

# Uplink

June 1999

A newsletter of the National Institute for Computer-Assisted Reporting

## DEADLY PRACTICE?

### Cosmetic Surgery

By Fred Schulte

*Sun-Sentinel, Ft. Lauderdale*

A few hours before dawn on April 1, 1998, police found a disoriented woman lying on the sidewalk outside a Fort Lauderdale medical office. Bandaged, bleeding and still attached to a catheter and heart monitor, she was suffering from seizures brought on by an overdose of anesthetic given during liposuction.

The woman had stayed overnight in a doctor's office to recover from surgery. She awoke in the middle of the night gasping for breath and panicked when

nobody answered her screams for help. Thinking she would die if she didn't get help, she crawled out to the street and collapsed. She spent five days in a hospital intensive care unit recovering.

I got a tip about the woman's ordeal from a source who said many poorly trained doctors did lengthy cosmetic operations—sometimes eight hours or more—in their offices and then let patients stay overnight to recuperate. This wasn't the first disaster, the source said.

#### Identifying Victims

The source was right. Courthouse records showed the surgeon whose patient had ended up on the sidewalk had nearly a dozen malpractice suits filed against him in just two years of practice in Fort Lauderdale. There were women who alleged they had been scarred horribly by botched breast-implant surgery and several people who couldn't close an eye or move their faces fully after a face lift.

I decided to sample other plastic surgeons who advertised in our newspaper or in the Yellow Pages; many of them also had high numbers of malpractice suits. (Florida health officials consider three or more claims in a five-year period to be a red flag.)

Next stop: Florida's Board of Medicine, which disciplines doctors and can be searched online. Some of the doctors in my sample also had been disciplined, in a few cases more than once, by the state medical board.

But it was the early interviews with several plastic surgeons that persuaded me to pursue this as a major project. I had never heard so many doctors so critical of their colleagues. They spoke openly about the need for tighter state regulation of office surgery and criticized the state medical board

**See Surgery, page four**

#### LEGAL DRUG HAZARDS

### Inside the FDA

By Stephen Fried

*Philadelphia Magazine*

"Bitter Pills: Inside the Hazardous World of Legal Drugs" began as a personal quest for answers after my wife had a severe adverse reaction to one sample pill of a new, heavily-marketed antibiotic. It grew into a magazine article, which triggered an FDA investigation and won the 1993 National Magazine Award for Public Service Journalism. It eventually led to a five-year investigation of the entire pharmaceutical-industrial complex. The book focuses on issues of international drug safety while exposing the inner workings of the Food & Drug Administration, the pharmaceutical industry, physicians, nurses, pharmacists, consumer advocates and the medical press. Unlike earlier books on these subjects, "Bitter Pills" weaves a personal narrative through the hardcore investigative reporting, putting a human face on the "legal drug problem." Even though

**See Pills, page three**

### Inside Uplink

This issue features five reporters honored in the 1998 IRE awards who used CAR to complete their work.

We start with Fred Schulte of the *Sun-Sentinel*, Ft. Lauderdale, who used a closed insurance claims database to find deaths from cosmetic surgery.

Read how Stephen Fried used extensive Internet research and drug databases to uncover hazards in the legal drug world for his book, "Bitter Pills."

We interview Vince Gonzales of CBS Evening News about his Internet and database work for "Tomb of the Unknown," which named an unidentified soldier.

Mark Fazlollah of *The Philadelphia Inquirer* expands on a previous *Uplink* story about their series, that proved police skewed crime statistics.

And Sherry Sontag tells how she and her co-authors built databases to track personnel from military submarines.

**PAGE 12**

*Tech Tip*

**PAGE 14**

*On The Internet*

**PAGE 15**

*Handout of the Month*

# Uplink

June 1999

Volume 11, Number 5

A newsletter of the National  
Institute for Computer-Assisted  
Reporting

## EDITOR

Brant Houston  
brant@nicar.org

## MANAGING EDITOR

Cindy Eberting  
cindy@nicar.org

## SENIOR CONTRIBUTING EDITORS

Len Bruzzese  
len@ire.org

Anna Brutzman  
anna@nicar.org  
Richard Mullins  
richard@nicar.org

## COPY EDITORS

Jessica Larson  
Jeanine Davis

## ART DIRECTOR

Wendy Charron

## DESIGNER

Kerrie Kirtland

## STAFF

Dawn Fallik  
Jason Grotto  
Seth Jayson  
Ted Peterson  
Noemi Ramirez  
Aaron Springer  
Neil Warner

*Uplink* is published every month  
by the National Institute for  
Computer-Assisted Reporting.  
138 Neff Hall Annex  
Columbia, MO 65211.

(573) 882-0684. Subscriptions are  
\$40 for IRE members, \$60 for  
nonmembers.

Postmaster: Please send address  
changes to NICAR.

Send email to [jeanine@nicar.org](mailto:jeanine@nicar.org).

NICAR is a joint effort of  
Investigative Reporters and  
Editors and the University of  
Missouri School of Journalism.  
NICAR services include hands-on  
newsroom training in computer-  
assisted reporting, special  
academic and advanced training  
in data analysis.

NICAR is supported by grants  
from The Freedom Forum and  
other foundations intended to  
help the institute deliver its  
services nationwide to news  
organizations and associations.

USING MILITARY DATABASES, WEB SITES AND LISTSERVES

# Naming the unknown

By Dawn Fallik  
NICAR Staff

Every day hundreds, if not thousands, of tourists gather at the Tomb of the Unknown Soldier in Washington D.C. to watch the changing of the guard and honor the memory of unknown men who died in battle.

So when CBS reporter Vince Gonzales began investigating whether the tomb's Vietnam era Unknown Soldier had been identified by military officials before being buried, he treaded softly around both memories and rituals.

One officer told us "He's not a human being anymore, he's a symbol," said Gonzales.

The initial seven-month investigation involved multiple military databases as well as Internet archives and personal interviews. Four months after the first story aired, the Pentagon ordered the soldier's body exhumed from the tomb. When DNA tests proved the unknown soldier was indeed decorated Air Force Lt. Michael Joseph Blassie, his St. Louis family reclaimed their son's body and buried him with his own tombstone.

## Military listservs

Gonzales said he began investigating the story after seeing an Internet posting by Ted Sampley, a reporter for U.S. Veteran Dispatch. Sampley's story described how documents pointed to Blassie and suggested that DNA testing could be used to confirm the suspicion. Gonzales went further, using interviews, documents and databases as confirmation until the government was forced to unearth the tomb's soldier.

"At first I thought it was another one of those Internet conspiracies," he said. But as he started looking into the details of the Unknown Soldier selection process and talking to the individuals involved, Gonzales was convinced Blassie's body was beneath the tomb.

Without mentioning the words "tomb" or "unknown," Gonzales quietly began requesting documents regarding Blassie, a 24-year-old pilot killed on May 11, 1972 during a mission. He requested all documents, including forensic reports, squadron logs and e-mail records that mentioned Blassie's name. During the investigation, Gonzales sent more than 20 FOIA request, to every branch of the

Department of Defense as well as the U.S. Department of State and many others. He received documents from almost all of the requests.

## Searching National Archives

In addition, the reporter turned to the National Archives in Washington, D.C. for electronic database help.

"We asked the Center for Electronic Records for the National Archives to run several searches for us on databases we didn't have. They were really intrigued by the project and so they helped us out a lot," Gonzales said. The center does charge for searches, and the station paid \$30 for documentation and eventually purchased the entire Combat Area Casualty File.

The two main databases Gonzales used were the Combat Air Summary and Detail Files (OPREA) and the Combat Area Casualty database, both available at the National Archives. OPREA includes information about all combat and non-combat air missions during the Vietnam War, including information about Cambodia and Laos. The casualty database included information on where and how military personnel died, their home of record, date of birth and other death details.

Gonzales also used The Situation Report Army File, a database of ground combat operations in Southeast Asia, looking for army officers who were supposed to be on the ground and recovered the remains.

