

Uplink

June 1998

A newsletter of the National Institute for Computer-Assisted Reporting

CFIC UPDATE

Moneyed interests

By Jack Dolan and
Aaron Rothenburger
CFIC

A direct link between campaign contributions and favoritism is the Holy Grail of campaign finance reporting.

In the 1997 series "Money and Influence," (<http://www.campaignfinance.org/stories/state.html>) *Milwaukee Journal Sentinel* reporters Steve Schultze and Daniel Bice connected campaign contributions to preferential treatment by Wisconsin Governor Tommy Thompson.

The two reporters examined records from

eight state agencies, state contracts, campaign contributions and expenditures, and Thompson's phone records. The eight-month investigation revealed a startling correlation between those who gave money to Thompson's campaign and those who got lucrative state contracts.

In one instance, Schultze and Bice found that Ron Van Den Heuvel, a Wisconsin businessman, donated \$10,000 to Thompson's campaign fund the day before a state agency awarded him \$24 million in tax-free bond financing to build a paper-manufacturing plant. Van Den Heuvel says a Thompson fund raiser asked him for the donation. The newspaper's inquiry forced Thompson to return the contribution.

Schultze and Bice say the Van Den Heuvel affair was not an isolated incident. "The [Van Den Heuvel] case and others reviewed by *The Journal Sentinel* in an eight-month investigation suggest a trend in which donors and well-connected firms enjoy a close and mutually beneficial relationship with the Thompson administration," they wrote.

You can investigate that claim for yourself by downloading current Wisconsin campaign finance data from the CFIC or by searching the Wisconsin Democracy Campaign Web site at <http://www.wisdc.org/>.

Creating diversions

Another strength of using campaign finance data is the ability to show how big money flows around the sometimes useless legal breakwaters that are meant to stem the tide.

Darrel Rowland of *The Columbus Dispatch* found that non-cash loans and in-kind contributions to parties or legislative campaign committees are exempt from state limits in Ohio. (The series can be accessed online

Continued on page five

Update

As the campaigning season draws near and the financing for it revs up, so, too, will be investigations into contribution tactics and motivations. Jack Dolan and Aaron Rothenburger of the CFIC describe recent campaign finance stories that use the type of data available from the CFIC Data Center.

Neill Borowski of *The Philadelphia Inquirer* contributes a statistics column showing how to conduct more thorough analyses of trends in home prices.

Wendell Cochran of American University dissects the possible stumbling blocks facing CAR specialists when year 2000 arrives. Cochran surveys the differing designs of hardware and software products to handle the turnover.

Jennifer LaFleur of the *San Jose Mercury News* discusses recent maneuvers in the ever-changing battlefield for data access.

In addition, Diane Renzulli of the Center for Public Integrity focuses on campaign contributions. The handouts offer advice on constructing databases from scratch.

PAGE 8

Tech Tsp

PAGE 12

On the Internet

PAGE 14

Handouts of the Month

STATS FROM PHILLY

Homes for sale

By Neill A. Borowski
The Philadelphia Inquirer

Your editor mumbles "real estate story."
You immediately head for the door.

Many reporters have little or no interest in writing about homes. Yet relevant news about the typical family's most valuable investment can make for a front-page story that has impact. This doesn't mean the filler material that tends to populate many newspaper real estate sections. The emphasis is on *relevant*—stories that are local and tell readers something they want to know.

And what can be more relevant than a story that explains trends in values of their home?

First, let's consider the role of the home in the United States, according to the American Housing Survey for the United States in 1995 (<http://www.census.gov/prod/2/constr/h150/h15095rv.pdf>). The house is the central asset, the biggest debt and the greatest

Continued on page six

Uplink

June 1998
Volume 10, Number 5
A newsletter of the National
Institute for Computer-Assisted
Reporting

EDITOR

Brant Houston
brant@nicar.org

MANAGING EDITOR

Brent Johnson
bjohnson@nicar.org

SENIOR CONTRIBUTING EDITORS

Len Bruzzese
len@ire.org
Sarah Cohen
sarah@nicar.org
Richard Mullins
richard@nicar.org

COPY EDITOR

Jeanine M. Davis

ART DIRECTOR

Wendy Charron

STAFF

Jack Dolan
Seth Hemmelgarn
Ben Lesser
Justin Mayo
Ted Peterson
Noemi Ramirez
Aaron Springer
Alistair White

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

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

Postmaster: Please send address
changes to NICAR.

Send e-mail to 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.

MILLENNIUM AD NAUSEUM

Trial of the century

By Wendell Cochran

American University

Nanosecond by nanosecond, the clock ticks toward the millennium and the biggest beta test in computer history – the rollover from 1999 to the year 2000. By some accounts, that first day of the new century is going to bring a silicon meltdown because millions of computers, literally, won't know what day it is.

Stock exchanges, banks, airlines, government agencies and major corporations are spending billions to make their systems compliant. The Information Technology Association of America estimates it will cost more than \$50 billion to fix computers in the United States, perhaps \$600 billion worldwide. A few economists have predicted that the costs of fixing computer hardware and software are so high that the U.S. economy will be thrown into a recession.

Some companies and governmental bodies, apparently hoping the problem will just go away, have barely begun to figure out what the Y2K issue means to their operations. A member of Congress recently reported that some "mission-critical" Pentagon systems won't be fixed until 2009.

Thankfully, those of us who work in computer-assisted journalism probably don't need to panic. Most of our hardware is relatively new; most of the applications we use regularly have tried to address the Year 2000 problem at least after a fashion; and most of our databases and spreadsheets don't depend on precise date calculations.

That doesn't mean CAR specialists are off the Y2K hook. A program that mishandles a date calculation, query or sort might well produce an error that is published or broadcast.

InfoWorld, a leading computer industry publication, cautioned last year against being too complacent. "The year 2000 will affect almost every machine on your network to some degree," the paper said. And nearly every computer expert warns against counting on hardware and software companies to diagnose and fix all the problems.

So you should understand just what Y2K issues you might face in your CAR shop. It would be a good idea to test your computers, to experiment with how your favorite software deals with dates such as 04/15/00, and

to review the structure and content of the date fields in your databases. You might also want to contact the folks who regularly supply you with data to see what changes they foresee making.

Bill Loving, computer-assisted reporting editor formerly at the *Minneapolis Star-Tribune*, assessed his newsroom's inventory of software and databases for Y2K risks but wasn't too concerned based on what he found. "My understanding is that all the major commercial apps that we use in CAR will be, or already have been, repaired by the vendors. I don't know yet what the database implications are."

Tom Boyer at the *Seattle Times* also thinks his computers and programs are in good shape for the next millennium. He says the problem will "probably affect some of the databases we acquire. But they're often so dirty anyway – this'll just be one thing more we have to clean up."

Do your computers recognize the Year 2000?

The first step is to determine whether you have hardware problems. Generally speaking, if you're using a Macintosh, you won't have any problems. Apple claims its operating system can handle dates from 30,081 B.C. to 29,940 A.D.

Most software applications take their time and date formats directly from the system, where time/date is kept at the BIOS level. BIOS stands for Basic Input Output System, a set of instructions automatically loaded into a computer before the operating system. It allows the computer to interact with the keyboard, ports, and other hardware devices. In other words, if the system clock doesn't properly adjust for the new century, your software probably won't either.

In general, the newer the system, the better the chance that your PCs won't choke on Jan. 1, 2000. But even Intel says, "Although, in general, the system BIOS revisions provided with Intel desktop motherboards and server baseboards manufactured since 1992 contain services capable of transitioning to the year 2000, this BIOS capability alone does not ensure that the system will handle the transition without error. The behavior of any given system dur-

Continued on page three

Continued from page two: **Compliant systems**

ing the year 2000 rollover is dependent on several factors alone or in combination, including, but not limited to, the system configuration, BIOS, operating system, and application software."

