July 1998

Uplink

A newsletter of the National Institute for Computer-Assisted Reporting

NURSING HOME NEGLECT

Unraveled care

By Mark J. Rochester
The Indianapolis Star and
The Indianapolis News

Normally, a newspaper pursues an investigative story with the hope that circumstances revealed by the probe eventually will change as a result of the public scrutiny brought by the story's publication.

So when *The Indianapolis Star* and *The Indianapolis News* concluded a year-long investigation of nursing homes with publication of a six-day series in late June, we were confident that the disturbing picture of elder care presented would generate an important public policy debate and could possibly lead to changes in the system.

What we *never* expected was that the series would result in significant changes to the system within *days* of its publication. Or that some of the state's most powerful legislators and governor would demand reforms and an end to the conditions described by the newspapers.

But that's exactly what happened immediately following publication of "The Tarnished Years," an investigative series in which computer-assisted reporting helped *The Star* and *The News* expose the pervasiveness of poor care in area nursing facilities and shortcomings in the state's nursing home regulatory system. The vows of reform announced by state officials demonstrate how CAR techniques, aggressive reporting and powerful writing can combine as effective catalysts for public policy change.

During its investigation, the newspapers documented disturbing accounts of neglect, broken bones, exploitation and other instances of poor care in Indianapolis area facilities. The series revealed that one-third of area nursing homes allowed harm to befall one or more seniors in their care. And when poor care is detected, the state makes little use of its own penalty system and instead relies on a federal system that gives facilities weeks or months to make corrections before sanctions are imposed. Neither system provides much incentive to stay in compliance with regulations. In fact, nearly nine out of 10 nursing homes statewide were not in compliance with federal regulations, the newspapers found.

Closing time

The investigation began simply enough. Two veteran journalists, Bonnie Harris and Joe Fahy, were curious about the same thing at the

Continued on page two

SPREADING HAZARDS

Toxic two-time

By Duff Wilson
The Seattle Times

It was the revolution in e-mail and World Wide Web publishing – more than the newsprint of *The Seattle Times* or wires of Associated Press — that gave sufficient power and reach to my series of articles on toxic wastes to be named a finalist for the 1997 Pulitzer Prize for Public Service.

While I wrote "Fear in the Fields – How hazardous wastes become fertilizer," it could easily have been ignored in the rest of America. After all, the story ran in a remote, rainy corner of the country during the July 4th weekend. It lacked the immediate national media follow-up that has attended more obvious environmental problems.

Heavy industries were saving millions by disposing of hazardous wastes as fertilizer instead of putting them in a landfill or recycling them in a safer manner.

Metal smelters, pulp and paper mills, cement kilns and their allied fertilizer

Continued on page four

Inside Uplink

The summer's almost over, but apparently neglect, deception and monetary troubles are heating up. The topics may get blood boiling, but at least the CAR work isn't cooling off.

Mark Rochester of *The Indianapolis Star and The Indianapolis News* recounts how one nursing home's closing triggered a statewide look at patterns of neglect and poor enforcement. "The Tarnished Years" series included an online searchable database of nursing home inspection data and prompted immediate calls for reform.

Based on his Pulitzer finalist work in "Fear in the Fields," Duff Wilson of *The Seattle Times* explains, among other online and spreadsheet tips, the value of maintaing e-mail databases of targets, sources and activists involved with an ongoing series of stories.

And Valerie Lilley of the *Peoria Journal Star* analyzes the CAR contributions to a week-long series explaining the cause of rising bankruptcies in a booming economy.

Uplink

July 1998
Volume 10, Number 6
A newsletter of the National Institute for Computer-Assisted Reporting

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.

From page one:

Disrespecting elders

same time. About a year earlier, the state had closed down a nursing home that had established a hideous care record – some residents had developed bedsores so serious that limbs had to be amputated. The reporters wondered about conditions in other nursing homes.

We knew this was going to be a difficult story to pin down. Documentation would be vital for the story to have any real impact.

The vows of reform announced by state officials demonstrate how CAR techniques, aggressive reporting and powerful writing can combine as effective catalysts for public policy change.

The first step was having Harris and Mark Nichols, our CAR coordinator, request a database of all annual state inspections of Indiana nursing homes from the federal Health Care Financing Administration.