"Not all of it was vastly useful; mostly we used it for some tidbits," he said.

## Data confirmed tips

Most of the data work was used to confirm information received from sources: who had gone down where on what day and whose remains were recovered.

All records related to the unknown selection process were supposed to be destroyed after the soldier was buried, but Gonzales found government officials who had refused to comply.

"Some people would hide the information in other files. One guy hid stuff in his garage. I now have a 5-foot tall filing cabinet filled with documents that were supposed to be destroyed," Gonzales said.

Although the Pentagon first told the reporters they would not receive any docu-

**See Soldier, page three**



*From page two:* **Soldier**

ments regarding the selection process, Gonzales said they were successful because they placed FOIAs with the original agencies and stayed away from using the Tomb of the Unknown Soldiers reference.

"It wasn't hard to get the databases, but it was hard to get clean data," he said, adding that the archive service tells users about potential obstacles. Data was put in wrong fields and names were spelled wrong.

Gonzales used the Internet to find researchers and veterans who knew about Blassie's death and the following cover-up. He suggests that reporters doing military stories use variations on the names they want when searching government documents. He spent three months trying to find retired Col. Bill Parnell, whose troops recovered Blassie's skeletal remains and his identification card. The search was stalled because documents had the name spelled Darnell.

Gonzales, teaming up with fellow reporter Eric Engberg and producer Dick

Meyer, held off from contacting the Blassie family until the initial investigation was ending.

"We didn't want to drag their hopes up, but we needed to get them to sign affidavits allowing us to get military documents that usually only family could get," Gonzales said.

On June 29, 1998, the United States Defense Department confirmed what the Blassie family and the CBS reporters had suspected for almost a year; DNA tests confirmed the remains buried at the tomb for 14 years were those of Michael Joseph Blassie. He is now buried at a veterans cemetery near the family's home in Florissant, Mo.

"It's what we had been working toward all along and the family were such great people you want to work even harder for them, Gonzales said."

The trio continues to investigate which military officials ordered Blassie's remains to be declared unidentifiable.

Vince Gonzales can be reached by e-mail

GONZALES SAID HE USES THE INTERNET TO RESEARCH STORIES AND FIND PEOPLE ALMOST EVERY DAY. HE RECOMMENDS USING A "METABROWSER" WHICH SEARCHES SEVERAL INTERNET ENGINES AT ONCE, TO GET BETTER COVERAGE OF THE WEB. HIS FAVORITES: WWW.DOG PILE.COM AND PROFUSION.COM

THERE ARE SEVERAL WEB SITES FOCUSING ON MILITARY RECORDS AND VETERANS' ISSUES.

WHILE INVESTIGATING THE TOMB OF THE UNKNOWN SOLDIER, GONZALES SAID HE USED THE POW/MIA ARCHIVE AVAILABLE THROUGH THE LIBRARY OF CONGRESS AT: [HTTP://LCWEB2.LOC.GOV/POW/POWHOME.HTML](http://lcweb2.loc.gov/pow/powhome.html)

*From page one:* **Pills**

most investigative journalism pieces focus on illegal drug use, the truth is that legal drugs kill many more people each year than illicit drugs. Adverse reactions to prescription and over-the-counter medications are the fourth-leading cause of death in America.

"Bitter Pills" is based on thousands of documents, many obtained from Freedom of Information requests from the FDA and other federal agencies. Others were rescued from discovery materials in lawsuits that were later settled and sealed. It is also based on an exhaustive survey of over 100 books and thousands of papers on various subjects pertaining to drug safety and the legal drug culture. But "Bitter Pills" also incorporates hundreds of interviews with government officials, drug company executives, lawyers, physicians, academics, consumer advocates, pharmacists, nurses and patients worldwide — as well as coverage of dozens of conferences in the United States and Europe.

I knew nothing about the drug business when I started this project. And I was faced with several problems: how to develop enough expertise on a highly technical field so I could

report and write credibly, how to journalistically handle the personal circumstances that got me interested in the subject so I might capture my passion for drug safety without being biased against drug makers, and how to humanize very technical issues. The drug industry is rife with potential conflicts of interest, but so is the world of investigative journalism. I tried to face those conflicts head-on in my reporting, and I'd advise other investigative reporters to do the same. Because my use of Internet resources was so extensive, I published a guide to the sources in the book, which was updated and expanded for the recently published trade paperback. There is also a large Web site for the book at [www.bitter-pills.com](http://www.bitter-pills.com), which allows for pretty extensive research by professionals and lay people. During the research for the book, my assistants and I obtained several electronic databases from IMS, the company that does the "Nielsen ratings" of drug use. We also obtained adverse reaction reporting databases for different drugs from the FDA, and deposition databases from most inves-

**See pills, page eight**

HERE ARE A FEW USEFUL  
SITES FOR INFORMATION  
ABOUT DOCTORS AND  
DOCTOR DISCIPLINE:

• **FEDERATION OF STATE  
MEDICAL BOARDS:** [HTTP://  
FSMB.ORG](http://FSMB.ORG).

THIS IS A GOOD SOURCE  
FOR GENERAL  
INFORMATION ON DOCTOR  
DISCIPLINE.

• **AMERICAN MEDICAL  
ASSOCIATION:** [HTTP://  
WWW.AMA-ASSN.ORG/APS/  
AMAHG.HTM](http://WWW.AMA-ASSN.ORG/APS/AMAHG.HTM).

THIS ONLINE DOCTOR  
FINDER INCLUDES THE  
MEDICAL SCHOOL  
ATTENDED AND TRAINING.

• **FLORIDA AGENCY FOR  
HEALTH CARE  
ADMINISTRATION:** [HTTP://  
WWW.FDHC.STATE.FL.US](http://WWW.FDHC.STATE.FL.US).  
ALL DISCIPLINARY  
ACTIONS AGAINST  
FLORIDA DOCTORS CAN BE  
SEARCHED ONLINE.

• **FLORIDA DEPARTMENT  
OF INSURANCE:** [HTTP://  
WWW.DOI.STATE.FL.US](http://WWW.DOI.STATE.FL.US).  
CLOSED MEDICAL  
MALPRACTICE CLAIMS CAN  
ALSO BE SEARCHED  
ONLINE.

*From page one:*

## Surgery

for being too lenient in punishing incompetent surgeons.

A couple of surgeons even said they knew of patients dying during what should have been routine cosmetic surgery. Unfortunately, none of these doctors could, or would, provide any details such as patient names or dates of death. Local medical examiners weren't much help either; they said they had no way to check without the name and date of death.

Senior writer Jenni Bergal and I began visiting morgues in South Florida and rummaging through thousands of cases by hand. One by one, we found deaths and began contacting families. We also tried to track down police reports on patients who had died in their homes a day or two after surgery. Some of these provided additional details about the operations.

By the time we published "Cosmetic Surgery: The Hidden Dangers" in November 1998, we had confirmed 34 deaths in roughly the last 12 years. Thirteen of them occurred in 1997 and 1998. We have since documented five more deaths since 1997, bringing the total to 18.

### Fatal Patter

The number of deaths stunned state regulators, medical examiners and even the gossip-prone plastic surgeons of South Florida.

A troubling pattern emerged in fatal cases. Many patients had more than one cosmetic procedure that took eight or 10 hours to perform in a medical office — longer than a liver transplant takes in a hospital. The victims tended to be in late middle age, were overweight or had chronic medical conditions that made them questionable candidates for such operations.

About half of those who had died since 1997 were men older than 50 even though women account for the vast majority of cosmetic operations in the state.

### Finding Fault

Still, we needed to keep the deaths in perspective. Nobody knew exactly how many cosmetic operations were performed in Florida every year, though it seemed from industry statistics to be at least 77,000.

Lacking any means to compute a death rate, we searched for data to bolster the case

that some cosmetic surgeons were the problem. We found the data in the Florida Department of Insurance closed-claims database, which listed all malpractice settlements paid by insurance companies. These claims include cases settled by insurance companies before any lawsuit is filed.