Translation: A lot of stuff could go wrong. What's more, many clones don't use Intel motherboards.

Different computer manufacturers report different levels of Y2K compliance. Dell systems made after October 1996 are fully compliant, the company says, and it has upgrades for older BIOSes. Gateway 2000 says its Pentium-based models work properly. 486 models will be fine until they are rebooted after the new century. Once they're rebooted, the system clock will need to be set again.

If you don't feel comfortable conducting your own test, several vendors are selling Year 2000 test programs. But, in case something goes wrong, don't test the computer with your most important databases until you've backed them up. And you're probably better off disconnecting from the network. Check network software and components later. (Expect problems to show up, especially if you do such things as force periodic password renewals.)

To test your computers, boot in DOS. At the C:\> prompt, type "date" and then "12/31/99". Also at the C:\> prompt, type "time" and set it for 11:48:00 p.m. Turn the computer off and wait long enough for the system to rollover. Reboot and check the system date.

If it says 01/01/2000, you are one step closer to home.

If the computer doesn't seem to know the date—for example, if it displays Jan. 4, 1980 or Jan. 1, 1900—you've got a problem. Contact the manufacturer of your machine to determine whether an updated BIOS is available. If it's an older model, you might want to scrap the computer.

Applications: Where the real issues lie

Hardware problems might turn out to be the easiest ones to handle. Software and data are likely to be tougher for many CAR newsrooms. At the least, you'll want to understand how your programs handle dates. Dates are stored as numeric values based on a fixed starting date that varies according to the program. Date fields can be formatted and displayed in a variety of ways, with the most common default being 00/00/00.

Most Year 2000 problems could have been prevented if date fields contained four-digit years. Few do. Two-digit years were designed into most applications and databases as a way to preserve scarce memory and disk space in the 1960s and 1970s, when most computer standards were set.

When did time begin? For personal computers, the short answer is Jan. 1, 1980. You cannot set most PC system clocks to an earlier day. Fortunately, most applications can deal with earlier dates by using a system known as Date Serial. Dates are assigned numeric values based on a fixed starting

Continued on page four

Date handling in popular CAR applications

PROGRAM	START DATE	END DATE	TWO-DIGIT YEAR DEFAULTS
Excel 97	Jan. 1, 1900	Dec. 31, 9999	00-29=2000 30-99=1900
Excel 98 for Macintosh	Jan. 2, 1904	Dec. 31, 9999	
Access for Windows 97		Dec. 31, 9999	Same as Excel
FoxPro		Dec. 31, 9999	
Quattro Pro 8	Jan. 1, 1600	Dec. 31, 9999	00-50=2000 51-99=1900
Paradox 8	9999 BC	Dec. 31, 9999	00-50=2000 51-99=1900
Lotus 123	Jan. 1, 1900		
Lotus Approach			

Sources: Microsoft, Corel, Lotus. Note: Earlier versions of these programs handle dates differently. For example, some earlier versions of Excel won't accept dates past 2078. Check your documentation.

TWO WEB SITES TO CHECK OUT FOR YEAR 2000 COMPUTER ISSUES INCLUDE:

- WWW.Y2K.COM
- WWW.YEAR2000.COM

THESE ISSUES HAVE FOSTERED LENGTHY DISCUSSIONS ON THE IRE AND NICAR MAILING LISTS. TO SEARCH THE ARCHIVES OF THESE MAILING LISTS BY AUTHOR, SUBJECT OR DATE, POINT YOUR BROWSER TO THE FOLLOWING WEB ADDRESSES.

- **IRE-L:**
WWW.IRE.ORG/RESOURCES/IRE-L/INDEX.HTML
- **NICAR-L:**
WWW.IRE.ORG/RESOURCES/NICAR/NICARL.HTML

**WENDELL COCHRAN
PROVIDES THE FOLLOWING
WEB SITES OF COMPUTER
COMPANIES ADDRESSING
YEAR 2000 COMPUTER
ISSUES:**

- **MICROSOFT:**
[WWW.MICROSOFT.COM/
CIO/ARTICLES/
YEAR2000FAQ.HTM](http://WWW.MICROSOFT.COM/CIO/ARTICLES/YEAR2000FAQ.HTM)
- **COREL:**
[WWW.COREL.COM/
2000.HTM](http://WWW.COREL.COM/2000.HTM)
- **LOTUS:**
WWW.LOTUS.COM
- **INTEL:** [HTTP://
SUPPORT.INTEL.COM/
YEAR2000/PAPER.HTM](http://HTTP://SUPPORT.INTEL.COM/YEAR2000/PAPER.HTM)
- **IBM:** [WWW.IBM.COM/
IBM/YEAR2000](http://WWW.IBM.COM/IBM/YEAR2000)
- **APPLE:**
[WWW.APPLE.COM/MACOS/
INFO/2000.HTML](http://WWW.APPLE.COM/MACOS/INFO/2000.HTML)
- **COMPAQ:**
[WWW.COMPAQ.COM/
YEAR2000](http://WWW.COMPAQ.COM/YEAR2000)
- **HEWLETT-PACKARD:**
[WWW.HP.COM/YEAR2000/
INDEX.HTML](http://WWW.HP.COM/YEAR2000/INDEX.HTML)
- **DELL:** [WWW.DELL.COM/
YEAR2000](http://WWW.DELL.COM/YEAR2000)
- **GATEWAY2000:**
[WWW.GATEWAY.COM/
HOME/SUPPORT/
C%5FTECHDOCS/Y2K/
DEFAULT.HTM](http://WWW.GATEWAY.COM/HOME/SUPPORT/C%5FTECHDOCS/Y2K/DEFAULT.HTM)
- **NOVELL:**
WWW.NOVELL.COM

Continued from page three:

Matters of time

point: Day 1=1, day 10=10, etc. This system permits date-formatted cells and fields to be used in calculations.

For spreadsheets, Jan. 1, 1900 seems to be the most common starting point for time. Lotus1-2-3 set this standard, and Excel uses it. An earlier date is not recognized as a date, will not be formatted as a date and cannot be used in a date calculation. QuattroPro, on the other hand, can handle dates back to Jan. 1, 1600, but day 1 is Dec. 31, 1899. And for Mac Excel users, time begins Jan. 2, 1904.

Steve Doig of Arizona State University points out that the Mac-PC difference means dates can be wrong by four years if you swap data between the platforms. If you move data containing dates between spreadsheet types, different versions of the same program or platforms, you might want to reconstruct your calculation formulas rather than importing them.

The concept of having a beginning for time also means you need to have an end. For you doomsday types, time ends on Dec. 31, 9999, according to Excel 97 and Mac Excel 98, but in 2078 for earlier versions of Excel. If you're using two-digit years in QuattroPro, the latest you can enter is Dec. 31, 2050. The latest four-digit year is 3199.

Database managers are a little more flexible. Access 97 can store and calculate dates in the range of Jan. 1, 1100, to Dec. 31, 9999. Paradox 8 can handle four-digit years from 9999 B.C. to 9999 A.D.; its two-digit range is Jan. 1, 1951 to Dec. 31, 2050.

Digit difficulty

The real test for date compatibility, however, comes in databases that have dates from multiple centuries, with years formatted as two digits. For example, the Federal Election Commission candidate table now lists some politicians running in 2000 and later, but it only uses two-digit years. The next election cycle covers 1999-2000, so the dates on contributions will be mixed between the 20th and 21st centuries.

For now, this doesn't cause great problems. As the table illustrates, most programs we use in CAR assume that two-digit years early in the century should be interpreted as coming after 2000. For example, in Excel and Access, all two-digit years less than 30 are assumed to be

in the 21st century.