That was no simple task. As with any database request, the reporter and CAR specialist had to immerse themselves in HCFA rules and regulations to know exactly what they were asking for. There were several conversations with HCFA PIOs to determine exactly how records were kept; in what formats they were available; what information was contained in each individual record; how the records were coded; and whether the records accurately reflected situations the newspapers were trying to analyze, i.e., the quality of care in each nursing home.

Data warning

A bit of forewarning here.

The original HCFA data was requested in late 1996. We didn't get it until spring 1997. At that point, Nichols, using AskSam, began sorting and cleaning the files. But the files we obtained were "flat" files – essentially electronic printouts of the files as they appeared on HCFA's network. In terms of creating relational database files, it was as if we were starting from scratch. Once the files were cleaned up, they were imported into Paradox for Windows.

Warning No. 2. The federal government revised regulations for nursing homes in mid-1995, so inspection records before then cannot be compared to those conducted afterward. Not every home had had an annual state inspection recorded by HCFA under the new rules, according to our database. To solve that problem, we went after more data.

But rather than going back to the federal government, we decided to request the same records from Indiana's health department, which happened to maintain its own set of inspection records. Indiana law requires government agencies that keep records on computer to provide them if they exist. In addition, the health department cooperated with our request after we explained our interest in the records. We obtained the new data within a matter of weeks.

Data interrogation

After a second round of cleaning and sorting, Nichols conducted numerous queries, guided in part by questions Harris and Fahy expected to ask during their reporting. Again, using Paradox, queries were run to chase after questions like: How many nursing homes had violations indicating they were substantially out of compliance with federal regulations? How many violations indicated residents were physically harmed? Did for-profit homes have more violations than nonprofits? Did urban facilities have better or worse records than suburban and rural facilities?

The answers helped steer the reporters in terms of looking at homes that consistently had quality of care problems. With the computer data in hand, they descended upon the records room at the health department and pulled the paper records on dozens of nursing homes – with both good and bad performance records – where the horrific accounts of poor nursing care were to be found.

Making public information truly public

Admittedly, we're not the first publication to examine the phenomenon of shock-

Continued on page three

From page two:

Online informant

ingly bad nursing home care, but the series had several unique twists that make it a rather unique effort. The series consumed more than 30 full news pages during publication, including a full-page nursing home "Report Card" that listed a "ranking" based on the number and severity of federal violations, a listing of the homes' worst violations and other information.

Of particular interest is the newspapers' online searchable database of nursing home inspection data.We presented it in a format that allows the public to conduct a number of information searches and get everything – not just what was published – the newspapers learned about each nursing home.

Of particular interest is the newspapers' online searchable database of nursing home inspection data. Working with our online department, we combined the various data used during the investigation – record analysis, personnel records, profitability studies, regression analysis. We presented it in a format that allows the public to conduct a number of information searches and get everything – not just what was published – the newspapers learned about each nursing home.

For example, readers can perform a search request for nursing homes ranked by the newspapers as "poor." For each home, they receive its name, address, ownership, number of violations, a description of each violation, the amount of administrative turnover, costs per patient day, and whether the home had a history of substandard quality of care.

Readers can call up information on homes in their neighborhood by typing in an address, zip code or city. The Web site also allows them to get maps and directions to nursing homes.

Attitude adjustment

During the investigation, the newspapers had asked state officials why such important consumer information was not being made available to the public. They responded that it wasn't a priority.

But, barely more than a week after the June 21-26 publication of the series, state officials announced that they would make an array of comparable consumer information on nursing homes available online soon. And that it would be updated *daily*.

Health officials plan to take a more aggressive stance against the state's worst nursing homes by threatening their licenses, banning new admissions and imposing other penalties. Top law enforcement officials, including the state's attorney general, also said they want to see a tougher posture in rooting out poor care.

Reforms by the Department of Health are a direct response to issues raised in the series, officials admitted. The newspapers' investigation "did stimulate us to refocus, to re-examine our programs here in the Department of Health," Indiana Health Commissioner Richard Feldman said in an interview after the series. "We're going to try to refocus, reallocate and make the best use of available personnel and resources to go after and pursue and litigate the worst repeat offenders."

Gov. Frank O'Bannon and state lawmakers also are endorsing change.