The state insurance department charged us \$150 for about 10 floppy disks, which we loaded into Access with the help of John Maines, our research center's database specialist.

The data included the doctor's name, location, specialty, training, severity of the injury and the money paid to the patient. Sorting by "plastic surgeons" gave us more than 1,100 cosmetic-surgery claims settled since 1990, including some deaths.

Unfortunately, the database didn't explain the type of injury or how it occurred. We had to pay 15 cents per page for paper copies of a detailed narrative that had this information. None of the documents included the patient names, which are confidential under state law. Still, we documented scores of injuries, ranging from permanent facial scars to grossly disfigured breasts after botched implant surgery.

Many doctors with multiple claims against them (several had a dozen or more) we already knew from our search of Department of Medicine disciplinary files.

I also used the insurance — claims data to create graphics. For example, we published charts showing the severity of injuries, the average size of claims and the board — certification status of doctors with large numbers of malpractice claims.

Board certification became important because plastic surgeons held that uncertified surgeons were injuring patients. In fact, the data showed that 80 percent of the injuries came at the hands of board-certified plastic surgeons.

Since our series, state officials have tightened oversight of office surgery. Florida's Board of Medicine banned overnight stays in medical offices and limited office surgery to six hours. The Florida Legislature also passed a bill that requires office-based surgeons to report deaths and serious injuries to the state.

Fred Schulte can be reached by e-mail at [fschulte@sun-sentinel.com](mailto:fschulte@sun-sentinel.com)



FBI REPORT PROVES PAPER'S FINDINGS

# Counting crime right

By Mark Fazlollah,  
Craig R. McCoy and  
Michael Matza

*The Philadelphia Inquirer*

When the FBI released its 1998 crime report for major U.S. cities last month, the Philadelphia numbers confirmed the enormity of the city's past statistical falsification—the focus of a series of articles by *The Inquirer* over the previous year. Major crime dropped 7 percent in the rest of America, but it was up 11 percent in Philadelphia. In raw numbers, Philadelphia reported 2,603 more violent crimes and 9,211 more property crimes than in 1997.

The mayor and the police commissioner explained it was not a crime wave, just the result of better police reporting after *The Inquirer* had hammered away at the issue for the past year. It was more than just an issue of dry statistics. *Inquirer* reporters had interviewed dozens of victims who were shot, stabbed, seriously beaten, raped or robbed, but the police had declared in nearly every case that no crime had occurred.

Some victims were denied state-funded compensation because the police said no crimes had been committed. And suspects stayed on the streets.

For decades, Philadelphia police engaged in a massive public relations conspiracy to make the city appear safer than it really was. The Philadelphia District Attorney recently characterized the falsification of crime stats as the city's "dirty little secret."

## Exposing the secret

The secret was exposed only after the city agreed to release zip drives with crime-by-crime data to *The Inquirer* and its sister paper, *The Philadelphia Daily News*. The database files provided a set of basic facts: dates, times, locations, crime-code classification, clearances information and more on 700,000 major and minor crimes from 1991 through 1997. Eventually, Tom Torok of the paper's computer-assisted reporting and analysis unit, put all that information on a fast Web site: <http://home.phillynews.com/crime> (See December 1998 *Uplink*).

While the police department had resisted releasing this data for several years, Philadelphia Mayor Edward Rendell finally ordered

it made public. *Inquirer* reporters sorted the data using the Access database and Excel spreadsheet programs, allowing them to produce crime analysis for each of the city's 23 precincts.

Several precinct captains were able to help us find the flaws in the numbers for their individual districts. The captains had kept their own crime statistics for their precincts, and their numbers conflicted with the official figures. Pulling together the flaws in the precinct data, we were able to report that about 3,000 property crimes had apparently been deleted from the officials' count. The giveaway was a suspicious drop in car ripoffs that appeared in only one half of the year.

The police department, after initially contending that the precincts' captains were wrong, would eventually concede that 2,559 car thefts and 560 larcenies were never counted in 1997.

It was an enormous admission. The undercount in stolen cars alone exceeded the entire number of heisted vehicles reported for many cities in 1997.

It was more, for example, than all the stolen cars reported for the year in Raleigh and Greensboro, N.C. (combined population: 454,000). It exceeded the combined 1997 total for Norfolk and Virginia Beach, VA (total population: 650,000).

The department's admission of under-reporting gave us credibility with town watch groups and victims' advocates – and even some police officials. They would all be important in the next phase of the project.

## Humanizing numbers

For our next round of stories, we wanted to put a human face on the numbers. We had to find victims whose crimes had been dumped.

Among the steps we took was to compare court records for hundreds of cases with the crime-by-crime police data. We found that far more suspects were criminally charged with aggravated assault than police said they had ever arrested for the crime.

The victims' advocates, rape counselors and town-watch groups gave us additional information about crimes they believed had never been properly reported by police.

**See Crime, page six**

THIS STORY EXPANDS ON  
A PREVIOUS UPLINK PIECE  
ABOUT THE INQUIRER'S  
SERIES IN THE DECEMBER  
1998 ISSUE.

VIEW THE CRIME  
DATABASES USED BY THE  
PHILADELPHIA INQUIRER  
AND READ THE STORIES  
AT [WWW.HOME.PHILLYNEWS.COM/CRIME](http://WWW.HOME.PHILLYNEWS.COM/CRIME)

ORDER THE LATEST FBI  
UNIFORM CRIME REPORT  
DATA FROM THE NICAR  
DATA LIBRARY. CALL  
(573)884-7332. SEE PAGE  
16 FOR A DETAILED  
DESCRIPTION, INCLUDING  
STRENGTHS AND  
WEAKNESSES, OF THE  
DATA SET.

YOU CAN ORDER A TAPE  
OF THE PANEL "USING  
CRIME STATISTICS:  
SUCCESES AND  
CAVEATS" FROM THE 1999  
IRE NATIONAL  
CONFERENCE FEATURING  
THE INQUIRER REPORTERS  
FAZLOLLAH AND MATZA  
FROM THE IRE STORE AT  
[WWW.IRE.ORG/STORE](http://WWW.IRE.ORG/STORE)

From page five:

## Crime

We interviewed civilian crime-watch leaders who had kept their own personal records of crime in their neighborhoods. Politicians alerted us to victims who had complained of police inaction on crimes.

Some police also helped us. They understood that dishonesty in crime reporting was hurting them. Because Philadelphia was supposed to be one of the safest urban areas in America, the bogus crime stats at times had been used to justify cuts in police funding.

Helpful police showed us how to work the bureaucracy to obtain the original crime reports for incidents they knew had never made it to the official count. In getting this aid, we were able to exploit a kind of glasnost brought on by the arrival of a new reformist police commissioner, John F. Timoney, who had made accurate figures a cornerstone of his crime-fighting strategy.

Timoney gyrated between two extremes as our reporting unfolded. On the one hand, he was blunt in his own lack of trust in Philadelphia crime figures. On the other, he often fought our reporting efforts when we pushed to illustrate the overall trends with concrete pictures of victims and concrete examples of police negligence.

### Back to documents

If an incident wasn't shown in the police data, we ordered the street cop's initial incident report. The city charges \$15 for each copy of an incident report. We ordered hundreds.

When we located victims, they often told us tragic stories of uniformed arrogance.

Rapes, for example, sometimes had been coded as "investigate person" — if investigators questioned the reliability of the victims. Incidents of shooting and stabbing were coded as "hospital case."

Even a murder case was affected by the downgrading. Melanie Alge, an 86-year-old widow, was critically beaten by a robber in May 1998. Police initially characterized the incident as a property crime, purse snatching. Two weeks after the beating Alge died. "Just because my sister was old is no reason to dismiss the incident as a mere purse snatch," Alge's sister would later write, demanding an investigation of police bungling.