Says Tom Torok of the *Philadelphia Inquirer*: "More of a pain are the different defaults some database programs use. For example, Access 7.0 and Access 97 assume any two-digit year that is less than or equal to 29 is in the 21st century ... Something that was converted from Access 2.0 to Access 7.0 to SQL might give you any two-digit year between 29 and 48 in the 21st century."

Knowing how your programs handle two-digit years might solve some data problems, but it would be better to create and display true four-digit years. Depending on the structure of your tables and the nature of your data, that can be fairly easy or somewhat complex. In any case, it will be easier in a database manager than in a spreadsheet.

Creating a four-digit year involving only one century is simple. Reformat your table to make the year field a four-character text field and add a date field formatted mm/dd/yyyy.

In Access, use an update query to add the desired century string to the existing year string. Then use another update query to put the combined contents of the month, day, and year fields, formatted as a date, into a true date field that displays all four digits.

In FoxPro, this would be a Replace query; in Paradox, use a Changeto query. In some cases, though, it might be safer to write a script or function to explicitly add the century to the year string. The key is to develop a reliable way to identify which date range belongs in which century.

Doig cautions that CAR specialists will need to be especially diligent when they get data from older systems. Stress the need to be vigilant about data coming from "legacy" systems. "Even if your machine is Y2K compliant, it doesn't mean that the data you're getting will be. The 'does-this-answer-make-sense?' test will be especially important to apply during the next couple of years."

The real solution to date-handling problems is to scrap the notion of formatting dates with two-digit years and create a date data type that only permits four-digit years. That will require changes in operating systems and in almost all applications.

Wendell Cochran can be reached at wcochran@ibm.net

Tales from the vault

Continued from page one:

at <http://www.campaignfinance.org/stories/divert.html>.) According to *The Dispatch*, the biggest beneficiary so far of this loophole is Cleveland area state Senator Robert A. Gardner. State Republicans funneled more than \$900,000 into Gardner's last campaign, or more than 90 percent of Gardner's total receipts. A sudden, last-minute \$384,000 advertising campaign paid for by the state party and "loaned" to Gardner may have been the deciding factor in Gardner's narrow victory. The loan was especially generous since there is no time line for Gardner to repay it, no state law that prohibits the party from simply forgiving it, and no requirement to report who repays it if it is in fact repaid.

Rowland also reports that a new headquarters for the Ohio Republican Party was "custom-designed around the state's campaign finance laws." The building was financed by a party trust fund that allowed corporations to exceed normal contribution limits. Rowland calls the new Republican headquarters "a veritable factory for in-kind contributions" since the TV and radio ads, computer services, and direct mailings that originate in the new headquarters count as in-kind contributions to party candidates.

Rowland's series also details how the Democratic and Republican national parties used Ohio's less restrictive campaign finance laws to their benefit. In 1996, the Democratic National Committee directed out-of-state contributors to send unsolicited contributions to the Ohio Democratic Party. The money is then transferred, or even sold back, to the national party. The biggest single contributor to the Ohio Democratic party in 1992 was Indonesian businessman James T. Riady, who gave \$75,000 to the state party six days before President Clinton's election.

The latest Ohio campaign finance data can be searched online at the Ohio Open Elections Project on the Web site of the Center for Responsive Politics (<http://www.crp.org/ooep/>).

Resisting choreography

When the bureaucrats in charge of computerizing state campaign finance records settle into their customary glacial rhythm, reporters have two choices: cover well-re-

hearsed sound bytes and photo opportunities while waiting for an information thaw, or start a fire by building their own campaign finance database.

A consortium of 29 newspapers and two TV stations in New York lit their own fire under local politicians this year when they set aside traditional rivalries to join forces and turn over 10,000 paper records into the most comprehensive campaign finance database in New York history.

The result, according to *Newsday's* Ford Fessenden, was "a blizzard of stories" putting state politicians on notice that the public is interested in their campaign finances. A telling example of the impact: A couple of years ago, a *Newsday* reporter noticed that the paper contribution reports from Governor Pataki's office came sorted alphabetically by the contributor's first name. Suspecting a computer was used to generate that order, the reporter called Gov. Pataki's press secretary, Zenia Mucha, to ask for the governor's campaign contributions in electronic form rather than on paper. Mucha laughed at the reporter. Today, after hundreds of stories on state campaign finances generated by consortium members, Pataki files contribution reports electronically.

CFIC offerings

The CFIC library, located online at www.campaignfinance.org/stories/index.html, is an archive of local, state and federal articles that use campaign contribution reports to "follow the money." The library also serves as a complement to the CFIC's online databases of state campaign finance contributions.

The three investigative series described above are just a few of the excellent examples of how campaign finance data can enhance state and local political reporting.

If you have a relevant article or series to submit to the CFIC's growing repository of stories, or if you know of an article or series that should be submitted, contact Jack Dolan at jack@nicar.org or Noemi Ramirez at noemi@nicar.org, or call (573) 884-1802.

Jack Dolan can be reached by email at jack@nicar.org. Aaron Rothenburger can be reached by email at aaron@nicar.org

SEE PAGE 15 FOR TIPS ON CREATING CAMPAIGN CONTRIBUTION DATABASES FROM PAPER RECORDS. THE TIPS ARE WRITTEN BY DIANE RENZULLI OF THE CENTER FOR PUBLIC INTEGRITY.

THE WEB SITE FOR THE CAMPAIGN FINANCE INFORMATION CENTER (WWW.CAMPAIGNFINANCE.ORG) IS A VALUABLE RESOURCE FOR ANYONE

INVESTIGATING CAMPAIGN FINANCES. IT INCLUDES:

- **DOWNLOADABLE DATABASES OF CONTRIBUTION DATA AND LINKS TO SEARCH ENGINES MAINTAINED BY NON-PROFIT ORGANIZATIONS AND STATE BOARDS OF ELECTION**

- **STORIES FROM PRIOR ISSUES OF TRACKER**

- **STORIES EXPLORING CAMPAIGN FINANCE AT THE LOCAL, STATE AND FEDERAL LEVELS**

- **INFORMATION ON THE CFIC-L MAILING LIST**

- **TIPSHEETS ON COVERING CAMPAIGN FINANCE**

- **LINKS TO POTENTIAL SOURCES FOR CAMPAIGN FINANCE STORIES AND OTHER RELEVANT SITES**

THE "GUIDE TO HOME PRICES" STORY CAN BE ACCESSED ONLINE AT WWW3.PHILLYNEWS.COM/ PACKAGES/HOMES/OVER29.ASP

THE SECTION'S GRAPHIC ARTIST/DESIGNER, MATTHEW ERICSON, TURNED THE DISTRIBUTION INTO A COMBINED BAR CHART, WITH FIVE SHADES OF GRAY SO READERS COULD SEE AT A GLANCE HOW PRICEY HOMES IN THEIR COMMUNITIES WERE. THESE DISTRIBUTIONS BY COUNTY AND CITY ARE AVAILABLE AT THE ABOVE WEB SITE.

Continued from page one:

Housing analysis

monthly expense for countless Americans:

- Out of 106 million year-round housing units in the nation, 60 percent are owner-occupied. The median value of owner-occupied homes is \$92,507.
- About six out of 10 homes have one mortgage outstanding. Seven percent have two or more mortgages and 13 percent have a home-equity loan. The median number of years to repay the mortgage is 20, with median principal of \$48,466 to go. The median monthly cost is \$593.
- The median ratio of home value to income is 2:3. However, the home value for about 25 percent of owners is four times their income or more.

Perceptions—and misperceptions—about trends in real estate prices can influence consumer spending and the economy.

At *The Inquirer*, I did my first home price survey in 1990, looking at 1989 prices. The data was cobbled together through the cooperation of different real estate associations. I did another one in 1991 and, in later years, other reporters carried on the tradition. In the last couple of years, though, I suggested that we forego the survey until we could get better data. We got that data this year and were able to offer a more balanced look at home values in the region.

