was obvious by looking at county numbers that the growth in New Jersey was occurring in places that were mostly rural and suburban. But why approximate? We got a list of Urban/Rural/Suburban town classifications as labeled by the state planning department. We simply tied that list to the Census data and voila – we were able to say precisely that the popu-

Continued on page twenty-four

Census Lessons:

We'd like to hear from those of you who had interesting experiences in working with the PL94 Census Data that was released in March.

Are there lessons you learned that would be beneficial to others?

If this was your first experience with Census data, was there something you wished you had known earlier?

Did your news organization come up with a unique story based on the data?

We'd like you to share your experiences in an upcoming issue of *Uplink*, preferably by writing a piece yourself. We could also arrange for a NICAR staff member to interview you for a story.

The stories can be anything from a few paragraphs to a few pages – whatever it takes to share what you learned.

Please contact us as soon as possible if you're interested. Contact Mike Sherry at mikes@nicar.org or (573) 884-7332.

CENSUS STORIES ONLINE:

The Los Angeles Times:
www.latimes.com/news/nation/reports/census/

Daily Camera (Boulder, CO)
www.thedailycamera.com/census

Tribune-Review (Western Pennsylvania)
www.triblive.com/census/

Newsday (Long Island, NY)
www.newsday.com/census

Journal News (Northern New York City suburbs)
www.thejournalnews.com/census

The Post and Courier (Charleston, SC)
www.charleston.net/census2000/

The Roanoke Times (Virginia)
www.roanoke.com/roatimes/census

Continued from page one:

Heavy lifting

5,054 that stop at the tract level. Those are just the counts. Our readers need to see percents and percent changes to comprehend the counts. And we need to sort counties, places, tracts, and maybe even block groups and blocks by those percents and percent changes to find the "most" and the "least" to salt and pepper our stories.

Maybe this will be where the journalists are separated from the demographers. How many of us really need to work through 15 tables of 209 variables of Sex-by-Age for separate Race/Ethnicity categories? There is one variable per year through 99, then years are grouped from 100 up.

But can we afford to ignore the nine tables of 49 variables each of Sex-by-Age for Race and Ethnicity categories, with ages grouped?

The last time around we all were in races to see how much we could get into tomorrow's paper from today's download. That approach won't work this summer. Instead, we need to learn how to cherry-pick.

The last time around we all were in races to see how much we could get into tomorrow's paper from today's download. That approach won't work this summer. Instead, we need to learn how to cherry-pick. That means learning how to reach in and pull out the most delectable morsels of data from an enormous pile that will crush us if we're not prepared.

Cherry-pick

There are some obvious entry points. For instance, Table PCT14, which has just seven variables describing "Unmarried-Partner Households by Sex of Partner" down to the tract level. It is the first time there has been such a table in census data.

In fact, few of the 104 tables in the 1990 Summary Tape File 1 will have matching corollaries in 2000. The same race and multiple race definitions that made PL94-171 difficult to compare will plague us in Summary File 1.

Even something so basic as comparing age groupings from 1990 to 2000 will be difficult because the age groupings are different in each census.

Where the stories from PL94-171 data largely focused on change over the decade, stories from Summary File 1 probably will paint more detailed pictures of who we were in 2000 with less of who we were in 1990. That's because of the richness of detail to be presented in the new data.

Where five fields described Hispanic Origin in 1990 there are 31 in 2000. Where there were no tables in 1990 describing American Indians and Alaska Natives, there are three tables with 47 variables each to do the job in 2000. Even Asian origin, which was detailed in 1990 will be more detailed in 2000 with the separation of Chinese and Taiwanese and the addition of Bangladeshi, Indonesian, Malaysian, Pakistani, and Sri Lankan.

The Group Quarters population also will be richly detailed, down to the tract level, this time, with categories of institutionalized and non-institutionalized persons broken down by male and female in three age categories and by race and Latino/Hispanic ethnicity.

All of those are in addition to the more detailed household and family composition tables coming this year compared to those provided in 1990.

What we learned from PL94-171 at the *Los Angeles Times* was that planning and practice made all the difference in our ability to deliver a comprehensive census r eport on deadline. I know that many of you learned the same lessons because it was apparent in the questions and answers shared on nicar-l and census-l.

That job is going to be tougher the next time around, despite assurances that we'll get several days to wrestle the data under embargo before we publish. It will be tougher because there will be so much more of it that we'll be buried by it if we don't have sound plans for managing it.

Planning

That planning can only begin by getting the documentation and studying it. That means not only the technical documentation for Summary File 1, but also the documentation for STF 1 from 1990.

Continued on page twenty-seven

Continued from page twenty-six:

Heavy lifting

Then comes the decision-making. How many stories are in there? How many geographies need to be covered? When do they need to be covered? Can you concentrate only on the summary levels for states, counties and places, leaving tracts, block-groups and blocks for later?

Which tables can you ignore? Which tables provide the big picture and which include details you'll never get into print or on the air anyway?