No suspect has ever been arrested in the murder. On May 17, the day after the FBI

released its report on crime in the largest American cities for 1998, the *Inquirer* published another crime story.

This time there was no zip drive. The police department had promised to give us new crime-by-crime data but did not complete the data before the story was published.

Still, we were able to show some apparent flaws in Philadelphia's newest numbers by simply comparing them to the reports of other major cities.

In the new numbers, Philadelphia police reported 8,701 aggravated assaults — a 40 percent increase over 1997. That still seemed low.

Even with the dramatic 40 percent jump, Philadelphia's figures were at serious variance with the number of aggravated assaults reported by other large U.S. cities.

Philadelphia was 22 percent below the average per-capita assault rate for the major U.S. cities in 1998.

Nine of the 10 major cities reported more aggravated assaults than robberies. Los Angeles, for one, reported twice as many assaults as holdups.

That makes sense, if you think about it. Surely, it is more likely to get beaten in a bar fight or other scuffle than to be held up.

Not in Philadelphia. It was the sole exception among the top 10 U.S. cities. The numbers said 11,435 people were robbed last year but only 8,701 assaulted.

University of Maryland criminologist Lawrence Sherman, named by the police commissioner to lead an audit of the department's reports, had previously said the ratio of aggravated assaults and robberies was important.

Sherman said a city's statistics probably were unreliable if the relationship between the two crimes was far out of step with the national norms.

And some officials never seemed to understand the impact of the chicanery. "This isn't about how fast police respond to 911 calls, or how they deal with civilians, or how they fight crime. It's a matter of numbers," the mayor's spokesman, Kevin Feeley, told reporters when the scandal was unfolding last year.

Mark Fazlollah can be reached by e-mail at [fazlo.pia@phillynews.com](mailto:fazlo.pia@phillynews.com)

# Digging up submarines

By Sherry Sontag and  
Christopher Drew  
Authors

Computer-assisted reporting: the phrase brings to mind complicated programs written to run analyses of miles of numerical data to create a litany of statistics to prove a trend, track the market, to do myriad tasks that have little to do with the old-fashioned investigative method of running around the country asking questions—lots and lots of questions.

Certainly when we began the six-year odyssey of researching and writing “Blind Man’s Bluff,” our tale of submarine espionage during the cold war, we knew that there were no statistics to analyze, at least none we could get our hands on. Just about everything we wanted to write about was classified, and the Navy had made it clear that it was all going to remain classified. If this book was going to be written—and we were determined that it was—we were going to have to dig out 40 years of undersea espionage the old-fashioned way: one source at a time. Our analysis would be similarly low-tech: if a submarine came back to port damaged at the same time another submarine was reported missing, maybe the two incidents were related.

Still, here we are, writing an IRE article on computer-assisted reporting, and for good reason. Even for an old fashioned information dig, we quickly discovered that all the modern trappings of modems, databases, and laptops were going to prove crucial. Had the infant Internet been worth talking about when we began, that might have been a big help as well. Instead, we ultimately used the net at the end of our project, to announce rather than to dig for findings. But on to that in a bit.

## Tracking interviews

First our start. The very first thing we did was build a database—not one of statistics but names, lots of names, all of them potential sources. This was a comprehensive list of just about every man who served on any submarine we had any reason to believe may have been involved in some incident that just might, maybe, if it panned out, end up in our book. There are all sorts of great database programs out there; D-Base, Foxpro and

Access are among them. We didn’t know about any of them at the time, and so we used an old DOS program called Profile. It worked great. It did just what we needed it to do: let us list men, their phone numbers, addresses, and just what subs they served on in which years.

By the way, when we say list, we mean type. Each profile was typed in by hand until we had thousands of names. We amassed those before we went out on a single interview. Then, as we traveled around the country—and we traveled a lot—we had the means of finding men by location, boat, by year, and we hoped these men just might be willing to tell us tales that they had never told their wives, children, parents or best friends.

It’s worth noting that an awful lot of our interviews were typed directly into a laptop computer. Ultimately, the laptop proved to be no more distracting to sources than a pad and pen. In fact, often a bit of electronic show-and-tell was the perfect way to break the ice. It was also the perfect way to keep interviews sorted and to prevent scrawled notes from becoming more illegible with time. Later, when we were looking for that particularly great quote, the one we found two or three years before we were ready to write, a simple computer search dug up the tidbit every time.

While the two of us ran around doing interviews, Chris’ wife Annette Lawrence Drew scoured books and old newspapers, creating a time line of events on dry land. When were the summit meetings held, the treaties signed, when was talk of detente taking place on land while spying heated up at sea, how often did the rhetoric match the reality? All of it now rested on keyboard access and helped us create context.

As we wrote, we used a modem and a direct computer-to-computer connection to bounce ideas and drafts back and forth. And we used the computer to encrypt everything we sent. It wasn’t that we really thought our phone lines were tapped. We just knew that it was possible. So we used Norton encryption—PGP is a better choice these days—to play what we called “I-

**See Submarine, page eight**

THE AUTHORS SPOKE AS PART OF THE “WRITING THE INVESTIGATIVE BOOK” PANEL AT THE IRE CONFERENCE IN KANSAS CITY, MO. THEY WERE JOINED BY INVESTIGATIVE REPORTER STEVE WEINBERG.

TO GET A TAPE OF THE PANEL, CALL THE IRE RESOURCE CENTER AT 573-882-3364 OR VISIT [WWW.IRE.ORG/STORE](http://WWW.IRE.ORG/STORE)



STEPHEN FRIED HAS ESTABLISHED A WEBSITE FOR HIS BOOK BITTER PILLS AT WWW.BITTER-PILLS.COM. THE SITE OFFERS A HANDY GUIDE TO DRUG INFORMATION.

DIRECT LINKS TO THE FDA, PHARMACEUTICAL COMPANY WEBSITES AND VARIOUS MEDICAL PUBLICATIONS IS AVAILABLE AT WWW.BITTER-PILLS.COM/INFO.HTML

YOU CAN ORDER A TAPE OF STEPHEN FRIED'S PANEL FROM THE 1999 IRE NATIONAL CONFERENCE, ENTITLED "DANGEROUS MAINSTREAM DRUGS: HOW REGULATION BREAKS DOWN," FROM THE IRE STORE AT WWW.IRE.ORG/STORE

*From page seven:*

## Submarines

Spy," maintaining protection against the unlikely.

### Net spreads word

By the time our book was researched and written, the Internet had grown up enough to become our first marketing tool. We found new Web pages written by and for submariners. We found their bulletin boards. We created long e-mail distribution lists. We spread the word, a new book with tales long untold was coming out.

We know the effort was effective. We became one of the top sellers at Amazon.com before we were ever published. At least one man who received our e-mail on the fifth forward between submariners and spies turned around and organized a book signing that drew more than 200 people and where we signed nearly

500 books. This was the first book for submariners and their relatives that showed that the men did more than guard the Golden Gate bridge. Soon we found heated discussions about "Blind Man's Bluff" on bulletin boards dedicated to the CIA and to Tom Clancy.

It was also through the Internet that submariners came to find us, to write their concern about secrets that no longer were, to thank us for ending the silence, to offer their congratulations for a book by a couple of unknowns that surfed the best sellers list for longer than six months.

Sherry Sontag can be reached by e-mail at [sherry@panix.ocm](mailto:sherry@panix.ocm)

Christopher Drew can be reached by e-mail at [drew6@com](mailto:drew6@com)

*From page three:*

## Pills

investigative journalism various lawsuits. Much of the material came for free or nominal charges (I'm good at begging, and the FDA will waive its FOI charges if you make the requests properly) but I did have to pay stenographers for some of the depositions.

In fact, the FDA is so strapped for cash in its investigative branches that I was asked to donate my copy of a deposition — which cost me \$500 — to the regulators because they couldn't afford it and couldn't go ahead with their investigation without it.