Our survey could be done anywhere. The data may be readily available in some areas and tougher to collect in others.

The old (and flawed) way

In the past, we used median values by municipality in Pennsylvania and average values by municipality in New Jersey. The data were based on analysis of statistics furnished to the states by each county. We presented the latest year's median (or average) sale price, the previous year's price and the percentage change. In addition, we showed how many units were sold one year compared with the previous year.

However, we were uncomfortable with what the data showed. There was no way to adjust for the type of housing sold each year.

If a large number of starter homes or townhouses were sold in one year, it could drive down the median or average compared with the year before. Using our method, we

would tell the reader that home prices dropped by "x" percent in that municipality. But this wasn't necessarily the case. The bottom didn't fall out of the market. Values might actually have been stronger. We were comparing two different sets of houses. The problem was most severe when looking at municipalities where there were only a handful of sales; however, the problem can be apparent even in larger areas.

Consider Montgomery County, a wealthy suburban county near Philadelphia. In 1996, the average price of a home in the county was \$168,392. In 1997, the average price was \$170,048 — an increase of 1 percent. However, the median price (the middle price when all prices are ranked) was down 2.5 percent. How could the data be so skewed? Part of the reason was an increase in the number of extremely high-priced properties sold in 1997 versus 1996. In 1997, the average of the top five prices was \$2.2 million, compared with \$1.7 million in 1996. Neither the median nor the average truly represented the change in home values in the county — and even they didn't agree.

The new (and improved) way

Last year, we contracted with Realist Inc., a Philadelphia-based real estate statistics company, to provide *The Inquirer* with several databases. One of those databases for the Pennsylvania suburbs was the sale price of every home sold, the sale date, the previous sale price and previous sale date. Realist came through with data for 1997 and 1996.

This allowed us to do a "matched pairs" or "sales pairs" analysis. The full database also allowed us to offer the reader other statistics about each municipality. Here's what we ran and how it was done:

Median Prices. They're still of interest. But we successfully resisted the temptation to calculate the percentage change between years. That was the flaw we were trying to avoid. We gave the reader the median for 1987, 1996 and 1997 and indexed all suburban median prices for 1997 to show how each municipality's median compared with the overall regional median.

Price Appreciation. This was the mea-

Continued on page seven

Continued from page six:

Revamped medians

sure we'd been after for years – and the one that can be done only if you have “matched pairs.” This answered the question: “What is happening to housing values when you can look at the values of specific properties?” The database provided us with matched pairs for an average of six out of 10 properties sold in 1997. The formula for appreciation is the compound average annual rate of growth formula:

$$\left(\left(\frac{1997 \text{ Price}}{\text{Former Price}} \right)^{\frac{1}{t}} \right) - 1$$

Where,

- the 1997 Price is just that for the given property
- the Former Price is the last available price for the given property
- t represents the number of years, found by subtracting the Former Price's sale date from the 1997 date and dividing by 365. Carry this to several decimal places.

For example, the 1997 price was \$105,000 on Jan. 3, 1997. The Former Price was \$52,500 on Aug. 1, 1983. The number of days is 4,904 and years is 13.4356. Plugging this into the equation above:

$$\left(\left(\frac{105,000}{52,500} \right)^{\frac{1}{13.4356}} \right) - 1 = 0.0529$$

This means the appreciation rate for this property was more than 5 percent a year – not a bad showing compared to most of the properties. The median appreciation rate for all of Montgomery County was 1 percent a year. This analysis guided us to the lede of the main story of the section:

The residential real estate market in the Philadelphia area last year underscored the reality of home ownership in the '90s: Think shelter, not profits.

In fact, a low-interest passbook bank account offered more promise of profit than a home bought in the '90s and sold last year, according to an Inquirer analysis of more than 57,000 real estate sales in the eight-county metropolitan area.

In Excel, the appreciation formula would be: $=((105000/52500)^(1/13.4356)) - 1$

In Access, the “Update to” order in the update query would be:

```
( ([ MONTCO] ![ SALE_AMT] /  
[ MONTCO] ![ SALE_PREV] ) ^ (1/  
[ MONTCO] ![ YEARS] ) ) -1
```

where the Sale_Amt field is the 1997 sale price, the Sale_Prev field is the former sale price and the Years field is the number of years between sales.

After the appreciation rate for each property is calculated, find the median appreciation rate for the municipality.

We also indexed appreciation rates in the region, with the median for the region equaling 100. Because you also have the 1997 price and the former price, you can calculate which homes were sold at a loss. We calculated this as a percentage of total homes sold to show the percent sold at a loss. In some communities, half of the properties for which we knew the previous sale price sold at a loss. Homes with the highest probability of being sold at a loss in 1997 were those purchased in 1989 and 1990, according to our database.

Price Distribution. The luxury of having every sale price allowed us to find price distributions in each municipality. The distributions we used were \$0 to \$50,000, \$50,000 to \$100,000, \$100,000 to \$150,000, \$150,000 to \$225,000 and \$225,000 and up.

Housing Market Measures. The “turn-over rate” for each municipality we used was the number of homes sold in 1997 as a percentage of total homes in the municipality. We also showed the median number of years owned for each municipality as well as the estimated 1997 median household income (from Claritas Inc.).

The 12-page section, called the Guide to Home Prices: What's selling – and why, was well-received. The day it appeared (March 29), our lead real estate writer was a guest on the radio show of a well-respected real estate authority. The radio host praised the section, calling it the “bible” of local real estate prices. He urged readers to go out and purchase a Sunday Inquirer if they hadn't already done so.

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

NICAR IS COLLECTING STORIES THAT DEPENDED IN PART ON STATISTICAL TECHNIQUES – ANYTHING FROM SIMPLE CROSSTABS TO AVERAGES VERSUS MEDIANS TO REGRESSIONS. SARAH COHEN IS PARTICULARLY INTERESTED IN STORIES IN WHICH THE NUMBERS NEVER WENT INTO THE PAPER OR ON THE AIR BUT WERE USED TO HELP FOCUS THE REPORTING.

SEND CLIPS TO SARAH COHEN AT IRE, 138 NEFF ANNEX, COLUMBIA, MO 65211 OR E-MAIL HER AT SARAH@NICAR.ORG WITH THE APPROPRIATE LINK.

IF YOU HAVE A PARTICULAR PROBLEM YOU WOULD LIKE TO SEE ADDRESSED IN THIS COLUMN, SEND AN EMAIL TO BRENT JOHNSON AT BJOHNSON@NICAR.ORG

SIGN UP TO BE AN IRE MEMBER. YOU CAN JOIN ONLINE AT OUR WEB SITE OR BY DOWNLOADING A MEMBERSHIP FORM AND FAXING IT TO (573) 882-5431.

FOR MORE INFORMATION, VISIT WWW.IRE.ORG/MEMBERSHIP.HTML

IRE MEMBERSHIPS ARE \$40 FOR PROFESSIONAL JOURNALISTS AND JOURNALISM EDUCATORS, \$25 FOR STUDENTS AND \$55 FOR INTERNATIONAL JOURNALISTS. FOR A DETAILED ACCOUNT OF WHAT IRE HAS TO OFFER, CONTACT THE IRE MAIN OFFICE AT (573) 882-2042.

TECH TIP

A more perfect union

By Andrew Lehren
Dateline NBC

For FoxPro 2.6 fans wondering whether to jump to Visual Fox, here's one reason: you can handle time without any fancy programming.

That's because Visual Fox includes a date time data type. It's separate from the old date data type, which remains in the newer versions. And along with the new category come some handy functions. Here's an overview with tips to get things done faster and avoid mistakes.