Is your computer up to the coming task? Is your software? If you were working in Excel or Access, can you do so again with so much more data involved? Or, do you have to figure out a way to divide and conquer?

Lessons learned

Here are a few things I learned the first time around that may be of help.

A quick way to cut down the data load is to start with a copy of the "Geo" file and discard all of the SUMLEV values you don't need right away. I got rid of everything below the county level except whole places, whole tracts and whole block groups and cut the file size to about a fifth of the original. Then when I merged it with the tables, they were cut down to size as well.

Cut local and regional files out of state files and use those smaller files for detailed queries. When I needed to know which tracts had the most or least in Los Angeles County, I looked only at Los Angeles County data.

Build universal query report forms. Mine were in Microsoft Access and fed from a series of make-table queries that each over-wrote the same table name. The reports ran from that intermediate table, so the field names were always the same. A big surprise was learning that Access can't make decent HTML tables from multi-column reports (true for both Access 97 and Access 2000). The solution was Adobe Acrobat. We printed the reports as PDF files with the Acrobat PDF Writer. They looked good, were easy to manage on both our library intranet site and our Web site, they were searchable, and they could be printed.

Adobe Acrobat was the tool of choice for mapping, too. The PDF files carried vector-graphic colored polygons from ArcView theme maps into Freehand where our artists could tweak them at will. Plenty of testing with 1990 data and then data from states released before California went into this process. The result was the ability to publish 14 maps on deadline.

But some things don't change, or at least change very slowly. Just as it was in 1990, the best way to get the data to our reporters in outlying bureaus in 2001 was the ubiquitous fax machine. Now if only I could figure out how to get everything I want on a report without resorting to 8-point type.

Richard O'Reilly can be reached by e-mail at Richard.Oreilly@latimes.com

The SF1 (short form) data that starts coming out this summer will be released to news media on an embargo basis (about 3-5 days).

Starting in early May, the Census Bureau planned to start releasing summary information from this short form data. These will be in one-page Excel spreadsheets or PDF images. Samples are available through Steve Doig's Reporting Census 2000 Web site, *http://cronkite.pp.asu.edu/census/*

(check under "What's New)

CENSUS-L

Lots of kids

The following is a recent discussion on CENSUS-L concerning a sliver of the PL94 data that could produce an interesting story:

I'm just starting to tackle a back-burner story from the PL94 data: Neighborhoods with lots of kids. I found one census tract in my area that had more kids than adults. Then I ran the numbers for the entire state, and found that this is the only census tract in all of Virginia with more kids than adults. My question: Has anyone else run across tracts in their area that have more kids than adults? That would be tracts with more than 50% people age 17 and under. I'm guessing this would be found in inner-city, public housing neighborhoods, which is where my special tract was.

A word of caution: make sure there aren't group homes in that tract. You might want to contact social service agencies (state and local) to find out for sure.

I'm working on a version of the where's the kids story, and two trends have come up. First, we're seeing significant flight away from our growing cities. Second, the kid concentrations are turning up in the same block groups where the Mormon temples or Indian reservation towns are.

Bits, Bytes and Barks

Data updates at NICAR

The following datasets have recently been updated at the NICAR Database Library:

The 1999 FBI Uniform Crime Reports: Includes summary information from law enforcement agencies throughout the country on index crimes. There are six tables to the database, including one on homicides and another on arsons. Cost for the entire U.S., all databases: $160, $320, $345 depending on size of news organization. State slices are available for: $55, $110, $135. You can also purchase just one of the six databases. More information is available at *www.nicar.org/data/fbi/*

The 2000 National Bridge Inventory: Includes the most recent inspection for all bridges in the United States. Bridges include those for highways, railroads, pedestrian-bicycle, overpasses, and those designated as historically significant. The data has been used by many news organizations to document problems with bridges in local areas. It is also useful to have available in case of a breaking news story involving a bridge. Cost for entire U.S.: $40, $80, $105. A single state is: $15, $30, $55. More information is available at *www.nicar.org/data/bridges/*

FAA Accidents and Incidents: Includes all accidents and incidents in the United States involving all types of aircraft. The most recent incidents included in the data are from early 2001, however not all incidents from 2000 were entered into the database by the FAA by the time NICAR received the data. The most current data, 1990 through 2000, costs: $40, $80, $105. NICAR also has data going back to 1973 that is available separately. More information is available at *www.nicar.org/data/faaai/index.html*

Reporting Conference Tipsheets

Tipsheets from the IRE Reporting Conference, held March 31 in Columbia, Mo., are available online at *www.ire.org/training/2001studentconf/tipsheets.html*

The one-day conference featured a variety of topics geared toward students and young journalists. Panels included:

- Finding out about licensed professionals
- Using property records
- Using FOIA and local sunshine laws
- Covering cops and courts
- Producing the story for television

Also available, is a transcript of the keynote speech by Diana B. Henriques of *The New York Times*. Her speech discussed, "What every reporter should know about business."

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