I had great difficulty, especially with FOI material. It reached the point where I had to find the person within FDA who had the material I wanted, get them to pull it and put it on their desk, and then re-contact the FOI office and tell them exactly where the material I was requesting could be found. All this FOI material, federal FDA data in particular, is deeply flawed because doctors don't report most adverse drug reactions, and the FDA is dramatically underfunded in their ability to analyze the data. I overcame these difficulties by doing extra reporting.

There has been very little good investigative journalism on the legal drug world because the work is so difficult; the FDA is

notorious for not filling FOI requests just as drug companies are notorious for ignoring the queries of journalists.

Because I wanted the book to both ex-

---

**The FDA is so strapped for cash in its investigative branches that I was asked to donate my copy of a deposition — which cost me \$500 — to the regulator because they couldn't afford it**

---

ploring the world of legal drugs and help other journalists do the same, I was careful to document in the book how certain information was obtained, leaving, I hope, a blueprint for other journalists to explore drug-related topics.

Stephen Fried can be reached by e-mail at [sfried@compuserve.com](mailto:sfried@compuserve.com)



# Look Before Touching

By Neill A. Borowski  
The Philadelphia Inquirer

You've arm-wrestled with the government agency for weeks. After haggling, they agree to make the data available to you. Finally, the shiny CD-ROM arrives in the mail.

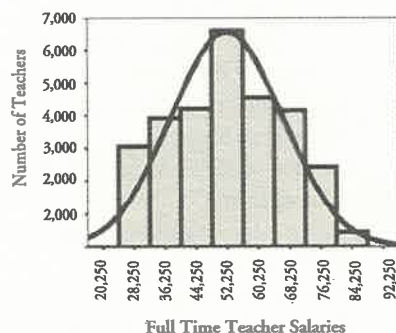
Quickly, you pour the file into a database. Like a surgeon, you wiggle your fingers, crack your knuckles and prepare to plunge your hands deep into the body of data.

Wait a moment. Drop the scalpel. Remove the gloves. No surgeon worth his license would operate without x-rays. The surgeon needs to look inside the patient and see before cutting.

And the data analyst also should visualize the data before poking around. The best applied statisticians are the ones who, for example, run clinical trials for pharmaceutical companies, advise analysts to study pictures of their data before they begin any calculations.

Too often, we start analyzing the data before sitting back and taking in the whole view. Taking in the view doesn't have to be time-consuming. And it can be even more revealing than any measurement you make.

The first step in visualization should be the histogram. The histogram is a bar chart, but the values are the frequencies of the key variable rather than the variable itself. Also, unlike the conventional bar chart, the histogram has no spaces between the bars. Here is a histogram showing full-time teachers by their salaries in the Philadelphia suburbs:

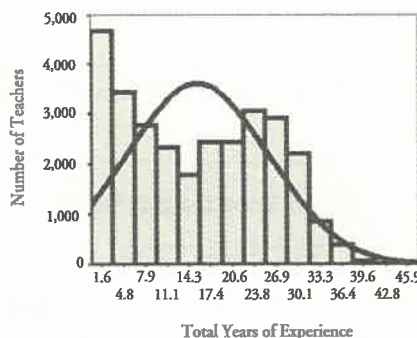


The data in the teacher-salary histogram are somewhat mound-shaped, but don't exactly follow the normal curve superimposed on the histogram. Many teachers are

paid around 52,250. However, the bars to either side of this middle bar are shorter than they would be if the data approximated the normal curve. There's a sharp fall-off to the left of the \$28,250 bar, but not as sharp a fall-off at the corresponding bar on the right side of the histogram. The \$84,250 bar sticks out and might warrant some attention.

Overall, the teacher salary data doesn't look too odd. You could design your histogram to exaggerate the oddities by adding more bars. And, conversely, by taking away bars, you mute the oddities. Generally, it makes little sense to have more than 15 bars in your histogram.

The data also don't appear to skew heavily to the right or left. A right skew, for example, would suggest there is a disproportionate number of teachers earning high salaries. In fact, when we run the usual statistics on the data, we find the mean is \$52,120 and the median is \$51,610 – not a very large difference. If the mean were much higher than the median, the data would be right-skewed.



Now take a look at another variable in the teacher salary database: total years of experience. These data aren't mound-shaped. They look almost bimodal (having two equally sized peaks). The two groups would be teachers with little experience (left mound) and teachers with a great deal of experience (right mound).

Usually the near-bimodal distribution would indicate the need for more research – and possibly a great story. Why would experience level plunge around year 14 and rise sharply in either direction?

The answer can be found in public-school enrollment totals. The data in this example are from 1997. Public school enrollment

**See Stats, page ten**

CONTACT THE IRE  
RESOURCE CENTER AT  
(573) 882-3364 OR VISIT  
[WWW.IRE.ORG/](http://WWW.IRE.ORG/)  
RESOURCECENTER TO  
SEARCH FOR TIPSHEETS

ALSO SEARCH THE  
LIBRARY OF MORE THAN  
15,000 INVESTIGATIVE  
STORIES ONLINE AT  
[WWW.IRE.ORG/](http://WWW.IRE.ORG/)  
RESOURCECENTER/  
STORYDB.HTML

THE RESOURCE CENTER'S  
WEB PAGES INCLUDE  
LINKS TO RESOURCES FOR  
SPECIFIC BEATS AS WELL  
AS INFORMATION ON  
BUYING AUDIO TAPES  
FROM CONFERENCE  
PANELS.

KEEP VISITING  
[WWW.IRE.ORG](http://WWW.IRE.ORG) FOR THE  
TIPSHEET OF THE WEEK  
AND THE STORY OF THE  
WEEK.

SEE RELATED TIPSHEET  
FROM 1998 INDIANAPOLIS  
NICAR CONFERENCE  
STATISTICS IN THE  
NEWSROOM, TIPSHEET  
# 797.

THIS TIPSHEET IS  
AVAILABLE ONLINE AT  
WWW.IRE.ORG.RESOURCECENTER

AN AUDIOTAPE OF THE  
PANEL IS ALSO AVAILABLE  
FOR \$11 FROM SOUND  
IMAGES, INC. AT  
(303) 649-1811. ASK FOR  
TAPE #40.

# Stats

*From page nine:*

since the early 1960s bottomed out in 1984, according to U.S. Education Department data. A teacher starting work in 1984 would have had about 13 years of experience by 1997. That corresponds with the valley in the histogram. Enrollment peaked in 1971. A teacher starting in 1971 would have had 26 years of experience in 1997—again agreeing with our histogram. Finally, there has been a steady increase in enrollments since 1984 (the peak is forecasted for 2006). Again, the growth in student numbers in the late 1980s and early 1990s is reflected in the left side of the experience histogram.

## Building the histogram

If you have SPSS or SAS, building a histogram can be relatively painless.

In SPSS (v. 9.0), from the pull-down menu:

“Graphs”

“Histogram...”

Set your variable (teacher salaries or experience, for example)

Click on “Display normal curve”

Click OK and SPSS will produce the histogram.

To modify your histogram, double-click on it in the SPSS output window. This will

---

**The best applied  
statisticians...  
advise analysts to study  
pictures of their data  
before they begin any  
calculations.**

---

open the chart in the SPSS chart editor.

The histogram can be edited for color, line width, etc. You also can tell SPSS to increase or reduce the number of bars. To do this, click on the value labels under the X-axis (horizontal). This opens up the Interval Axis box. The default in this box is “Automatic” intervals. Click on “Custom” to change this and to make the “Define” button available. Click on this button and the “Define Custom Interval” box will open. In “Definition,” SPSS asks for the “# of inter-

vals.” Try 10 to begin with. If too wide, try 6 or 7. If too narrow, try 11 or 12.

If your CAR shop doesn’t have SPSS, you can build a histogram in Microsoft Excel. But it’s much more cumbersome than statistical programs.

In Microsoft Excel (97), from the pull-down menu:

- Your data is on one worksheet (record limit is 65,536, based on the number of rows in the spreadsheet).