Converting your data

Let's assume you want to convert data stored as character. You can create a new date time field by taking your mouse to Table... Properties... and modifying the structure. After inserting a new column, go to data type and pick date time. That's a choice you didn't have in earlier versions of Fox. Now go back to your command window. To bring in data from a character column, you'll need to use the character-to-time function - ctot (). Let's look at how it works.

Let's say you're starting with three separate character columns, one with dates, another with times, and a third designating a.m. or p.m. Pretend this is part of a database showing the times that city officials made telephone calls:

CALLDATE	CALLTIME	AMorPM
12/31/95	12:45	AM
09/14/96	7:35	PM
01/10/97	7:40	AM
10/8/96	11:35	AM

(Notice the data is a little inconsistent. Sometimes you've got a leading zero in your day or time. Sometimes you don't. It doesn't matter - Fox will cut you slack. You also won't need to trim any leading spaces.) In this example, we'll pretend you've created a new date time column called CALLTIME.

Now we're ready to use the ctot () function. Have no fear. It should be familiar territory to anyone who's messed with Fox functions. To fill in the times in the new column, run this command:

```
Replace all CALLTIME with ;
ctot (CALLDATE + CALLTIME +
AMorPM)
```

The result should look something like this:

```
12/31/95 12:45:00 AM
09/14/96 07:35:00 PM
01/10/97 07:40:00 AM
10/08/96 11:35:00 PM
```

And you're done converting your data.

Missing or bizarre entries

If, for some reason your data includes a time that cannot exist, like 25:00, the replacement for that record won't work. Nothing will be inserted. The same goes for an invalid date, even if you have a valid time. Nothing will be inserted.

However, if you have a valid date and nothing in your time field, Fox will insert the date plus 00:00:00, which is midnight. And, if you have a valid time but no date, Fox will insert 12/30/1899. (If you forgot how to see what century it is, simply type in your command window: SET CENTURY ON.) If these seem like boring details, keep in mind that this can mean the difference between a query that gets a right or wrong answer.

How do I look?

The default setting shows time with seconds. If you don't like this - your data doesn't have seconds and you don't like looking at all those zeros - type SET SECONDS OFF in your command window. Away go all those pesky :00s.

In addition, you may find it easier to ask some questions by using military time. That's where the clock goes all the way to 24:00 and you no longer need a.m. or p.m. To do that, type SET HOURS TO 24. Now everything will be in military time. Then again, if you hate seeing things in military time, the opposite command is SET HOURS TO 12.

A query sampler

Let's say you want to find all calls after 2 p.m. on Sept. 12, 1996, when a big piece of legislation was introduced:

Continued on page nine

Continued from page eight:

Time for a date

```
Select * ;
From PhoneCalls ;
Where CALLTIME > { 9/14/96 14:00}
```

Here you need to use the curly braces, the same way you needed them for the original date data type.

Let's say you want to find all calls between 2 p.m. on Sept. 12, 1996, and 10 a.m. on Jan. 15, 1997, when the big piece of legislation finally came up for a vote:

```
Select * ;
From PhoneCalls ;
```

```
Where between (CALLTIME, {9/12/96
14:00}, {1/15/97 10:00})
```

Again, you need to use the curly braces.

Now let's say you want to find out who's so important that the mayor is calling after 10:00 at night. If you have things set to military time, the query is simply:

```
Select * ;
From PhoneCalls ;
Where hour (CALLTIME) > 22
```

Notice that when you use the hour () function, the criteria is numeric. You don't need any punctuation, like quotes or curly braces.

Time differences

Let's say you have the starting time of a phone call and the ending time, and you want to know how long a city council member called that 900 number. Remember that the date time data type answers with the number of seconds – even if you've set seconds off. So if you want your answer in seconds, don't do a thing. If you want it in minutes, divide by 60. In hours, by 3600. And so on. For example:

```
Select *, (ENDCALL - BEGINCALL) /
60 as CallLength ;
From PhoneCalls
```

This will tell you the call's length, in minutes. Notice that you need to put the EndCall – BeginCall in parentheses. The reason that must be done is to first generate

the difference a result that will be numeric and can then be divided by 60.

Let's say that, for some reason, you have one column as a date field, and another as a date time field and you want to find the difference between them. Let's say it's an arrest database. The date field is Birthday, and the date time field is Arrested, noting when someone was busted. For this mixing and matching, you can use the ttod() function. Here's an example. (This time, dividing converts our answer into years, to see how old they were when arrested):

```
Select *, (ttod(Arrested) -
Birthday)/365.25 as ArrestAge ;
From Crimes
```

Catching nulls

Remember in the section on missing and bizarre entries? Let's say you want to find an instance when no data is in the field. There are two ways to do that: with the isblank () or empty () functions. The isnull () function will not work because Fox has yet to allow true null designations for these cases. So here's an example to catch them:

```
Select * ;
From PhoneCalls;
Where empty (EndCall)
```

And, since you know what Fox inserts when you have a valid time but no date, or vice versa, you can run filters to catch or ignore those entries.

Watch out! A caution about averages

Remember how Fox will convert proper days with no time (by adding midnight) or no days with proper time (by assuming the day is 12/30/1899)? Well, don't forget that can screw up any summary calculations. For some reason, Fox will consider empty entries when running averages unless you tell it otherwise. Regardless, be careful to filter for these kinds of values if you don't want to include them.

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

UPCOMING BOOT CAMPS:

**JULY 12-17, 1998 –
BASIC BOOT CAMP IN
COLUMBIA, Mo.**

**AUGUST 9-14, 1998 –
BASIC BOOT CAMP
IN COLUMBIA, Mo.**

**AUGUST 14-16, 1998 –
ADVANCED BOOT CAMP
IN COLUMBIA, Mo.**

**FOR A COMPLETE LISTING
OF UPCOMING IRE AND
NICAR EVENTS,
INCLUDING BOOT CAMPS,
ON-THE-ROAD TRAINING,
SEMINARS, AND REGIONAL
CONFERENCES, VISIT OUR
WEB SITE AT
[WWW.IRE.ORG/
CALENDAR.HTML](http://WWW.IRE.ORG/CALENDAR.HTML)**

Battling for access

By Jennifer LaFleur

San Jose Mercury News

— *Jennifer assumed that the press needs access to government databases... seems a little cocky.*

I recently spoke at a conference for the organization for federal FOIA Officers (called ASAP). Although my evaluations varied, the above comment stood out for me:

The battle for records access is heating up. It is no longer a given that government records are easily available. The hurdles are many, but three key problems that are more frequent and of grave concern include:

- Privatization of records: Government agencies are more frequently having private companies maintain their databases. It makes access more difficult and muddies the definitions of what is owned by the public. Sometimes government agencies themselves have problems getting data in these situations.

- Escalating costs for records, particularly electronic records, which prevent or greatly delay a news organization's ability to gain access to records.

- Violating public records laws in the name of "protecting the privacy of citizens." In many states, the courts have flipped public records laws on their heads. They make it the burden of news organizations to show public interest over privacy, which was not the intention of those laws.

Although there is a lot of hype about how the Internet is causing everyone's privacy to be invaded (just like cable television was when I was younger), the public doesn't always agree with government's protection. At a recent *Mercury News* brown bag with readers on a totally different topic, readers spoke up about how they want their newspaper to get access to government records. They even volunteered to get a packet of letters together in support of public records access. If not for ourselves, we owe it to our readers to demand access to government records.

Is it our fault?

Rather than dwell on the usual ways government agencies bamboozle reporters into thinking they shouldn't have access to data, let's think about some of the things we do or don't do when it comes to public records.

- Cover citizen's access battles. We don't do

very many stories about how citizen's can't get records or how one government agency can't get information from another.

- Cover how candidates stand on public records. How many know how your local candidates stand on public records? Did your legislators vote for the Drivers Privacy Protection Act?

- Do we follow legislation that affects open records? Do a search of legislation in your state to find those affecting your state's open records act — you might be surprised.

- Provide information for readers about how to get information. Once a year the American Society of Newspaper Editors does a project called "Your Right to Know" where they provide information to newspapers (and encourage papers to do their own work) on open records. What else could you do?

- Do reporters follow through with requests and make agencies accountable? Or do they get tired of asking and move on to something else? We are all guilty of giving up some fights — but that sends a message to government agencies that they've "won." If you don't have legal help at your newspaper, and you've tried every means yourself — contact IRE's access committee or your local SPJ chapter.

- Duplicate the Indiana Project. Seven newspapers in Indiana did an audit of government records in their state. Search the Indiana CAR handouts at www.ire.org for more information on this project.

Guidelines to request records

1. Know the law. Know how FOIA/open records law treats the information you're requesting and what the exemptions are. If there are exemptions, the agency must be able to cite them.

2. Know what information you want. Don't ask an agency to provide you everything they have. Make your request specific. It's likely that a reporter may need the actual data that was used for a report. Before asking for the data, request the record layout for that database. Documentation is available under the law.

3. Know how the information is kept. Try to find someone in the information systems department at an agency who knows. This is

Continued on page eleven

NICAR'S DATABASE LIBRARY CAN CONVERT DATA FROM ALL TYPES OF GOVERNMENT SYSTEMS. THE STAFF CAN TAKE NINE-TRACK TAPES, 4MM DAT TAPES, 3480 CARTRIDGES, PRINT IMAGE FILES OR TR1/TR2 TAPES, AND PUT THEM ONTO CD-ROM IN THE DATABASE FORMAT OF YOUR CHOICE. NICAR STAFF CLEANS AND CONVERTS THE DATA, AND PERFORMS INTEGRITY CHECKS. PRICES ARE BASED ON THE SIZE OF YOUR NEWS ORGANIZATION. PLEASE CALL THE DATABASE LIBRARY AT 573-882-0684 FOR MORE INFORMATION.

Continued from page ten:

Custodian plays

not to avoid the public relations person you're dealing with, but sometimes it's necessary to talk nerd-to-nerd. Reporters using computer databases are actually quite computer-savvy.

4. Know what the appropriate cost should be. You really should have to pay only duplication costs. Ask for an itemization for those costs. This is an area that gets difficult when an agency has a set price for databases.

5. Know the source of the data – be familiar with how the records are created.

6. Know who administers the data. The person in charge of the database can be helpful for understanding the data.

7. Get hard copy summary reports. This will give you a way to check your data.

8. Know how many records or pieces of information are in the database. When you get the database, make sure you have the right number of records.

9. Know how large the database is. You don't have to know a lot of technical jargon. Most PCs will hold at least 250 megabytes of information or roughly 250 million characters. If they say the file is 900 megabytes, you won't be able to do much with it unless you find a computer with a bigger hard drive. Most newsrooms have the capability of reading nine-track tapes from mainframes and many other data formats.

10. Confidential information. A government agency may claim that certain pieces of the information are confidential. There very well may be confidential information in the file, but that doesn't mean you shouldn't get the rest of it.

11. Get involved with local group meetings of software users. These groups bring together folks from all different disciplines who work together to solve problems and share tips.

Non-viable denial

The previous guidelines are all dandy, until you actually go to ask for the data. Many times you'll run into reasons why an agency can't give you information, probably because 8 out of 10 reporters will just say "oh, okay" and go away. Here is a sample of reasons I have received for why an agency cannot fill my request:

• *Listen Missy (yes, this really happened): Our database is very complicated. You probably won't understand it.* Most newsrooms have the

capacity to handle very large and complex databases. If you're not dealing with someone who knows the data, you need to.

• *Our computer system can't do that.* This is unusual for any database system. It may take talking with the software vendor to figure out the best option for both sides.

• *The person that knows how to do that is on vacation for two weeks/doesn't work here anymore.* This is a problem.

• *It will cost you \$20,000.* Ask for an itemized estimate of charges. Remember that the list of other recipients of the data and what those recipients paid is available under FOIA.

• *The database is not public record.* The burden is on the agency to show where in the law that information is not public record. If you have done your homework, this should be less of a problem.

• *We don't like what you plan to do with it.* Interesting. On the other hand, when you're able to cooperate with an agency, they frequently can help point you in the right direction.

• *The database contains confidential information.* Everyone needs to do their homework on this one. Redacting with a computer is possible just like on paper with a black marker.

• *We don't keep that on computer.* Although this is true in some cases with smaller jurisdictions or small offices federal organizations, it's uncommon.

• *If we give it to you, we'll have to give it to everyone.* So what's the problem with this.

• *That uses proprietary software.* You don't want software, you want the data. This may require getting together with the vendor.

• *We don't mind giving you a few records, we just don't want to give you the whole database.* To which the judge asked: You mean if they wanted *one* record from *one* person it would be okay? Their answer: depends on the person.

Visit these relevant Web sites:

• American Society of Newspaper Editors "Your Right to Know" project: www.asne.org/ideas/rtk/rtkmain.htm

• Freedom of Information Center: www.missouri.edu/~foiwww

• Reporters Committee for Freedom of the Press: www.rcfp.org
Jennifer LaFleur can be reached by e-mail at jlafleur@sjmercury.com

NICAR'S DATABASE LIBRARY HAS RECENTLY UPDATED SEVERAL DATABASES: THE INS LEGAL RESIDENCY DATABASE FOR 1996, THE HAZARDOUS MATERIALS DATABASE FOR 1997, AND THE LATEST QUARTERLY UPDATE FOR IRS EXEMPT ORGANIZATIONS.

THE LIBRARY CAN BE REACHED AT (573) 884-7332.

A DOWNLOADABLE ORDER FORM IS AVAILABLE AT WWW.NICAR.ORG/DATA

Beyond plain Boolean

By Debbie Wolfe

St. Petersburg Times

Searching the Web with pizzazz and aplomb requires going beyond (yawn) plain Boolean.

By some estimates, the Web contains more than 150 million pages—some 50-60 billion words. (Let's not get too picky over the "exact" size. The point is the Web is *huge* and growing exponentially. It's time to learn advanced searching tips and tricks and to keep current with searching advancements.)

If you're still doing searches only with plain "and" and "or" connectors/operators, you must *love* to plow through junk hits. Since search engines and directories offer some of the same advanced Boolean logic options available in Nexis, Dialog and DowJones, why not take advantage to save time and frustration?

Serendipity has its place in any searcher's toolkit, but when you're on deadline—or just not in the mood to plow—try focusing your search with supercharging techniques.

Sample supercharges

- *Making sure "and" means AND!*

Add a "+" (plus) sign in front of each word you absolutely want to include in the results:

+pet+care finds both
pet+care finds "care" and may find "pet"

- *Proximity: "phrase" and "near" searches*

Add quote marks around words you want to search as a phrase:

"pet care" looks for both as a phrase
"+pet+care" finds both as a phrase
pet near care looks for "pet" within a certain number of words of "care"

- *Order can make a difference*

Change the order of your search terms to adjust the weight assigned to each part: breeders and title:"maltese" may find different results than title:"maltese" and breeders

Test your searching savvy

- Alta Vista (<http://www.altavista.digital.com>)

I'll either use this or Yahoo (<http://www.yahoo.com>) to start a Web search quest. The advanced search page "Help" area reveals one of the most comprehensive tip sheets out there. Run the same query on the simple and advanced pages and you'll get different results. Search words are automatically weighted on the simple page—you're in charge on the advanced page. Caution: If you don't specify a word to sort by, the advanced results will be in no particular order.

- FindLaw — Internet Legal Resources (<http://www.findlaw.com>)

• DejaNews (<http://www.dejanews.com>)
The best place to search Usenet newsgroup postings, including complete threads.