- Sort the data and, after looking at it, determine how many bars you want in the histogram. Excel can automatically set the number of bars, but the program sometimes uses too many bars. Try to keep the number under 15. In the data, look at the smallest and largest records. Without sorting, you could use =min(range) and =max(range). Subtract the smallest from the largest and divide by one less than the number of bar desired. In the teacher salary example, the smallest is 18,500 and the largest is 86,828. The difference is 68,328. If we want 10 bars, each bar would be 7,592 wide.

- If we start the first bin at 18,500, Excel will set this as the midpoint of the bar. We don’t want this. We want the lowest number (18,500) to be at the extreme left of the histogram. The midpoint of the first bar, then, would be  $18,500 + (7,592/2)$  or 22,296. The second bin would be  $22,296 + 7,592$  or 29,888 and so on. Go to another part of the worksheet. Fill in the midpoints of the bins for Excel.

- From the pull-down menu, click on “Tools” and then “Data Analysis. . .” This opens the Data Analysis box. Scroll down to “Histogram” and click OK.

- Set the “Input Range,” which is the spreadsheet range of the variable of interest, such as teacher salaries. Set the “Bin Range,” which is the spreadsheet range of the bins. Then select the “Labels” option and click OK. (Do not click the “Labels” box unless you’ve included column headers in the spreadsheet ranges you’ve highlighted).

- Excel will insert another worksheet, which will include the bin midpoints in the A column and the frequency in the B column. Use the frequency count to order a column chart from Excel. The A column (labeled “Bin”) will be the X-axis labels.

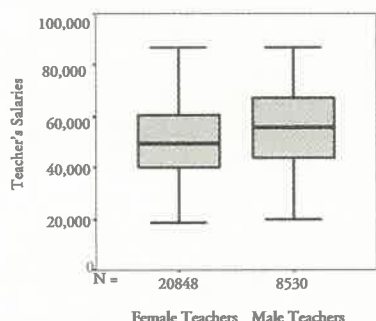
**See Stats, page eleven**



# Stats

From page ten:

• After building the column chart, notice that the bars are not touching. If you're simply trying to visualize the data, you don't have to make it a true histogram with the bars touching. If you're publishing the histogram, you'll want the bars to touch. To do so, click on the bars in the chart. Go to the Excel pull-down menu and click "Format" and "Selected Data Series..." This opens the "Format Data Series" box. Click on the "Options" tab. Set the "Gap Width" to "0." The result will be no space between the bars.



Other chart options will help to visualize the data. A good choice when comparing one variable between two groups is the box plot or "box and whisker" chart.

In the example, the variable on the X axis is whether the teacher is male or female. The Y axis is the salary level. The whiskers of the box plot show the end, or extreme, values.

**Other chart options also will help to visualize the data. A good choice when comparing on variable between two groups is the box plot or "box and whisker" chart.**

In this case, they're nearly equal for each sex. The shaded box is the interquartile range – the range between the 25th and 75th percentiles. The line through the box is the

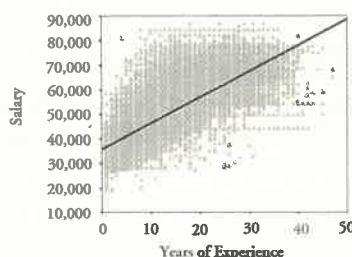
median.

At a glance, the median for males clearly is higher than for females (\$55,944 versus \$49,616). In fact, the male median is just under the top of the female interquartile range, suggesting relatively sharp pay differences. The difference between the sexes might be explained in part by the experience levels (median 22 years for males versus 13 years for females).

In SPSS, go to "Graphs" and "Boxplot" from the pull-down menu.

The process is much more complicated in MS Excel. There is an explanation of how to do a box plot at the Microsoft site (<http://support.microsoft.com/support/kb/articles/q1551130.asp>) or search for "whisker" in the Microsoft Personal Support Center.

For visualizing relationships between two variables before running a regression analysis, try the scatterplot. If we wanted to see a relationship between years of experience and salary, a scatterplot would look like this:



The data points suggest a linear relationship. The dark line is the fitted line (the line derived from the regression analysis), which you can order up as an option on the chart. In SPSS (9.0) you also can point to a datapoint on the scatterplot and click on it. It will embed the record number in the plot (that's what the 11,505 is above). The SPSS Data Editor also will automatically go to each record in question as you point.

The scatterplot in MS Excel can be ordered from the Chart Wizard as an XY (Scatter) chart. When the chart is complete, from the pulldown menu click on "Chart" and "Add Trendline ..." and a line will be fitted to the data.

Neill Borowski can be reached by e-mail at [nborowski@phillynews.com](mailto:nborowski@phillynews.com).

**UPCOMING IRE REGIONAL CONFERENCE IN LOS ANGELES, SEPTEMBER 9-12, 1999. SEE [WWW.IRE.ORG/RESOURCES/CONFERENCES.2](http://WWW.IRE.ORG/RESOURCES/CONFERENCES.2) FOR DETAILS.**

**UPCOMING CAR BOOT CAMPS JULY 18-23 AND AUG. 8-13 WILL BE HELD AT THE MISSOURI SCHOOL OF JOURNALISM TO TEACH JOURNALISTS THE BASICS OF USING DATABASES AND SPREADSHEETS. FOR MORE INFORMATION, CALL IRE/NICAR AT (573) 882-2042 OR VISIT [WWW.IRE.ORG](http://WWW.IRE.ORG).**

CHECK OUT [HTTP://ZIPINFO.COM/PRODUCTS/PLFIPS/PLFIPS.HTM](http://ZIPINFO.COM/PRODUCTS/PLFIPS/PLFIPS.HTM) IF YOU'RE INTERESTED IN BUYING A FIPS DATABASE. PRICES START AT \$29.95.

JEFF SOUTH, ALONG WITH TERESA ALLEN AND BRANT HOUSTON, PRESENTED A PANEL CALLED "FROM NEWSROOM TO CLASSROOM: TRANSLATING WHAT YOU'VE LEARNED FOR THE NEXT GENERATION" AT THE NICAR CONFERENCE IN BOSTON. YOU CAN ORDER A TAPE OF THE PANEL AT [WWW.IRE.ORG/STORE](http://WWW.IRE.ORG/STORE)

## TECH TIP

# FIPS fixes

By Jeff South

*Virginia Commonwealth University*

To you, it's Texas. But to the Census Bureau, it's "48."

You say New York's capital is Albany. Data heads call it "3601000."

The Seattle Seahawks' stadium is in King County, Wash. Geo-geekssay the Kingdome is in "53033."

Welcome to the digit-laden world of FIPS codes – the numbers the federal government uses for every state, county, city and metropolitan area. These codes are not only shorter, but they are also more precise: After all, is that South Texas county spelled La Salle, LaSalle or Lasalle?

But FIPS codes can pose problems, too. Is the code for Alaska "02" or just "2"? Is California "06" or "6" – and within that state, is Orange County "059" or simply "59"? Moreover, what's the data type for these codes – numeric or character?

Those are important questions if you're using Access, FoxPro and other tools for doing data analysis.

If one table has its FIPS codes as numeric fields and another table's codes are character strings, you can't automatically join the tables.

FIPS codes – the acronym stands for Federal Information Processing Standards – are used in datasets on a variety of subjects, from population and home loans to pollution and traffic accidents.

But they're stored in an annoying array of ways by the Census Bureau, the IRS and other government agencies.

---

**If one table has its FIPS codes as numeric fields and another table's codes are character strings, you can't automatically join the tables.**

---

Sometimes FIPS codes are numbers; other times they're text. Even when the codes are character strings, they can have different formats, such as "007" (leading zeroes) or

"7 " (a number followed by spaces or blanks) or " 7" (spaces before the number).

Here are some tips on dealing with variations in FIPS code fields.

One tip is a workaround: how to join on the fly tables that have their codes stored in different ways.

The other tip is a fix: how to add permanent fields that present FIPS codes the right way.