- California State University at Stanislaus (<http://www.library.csustan.edu>)

Links to search engines maintained by librarians.

- Lycos (<http://www.lycos.com>)
Search for pictures or sound files through the drop-down menu. Pro search help area has a link to "Web Guides." Choose "Education" and browse the "Reference" link to locate electronic favorites.

- HotBot (<http://www.hotbot.com>)
The form-based interface looks simple to

Continued on page thirteen

WHERE TO DO

SUPERCHARGED SEARCHES:

- ALTA VISTA ([HTTP://WWW.ALTAVISTA.DIGITAL.COM](http://WWW.ALTAVISTA.DIGITAL.COM))
- HOTBOT ([HTTP://WWW.HOTBOT.COM](http://WWW.HOTBOT.COM))
- INFOSEEK ([HTTP://WWW.INFOSEEK.COM](http://WWW.INFOSEEK.COM))
- WEBCRAWLER ([HTTP://WWW.WEBCRAWLER.COM](http://WWW.WEBCRAWLER.COM))

FOR THE BEST AND MOST EXACT RESULTS, READ THE HELP AREAS AND STUDY THE SAMPLE SEARCHES AT EACH SITE.

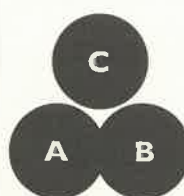
PLAIN BOOLEAN

AND, OR

SUPERCHARGED BOOLEAN

ADJACENCY, PROXIMITY, NOT, RELEVANCY (WEIGHT VIA WORD ORDER, SYNONYMS, SORTING), META WORDS (FIELDS OR SEGMENTS)

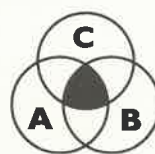
Boolean Logic Primer



OR

broadens the results (results must contain at least one item)*

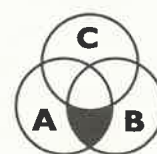
A or B or C
hurricane or surge or damage



AND

narrows the results (results must contain all items)

A and B and C
hurricane and surge and damage



NOT

narrows the results (results must contain A & B but no C's)

(A and B) not C
(hurricane and surge) not damage

*An "OR" search would look for 'hurricane' or 'surge' or 'damage' or any combination of the three words. In the illustration above, the results indicate every possible result except the combination of all three words.

Continued from page twelve: **Set your sites**

use, but spend some quality time checking out the options since documentation is lacking. The "Browse by Category" area has a unique cascading interface. Check out the "Reference" link to see the effect.

- InfoSeek (<http://www.infoseek.com>)

The tip sheet is a fast scan, the interface is simple, and the results come back at rocket speed.

- The Virtual Reference Desk (<http://thorplus.lib.purdue.edu/reference/index.html>)

I know this is not a search engine or directory, but I couldn't help myself. This page is so wonderful, it will convince any jaded journalist that the Web is a tool that belongs at everyone's desk!

- Magellan (<http://www.mckinley.com>)

Simple to use. What's best about this site is the "Find Similar" link on the results page. This feature helps you modify your original search with one click. Reviews are also included.

- Excite (<http://www.excite.com>)

Once you get a results page, the "List by Web site" feature is a handy way to look at the top 40 hits as multiple links from the same site address are grouped together for easy evaluation. The "More Like This" link on the results page is a quick way to modify a strong link for further results.

- WebCrawler (<http://www.webcrawler.com>)

Fast and simple to use. The "Similar Pages" feature on the summary results page helps to focus the next search. Choose between "titles" or "summaries" for display options. Don't have enough time to check out all the links? Save the page in HTML format and explore them later!

- Reference.com (<http://www.reference.com>)

Metasearch sites

- Internet Sleuth (<http://www.isleuth.com>)

Allows you to search one-six engines/directories at a time and completes the task in seconds, but the results display takes time. The interface also likes the lowest-common-denominator search strategy, which frequently limits what you are looking for to plain Boolean. The service indexes only links to what it considers "searchable databases."

- All4One (<http://www.all4one.com>)

Search four engines simultaneously: AltaVista, Lycos, HotBot and Excite. Each engine is displayed in a frame taking up a quarter of the screen.

- Highway61 (<http://www.highway61.com>)

This is a down-to-earth site, and when you get there, you'll see why. It searches six engines at once: Excite, Infoseek, Lycos, WebCrawler and Yahoo. The display intermingles the results and ranks each hit on a scale of 100. Response time is fast.

- MetaCrawler (<http://www.metacrawler.com>)

Search six engines at once: Excite, Lycos, Yahoo, WebCrawler, AltaVista and Infoseek. The display is intermingled, and the results are ranked by confidence level on a scale of 1,000. Check out the "MetaSpy" link for a fun, red-light district of what other folks are searching for. Response time is extremely fast.

- ProFusion (<http://www.designlab.ukans.edu/profusion>)

Search nine sites, including Excite, AltaVista, WebCrawler, Lycos, Magellan, Infoseek and Yahoo. You can ask the site to pick the best three, the fastest three, or all nine. The interface is very thorough and user-friendly. Ranked results are displayed by title and URL only.

- Dogpile (<http://www.dogpile.com>)

Quickly search 13 Web engines and directories, four Usenet newsgroup sources, and 2 FTP archives including Filez (<http://www.filez.com>). Results are displayed ordered by the engine where they were found. *This is my top pick for metasearch sites.*

- Verio MetaSearch, previously known as OnRamp (<http://search.verio.net>)

Search eight sites including Excite, HotBot, Infoseek, Lycos, WebCrawler and Yahoo. Be sure to read "About" and "Origins," then go to the advanced search area and experiment with the weight options—very cool!

- BONUS SITE: Search Engine Watch (<http://searchenginewatch.com>)

This is your homework!

Debbie Wolfe can be reached by email at dpwolfe@sptimes.com

**TO STEER YOUR CAR
PROJECTS IN THE RIGHT
DIRECTION, ORDER
"COMPUTER-ASSISTED
REPORTING: A PRACTICAL
GUIDE" BY BRANT
HOUSTON. IT CAN BE
ORDERED FROM NICAR
AND INVESTIGATIVE
REPORTERS AND EDITORS
FOR \$26 FOR IRE MEMBERS
OR \$30 FOR NON-MEMBERS
PLUS SHIPPING.
CALL (573) 882-2042.**

**FOR A LISTING OF OTHER
TITLES AVAILABLE
CONCERNING COMPUTER-
ASSISTED AND
INVESTIGATIVE REPORTING
TECHNIQUES, VISIT OUR
WEB SITE AT
WWW.IRE.ORG/
PUBLICATIONS/
BOOKSTORE**

Assembly required

THESE HANDOUTS WERE PROVIDED AT THE IRE NATIONAL CONFERENCE IN NEW ORLEANS.

TO SEARCH THE COLLECTION OF CONFERENCE HANDOUTS AVAILABLE FROM THE IRE RESOURCE CENTER, VISIT OUR WEB SITE AT WWW.IRE.ORG/RESOURCES/CENTER/HANDSEARCH.HTML

TO SEARCH THE LIBRARY OF MORE THAN 12,000 INVESTIGATIVE STORIES AVAILABLE, VISIT WWW.IRE.ORG/RESOURCES/CENTER/SEARCH.HTML

THE RESOURCE CENTER CAN BE REACHED AT (573) 882-3364 OR BY EMAIL AT RESCNTR@IRE.ORG

By Jeffrey Meitrodt, *Times-Picayune*
Carolyn Tuft, *St. Louis Post Dispatch*
David Washburn, *The Morning Call*
"The Ten Commandments of Building Your Own Database"

1. Use the computer as a source.

The most compelling stories can be put together with readily available information that merely needs to be organized and examined. Such records include standardized test scores available to every parent, traffic tickets, lawsuit records, public contracts, campaign finance reports, city/county school budgets, personnel files, public official voting records, billing records, check files.