## What is the right way?

FIPS codes should be character fields. State-code fields should be two characters wide; county-code fields, three; metropolitan areas, four; cities, county divisions and other populated places, five.

The codes should have leading zeroes, so that Alabama is "01," Alaska "02," Arizona "04" ...

---

**The codes are stored in an annoying array of ways by the Census Bureau, the IRS and other government agencies.**

---

(Curiously, there is no "03" among the state codes, and the codes for counties are all odd numbers.)

But what do you do when a government agency provides a table with FIPS codes stored differently? I encountered this recently when I got 1996-97 county-to-county migration data from the IRS: All the FIPS codes were numeric (California was "6," and Los Angeles County was "37"). In years past, the codes were all character fields (California was "06," and Los Angeles County was "037").

To join the new and old data – that is, to make the records match – I had to persuade FoxPro to consider the number "6" the same as the character string "06" and the number "37" the same as the character string "037."

The work around: the VAL() function. VAL() converts a character string into a numeric value. In the migration case, I wanted to match my FIPS-as-text fields

**See FIPS, page thirteen**



# FIPS

From page two:

(st\_txt, cnty\_txt) in my old data with my FIPS-as-numeric fields (st\_num, cnty\_num) in my new data. So I used this as my join condition:

where new.st\_num = val(old.st\_txt) and  
new.cnty\_num = val(old.cnty\_txt)

That says: Match the numeric FIPS code with the numeric value of the character FIPS codes. The VAL() function works in Access as well as in FoxPro.

VAL() also can be used to join two text fields that have different FIPS code formats. A Virginia state agency once gave me data in which the county FIPS codes were character strings but without leading zeroes.

---

**Often you'll want to permanently change how FIPS codes are stored in your data. For instance, it will be faster and easier to join tables if your codes are consistent. More over, your codes must use the same format if you want to append records from one table into another table.**

---

I wanted to join this with Census Bureau data in which the county codes did have leading zeroes. So my join condition was:

where val(table1.fips) = val(table2.fips)

That way, "1" in the state's FIPS code field matched "001" in the Census Bureau's field, "3" matched "003," and so on.

**The fix: STR, LTRIM and PADL functions.**

Often you'll want to permanently change how FIPS codes are stored in your data. For instance, it will be faster and easier to join tables if your codes are consistent. Moreover, your codes must use the same format if you want to append records from one table into another table.

Let's say you have a table in which the state FIPS code (st\_num) and county FIPS code (cnty\_num) fields are both numeric. In FoxPro's Table Designer, insert two character fields: st\_fips, with a width of 2; and cnty\_fips, with a width of 3. Then run this command:

```
replace all st_fips with;  
padl(ltrim(str(st_num)),2,'0')
```

This takes a numeric "1" in the st\_num field and turns it into the character string "01" in the st\_fips field.

Here's how the command works, starting from the middle of the second line:

- str(st\_num)  
means convert the number into a character string.
- ltrim(str(st\_num))  
means trim all the blanks from the left side of this character string.
- padl(ltrim(str(st\_num)),2,'0')  
means pad the string with zeroes so it fills up a width of 2 characters.

You can change the numeric cnty\_num field into cnty\_fips, a character field with leading zeroes, the same way:

```
replace all cnty_fips with;  
padl(ltrim(str(cnty_num)),3,'0')
```

Use a similar command to change a text field without leading zeroes (cnty\_no0) into one with leading zeroes (cnty\_fips):

```
replace all cnty_fips with;  
padl(alltrim(cnty_no0),3,'0')
```

This says: Take the cnty\_no0 field; trim all the blanks at the beginning and end; then pad it with zeroes to a width of three characters.

## Conversions in Access

Access does not support PADL, but it has an even easier function for converting the data. If your county FIPS codes are in a numeric field (cnty\_num) and you want them in a text field with leading zeroes (cnty\_fips), update the cnty\_fips field to:

```
Format(cnty_num,"000")
```

The same function will convert a text field without leading zeroes into one with leading zeroes.

Jeff South can be reached by e-mail at [jcsouth@vcu.edu](mailto:jcsouth@vcu.edu).

**THE LATEST SET OF FIPS  
CODES FOR ALL  
GEOGRAPHIC ENTITIES  
IS AT  
[HTTP://WWW.NIST.GOV/  
ITL/DIV897/PUBS/55NEW/  
INDEX.HTM](http://www.nist.gov/ITL/DIV897/PUBS/55NEW/INDEX.HTM)**

**THE CENSUS BUREAU HAS  
DOWNLOADABLE LISTS OF  
CODES AT  
[HTTP://FTP.CENSUS.GOV/  
GEO/WWW/GAZETTEER/  
PLACES.HTML](http://ftp.census.gov/GEO/WWW/GAZETTEER/PLACES.HTML)**

# Geo-spatial sites

**By Christopher Simpson**  
American University

Simpson presented this handout at the 1999 NICAR National Conference. It suggests great web sites to get you started with satellite imagery and geospatial data.

<http://eduwww.cr.usgs.gov/webglis>

- The U.S. Geological Survey's EROS DATA Center is the government's primary unclassified intelligence imagery from the 1960s and 1970s. GLIS is the primary index to the collection and is relatively easy to use considering the vast scope of the material. Declassification of certain intelligence satellite imagery from the later 1970s and 1980s is presently underway and will most likely be available via USGS site.

<http://gcmd.gsfc.nasa.gov> and <http://www.eosdis.ornl.gov/daacpages/otherdaacs.html>

- These NASA sites are gateways to a giant, evolving effort, known as EOSDIS, to link indexes to imagery around the world. The effort is by no means complete and sometimes operates awkwardly, but the sites permit access to a wide range of unclassified U.S. government imagery and scientific data gathered from outer space.

<http://hwa.saa.noaa.gov>

- The National Oceanographic and Atmospheric Administration's index of its weather imagery provides high-quality and, often, free imagery of continent-sized parts of earth.

In addition to overhead images of hurricanes and typhoons, it provides the wide-angle imagery necessary to see the scope of forest fires throughout the northern Amazon, environmental crises in northern Africa and similar very-large scale phenomena.

<http://www.crisp.nus.edu.sg/crisp.html>

- This Singapore-based ground station provides an index of imagery collected over Southeast and much of South Asia by satellites from the United States, Canada, Europe, SPOT, Japan and other sources.

<http://isis.dlr.delservices/ISIS/ISIS-query.html>

- This index, sponsored by the German Aerospace Center, provides convenient access to imagery of Europe from several sources, as well as links to other indexes worldwide.

[http://www.eurimage.it/einet/einet\\_home.shtml](http://www.eurimage.it/einet/einet_home.shtml)

- The European Space Agency operates this user-friendly index imagery collected over Europe, much of the Middle East and the western territories of the former USSR.

<http://www.fas.org/irpl/index.html>

- Operated by the Federation of American Scientists, this index specializes in declassified "spy satellites" imagery and support for treaty-verification projects.

<http://www.spaceimaging.com/home/products/carterra/worldview/index> and <http://www.spaceimaging.com/home/products/carterra/carterra3dl/index.html>

- This is Space Imaging Corporation's Carterra imagery index, which provides access to commercial and unclassified government imagery gathered over the United States. The second URL provides the first access to "assembly line" (as distinct from custom-built) fly-through, moving satellite images that simulate the view that low-flying aircraft might see over, say, China's nuclear test range at Lop Nor.

<http://www.spot.com> and <http://www.spot.com/spot/home/sirius/welcome.htm>

- SPOT Image's commercial DALI index presently costs \$100 annually to subscribe. It provides an unusually well-organized and complete index from medium-resolution imagery (about 30 meters per pixel) for almost any point on earth.

SPOT's imagery is often more complete and more current than its competitors, though not as useful for close-up shots of the ground. The new Sirius index promises to maintain DALI's functionality, offer new services and be available for free. The Sirius product should be available this spring.

<http://terraserver.microsoft.com>

- This very large data collection combines USGS aerial photography of the continental United States with very high-resolution imagery gathered throughout the world by Russian Sovinformspутnik satellites. The USGS data can be downloaded for free; the high-quality Russian data is offered at a modest cost but is spotty in its coverage.

Chris Simpson can be reached by e-mail at [simpson@american.edu](mailto:simpson@american.edu)

SIMPSON, ALONG WITH ROBERT CHEN, PRESENTED A PANEL CALLED "SATELLITE IMAGERY ON YOUR DESKTOP: PRACTICAL PROJECTS FOR JOURNALISTS" AT THE BOSTON CONFERENCE. CONTACT THE IRE RESOURCE CENTER AT (573)882-3364.

SEE RELATED TIPSHEETS FROM 1999 NICAR CONFERENCE SATELLITE IMAGERY - ON YOUR DESKTOP, TIPSHEET #940 IMAGE INDEXES, DATA SET INDEXES, GAZETTEERS AND SATELLITE TRACKING SOFTWARE, TIPSHEET #941 THESE TIPSHEETS ARE AVAILABLE ONLINE AT [WWW.IRE.ORG/RESOURCECENTER](http://WWW.IRE.ORG/RESOURCECENTER)



# Usable Data

By Hal Straus

The Washington Post

*Straus provided this handout at the National Computer-Assisted Reporting Conference held in Boston this past March.*

## Three Ways to Build Compelling Web Databases

Web databases aren't CAR databases. And the fact that you need different software to deliver them is only one of the reasons. Here are three un-CAR things you can do to keep your work from landing on the Web with a loud thud.

### 1. Use the Right Software

- Perl and VBA/Active Server pages are good options for small and medium-sized sites. They don't work well when your volume is high (10 concurrent queries, 100,000 page views of a given feature). The problem: Forking.

- You can speed things up a lot (and minimize forking) by building a Unix system from scratch. If you're AOL or Yahoo, with lots of programmers and a "tech" culture, that's a good plan.

- If you're a newspaper or local broadcast Web site, a better solution is to go out and buy a Unix-based Web Application Server. The biggest names are Sapphire/Web and Net Dynamics. At *Washingtonpost.com*, we decided those cost too much. We bought Oracle Application Server and Netscape Application Server. From a database editor's point of view, they're about as easy to use as ASP or Cold Fusion, but they're faster.

### 2. Forget All That Relational-Database Nonsense

This is simple stuff: You're probably already doing it. But if you're not, you can really speed up your Web databases and make them a lot more pleasant for users if you break some database design rules.

- Don't use joins. Joins work fine when you're using Access to do a CAR project. They're recommended in all the relational-database books. But even if they're perfectly indexed and you're using grossly overpowered hardware, they take too long on the Web.

- Put data for different Web pages in separate tables. Then, store the tables in system memory or put each of them on different hard drives. If possible, put the

indexes your queries use on different hard drives as well. For example, if you're posting a school's database and have the test scores on a different page than the principal's name, make a different table for the scores that only has the scores and the basics that you need on that page (school name, etc.).

- Try to avoid ODBC drivers to access your database. Native drivers are faster.

### 3. Keep It Simple

Most of us in the news business got into using databases to produce stories. We're used to mining them to death to make sure that we've gotten every last bit of information out of them, then distilling the results down to understandable chunks when we write our stories. Or we are letting another reporter or an editor do it. That's great, but on the Web, you'll inevitably run into one big problem: *2.5 seconds*

Our server logs tell us that site visitors spend less than three seconds on an average interior Web page, which is where your database will end up. If you want Web users to actually explore your databases, you've got to do a few simple things:

- Design your features so users can go somewhere productive with a single click.

You can give them lots of radio buttons and text boxes and search options, but you can't count on them because 2.5 seconds isn't a lot of time.

Example of what not to do: The General Social Survey <http://gort.ucsd.edu/gss/extract.html>.

Example of what to do: The Washington Post Poll Vault <http://www.washingtonpost.com/wp-srv/politics/polls/vault/vault.htm>.

Number-heavy pages don't work any better on the Web than they work in print, even if every number is a link.

Example of what not to do: The Toxic Release Inventory Scorecard [http://scorecard.org/env-releases/us.tcl#major\\_chemical\\_releases](http://scorecard.org/env-releases/us.tcl#major_chemical_releases).

Example of what to do: The Seattle Times School Guide <http://texis.seattletimes.com/schoolguide>.

Hal Straus can be reached by e-mail at [Hal.Straus@washingtonpost.com](mailto:Hal.Straus@washingtonpost.com).

## UPCOMING HANDS-ON

### CAR TRAINING ON

### COVERING CAMPAIGN

FINANCE: JULY 23-25 IN CHICAGO. "GEARING UP FOR THE 2000 ELECTION."

CONTACT RTNDF FOR MORE INFORMATION (202) 467-5215 OR CHECK THE

INTERNET AT:

[HTTP://WWW.RTNDF.ORG/](http://www.rtndf.org/prodev/rtndf/)

PRODEV/RTNDF/

CHICAGO99/

CHICAGO99.HTM

SEPT. 17-18 IN DES

MOINES. NICAR, CENTER

FOR RESPONSIVE POLITICS,

CENTER FOR PUBLIC

INTEGRITY, MIDWEST

STATES CENTER HOST A

WEEKEND ELECTION

COVERAGE

SEMINAR FOR POLITICAL

REPORTERS IN THE

MIDWEST. CONTACT

BRAD LINT AT

515-255-8420, OR BY E-

MAIL AT

[BLINT@NETINS.NET](mailto:BLINT@NETINS.NET).

# Bits, Bytes and Barks

## NICAR data update

Updates of the following data set are now available through NICAR's Data Library:

FBI's Uniform Crime Reports (1980, 1990, 1993-97) contain detailed crime data gathered by the FBI from law enforcement agencies across the country. UCR comes in six tables:

- **The Return A database** gives the number of crimes in 27 categories, including murder, rape, robbery, assault, burglary and larceny. Crimes are reported by month and broken down into actual offenses, offenses cleared by arrest and clearances under 18.

- **Supplement to Return A** provides greater detail on property crimes. Specific types of crimes (pocket picking, gas station robbery) are given, along with amounts stolen and recovered.

- **The Age, Sex and Race database** provides male and female totals for various crimes. Race is given for juveniles and adults. The crimes are more specific, including drug offenses, gambling, DUIs, runaways and others.

- **The Supplemental Homicide Report.** For each homicide, the data includes age, sex and race of offenders and victims, as well as their relationship and the circumstances of the crime. It also includes justifiable homicides by police and citizens.

- **Arson** is set up like the Return A and Return A Supplement databases. Information includes fires in uninhabited structures and property damage estimates.

- **The Police database** provides information about assaults and murders of police officers.

**Strengths:** FBI Uniform Crime Reports are the only standardized source of detailed, agency-by-agency crime statistics for the whole United States. The reports don't identify criminals or victims, but they do provide interesting information about them, including age, sex, race, and sometimes crime circumstances. Do the numbers for your area look suspicious? Interesting stories have been written about discrepancies between what really happens and what gets reported to the FBI. (See "Counting Crime Right?" story, page 5).

**Weaknesses:** All UCR data is self-reported; if a law-enforcement agency wants to fudge the numbers, it's easy to do. The FBI has integrity checks in place, but it mostly takes agencies at their word. Even the checks can cause problems: If the FBI deems an agency's (or entire state's) reports unreliable, it may delete them without telling anybody. Because different jurisdictions have different classifications for some crimes, these crimes are sometimes left out of the database without notation (i.e. Minnesota was listed as having no rapes for some years). So it is, important to check values against local numbers.

Also coming this summer, HAZMAT (hazardous material spills), OSCAR (nursing home inspections), Federal Procurements (federal contracts), USDA (meat-packing plant inspections) and more.

To order, call NICAR at (573) 884-7332.

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

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