2. No amount of data is too small an amount.

How big does a story have to be before you plug data into a spreadsheet or database? The answer: "smaller than you'd think." Reporters love to say, "We analyzed a gazillion records and found..." That is well and good, but you can add meat to almost any story — and improve your chances of landing on Page One — by plugging a dozen records into a spreadsheet.

3. Don't be discouraged if you have to do it the hard way.

Sure, it's great when there is a ready-made database waiting to be analyzed for investigative stories and a cooperative public official willing to share it. That's not always the case. Besides, if the agency keeps records on a database, the target is already ahead of you. Chances are, if you're the first person who ever took a look at the numbers, there's a good story to tell. If the agency can't analyze its own data, it has no idea what's really going on.

4. Think about the best way to get data before you go and get it.

A low profile works best when you're taking on a new target. You don't want to give them time to circle the wagons and take action that would kill your story. So, find out if they release the data to the general public in some fashion and then get those reports acting as Joe Citizen. It's more work, but no one will know what you are up to.

5. Organize your data before you analyze it.

Think about the questions you want your data to answer before you start working on your spreadsheet, and then customize the columns to fit your needs. Think in terms of yes/no questions. Find ways to create numbers that will allow you to rank your data.

6. The "sort function" is your friend.

The sort function may sound boring, but it isn't. It is the most powerful key on the computer when it comes to CAR. Do you want to find out who made the biggest campaign contributions? Or which school had the largest test-score swings? By sorting the data, you'll often find your most powerful anecdotes.

7. Keep your data consistent.

If you're doing a column of people's names, always put the last name first. Don't use 9 and nine. Don't do anything that will make it harder to rank or sort data or to find data that matches.

8. Double-check the data.

You need to be bullet proof. Mistakes are deadly. They give your target a chance to discredit your research.

9. Do it yourself whenever possible.

Sometimes you have no choice about who inputs the data: there is simply too much to handle by yourself or with another reporter. But whenever possible, type in the data yourself. Two advantages: you see stuff you wouldn't have found otherwise, and you eliminate the errors that other people with much less invested in the story are going to make.

10. Buy yourself enough time.

Editors think CAR is easy. It's your job to make sure they know you're going to spend two weeks doing nothing but punching numbers into a spreadsheet. And then maybe another week checking your work and playing with the data before you're ready to make a single phone call. Tell them up front. You need to sell them on WHY there's no other way to do it.

Paper riches

By Diane Renzulli

Center for Public Integrity

"Tips for Creating Campaign Databases
from Paper Records"

Start small

- Don't take on the whole state legislature on the first try. State legislators in leadership positions, committee chairs, or candidates from local races are much more manageable subsets.

Staffing

- More people does not mean better data quality.
- If you can, avoid in-house data entry for major projects.
- If you can't, test potential data-entry personnel for attention to detail before hiring them.

Gathering the data

- Make friends with the staff at the local elections board.
- Create an index of the candidates you want to include in your database.
- Keep copies of the records on site.

Data entry

- Use a double-entry system.
- Give each contribution a unique identification number that allows you to quickly retrieve it in paper form.
- Enter the name of the contributor exactly as it was written on the campaign records, even if it is incorrect.

Editing the data

- Keep both an "edited contributor" and an "original contributor" field in your database.
- To identify which acronym belongs to which organization, sort your database by address.
- Never guess. If a contributor has two different addresses, confirm that it is the same person or organization.
- If family members give from the same address, you can combine them by using an "and/or" notation in the "edited contributor" field.

Coding contributions

- Code contributions AFTER the database has been edited.
- Use the Center for Responsive Politics coding system.

- Avoid identifying organizations by name only.
- Create a field on the database that lists your source for the coding.
- After coding, sort by industry code and check for mistakes.

Identifying individuals

- Prioritize. Identify top donors first.
- Always confirm that an individual at two different addresses is the same person. Call them directly if you need to.
- If your state or locality doesn't require employer or occupation information, you can match the following with your database:
 - lists of professionals regulated by state agencies
 - information from a CD-ROM telephone directory (for example, ProCD)
 - local companies' board of directors and officers
 - local companies' contributors to state and federal PACs
 - your newspaper's press files
- Sort your database by address to find individuals who work at a particular organization.
- Create a field in the database for the initials of the staff person who identified the individual.
- If you're in a state or locality that mandates employer or occupation information by law, contact the campaign for more information on unidentified donors.

Fact checking

- Check your database with the candidates' cover pages. If the totals for a candidate don't match your database, check each contribution individually.
- Treat your database suspiciously before writing your story: sort, sort, and sort some more.
- If you are using a campaign database provided by the state, first check it with paper records to ensure its accuracy.

Final note

- While you should do everything you can to ensure your database's accuracy, don't let the idea of the "flawless" database control you or your project.

Diane Renzulli can be reached at
(202) 783-3900 or by email at
drenzulli@publicintegrity.org

THE WEB SITE FOR THE
CENTER FOR PUBLIC
INTEGRITY IS

WWW.PUBLICINTEGRITY.ORG

ONLINE CAR PROJECTS:

TO VIEW A LISTING OF

LINKS OF RECENT

COMPUTER-ASSISTED

REPORTING STORIES

POSTED ON THE WEB,

POINT YOUR INTERNET

BROWSER TO

WWW.IRE.ORG/RESOURCES/

CONFERENCES/TRAINING/

CARPROJECTS.HTML

THE SITE INCLUDES A

DESCRIPTION OF THE

STORIES AS WELL AS LINKS

TO IRE AWARD WINNERS.

IF YOU WOULD LIKE TO SEE

A STORY ADDED TO THE

LIST OF LINKS, SEND JACK

DOLAN AN E-MAIL AT

JACK@NICAR.ORG.

Bits, Bytes and Barks

Indiana CAR Audio Tapes

Audio tapes from the sessions at Indiana CAR are now available for ordering. For more information or an order form, point your browser to www.ire.org/resources/nicar/conferences/indianal/audio.html

Audio tapes from the IRE National Conference in New Orleans are also available from Sound Images, which can be reached at (303) 649-1811 or by email at sitapes@aol.com

Campaign Finance Information Center

Campaign finance data from thirteen states is freely downloadable from the CFIC at www.campaignfinance.org. We also have links to twelve online search engines hosted by other non-profits and state boards of election. We will soon release our "universal" online search engine so you can type in a contributor from your state and see where else they are giving.

The more inclusive this database, the better for everyone. So, if you have state or local campaign data you want included, contact CFIC Coordinator Jack Dolan at jack@nicar.org or (573) 884-1802. The CFIC credits everyone who contributes data.

Subscribe Online

You can now subscribe to *Uplink* online. Point your browser to www.ire.org/resources/nicar/uplink

Moving On

Bill Loving recently moved from the *Minneapolis Star Tribune* to the *Los Angeles Times*. He can be reached by email at bill.loving@latimes.com. If you have switched locales, please let us know. Contact Brent Johnson at bjohnson@nicar.org.

New Deputy Director at IRE

Len Bruzzese has joined IRE as its new deputy director. Bruzzese comes to IRE after serving as the editor of *The Olympian*, an award-winning state capital daily in Olympia, Wash. Bruzzese will oversee IRE's publications, its World Wide Web site, its resource center and the Campaign Finance Information Center. He also will assist in the administration of IRE and its conferences.

International Investigative Reporting

A new \$20,000 award recognizing the best international investigative reporting in the world has been announced by the Center for Public Integrity. The award, part of the Center's new International Consortium of Investigative Journalists (ICIJ), will go to a journalist or team of journalists who have demonstrated excellence in transnational investigative reporting during the previous year.

ENTRIES MUST BE POSTMARKED BY AUG. 15. For more information on the award, including how to apply, visit the ICIJ Web site at www.icij.org or call (202) 783-3900.

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