In fact, one of the most powerful members of the Indiana legislature vowed that, if health officials and the governor don't fix problems in state nursing homes, the legislature will. That's a powerful statement coming from the Republican leader of the Indiana Senate. The nursing home industry pumps significant amounts of campaign cash into their coffers each election cycle.

Such swift and resounding moves toward reform serve as effective reminders of the power we as journalists have to bring about change.

Mark J. Rochester can be reached at
1-800-669-7827 ext. 9288 or
mrochester@starnews.com

THOSE INTERESTED IN THE "TARNISHED YEARS" STORIES CAN GO DIRECTLY TO A SERIES DIRECTORY:

- **HTTP://WWW.STARNEWS.COM/NEWS/SPECIAL/NURSING.HTML**

TO GO DIRECTLY TO THE SEARCHABLE DATABASE:

- **HTTP://SEARCH.STARNEWS.COM/DB/NURSING.QRY**

From page one:

Fooling farmers

companies would advertise the minor plant foods zinc, iron and lime in their wastes. But they wouldn't tell the ultimate buyers about the arsenic, cadmium, lead or dioxins.

Nobody was regulating, testing or disclosing the tag-along toxics. Nobody knew what level was safe. Some experts said the metals build up in soil, plants and animals. Industry controlled the research.

Industries were shifting the risk from their wastes onto farmers and gardeners. The industries saved $100 per ton. The unknowing farmers might save dimes per acre. When the mayor of Quincy, Wash., raised questions, the local industry tried to force her to shut up.

Knowing you're going to initiate contact with your targets and sources with this question does wonders to make you careful to be fair, accurate and complete.

Beyond newsprint

But instead of fading in the drizzle or sinking in Puget Sound, the story kept growing, and growing, spreading by bits, bytes and word of mouth.

Thanks to e-mail and the Web (and of course my editor and publisher), the story gained national and foreign attention and, finally, prompted significant reforms in laws, public understanding and industry practice.

Like any self-respecting investigative project nowadays, "Fear in the Fields" was documented with help from spreadsheets, a database, the Web and e-mail. More importantly, it gained credibility and impact from what would have to be termed computerized, Web-based, informal publishing strategies following the newsprint version of the story.

Starting from the end: Every time I wrote a follow-up article, I sent it by e-mail to 100-200 experts and activists on every side of the issue.

I kept a contact list in Microsoft Outlook Express 4.0 Address Book. This is a fully searchable mini-database that some people say is as good as Eudora Pro for e-mail, and it's free. Outlook Express makes it easy to build groups of people in e-mail lists.

Often these experts and activists would get the article in e-mail before they could buy it in the newsstand.

I asked these people if the articles were completely fair, accurate and complete – a practice I have previously done by telephone. Knowing you're going to initiate contact with your targets and sources with this question does wonders to make you careful to be fair, accurate and complete.

The cover note on the e-mail would summarize the new article in a sentence, provide a link to the article, ask for feedback and follow-up ideas, and provide links to source materials.

Electronic contacts

This added credibility. It also meant the insiders' first response came directly back to me. So I knew the buzz inside industry and government. I harvested the follow-up suggestions. I knew whether any questions were being raised and, in one case, corrected a minor error before the final deadline.

The Associated Press and Knight Ridder wire picked up on parts of "Fear in the Fields," but this e-mail list extended the immediate distribution to 20 other states and five foreign countries. It pushed the articles to the desktops of scientists, industry officials, environmental activists, trade journals and government policy makers.

They could always look for more information on the Web, too. Our site (*http://www.seattletimes.com*) linked all the articles from a Special Projects area that got more than 1,000 hits a day for weeks.

We did not have time to put source documents on our Web site in Adobe Acrobat format. I wish we had. It adds credibility. It adds punch. The *Miami Herald*'s "Collars for Dollars" is a model.

Post-publication tips

These were some other uses of the computer after publication:

• "Fear in the Fields" generated more comments (most positive) than anything else we've done, with more than 200 e-mails, 100 phone calls, and 15,000 reprints. I sent a mass e-mail

Continued on page five

WILSON WRITES:

THE RIGHT-TO-KNOW NETWORK (HTTP://WWW.RTK.NET) DOES A GREAT JOB OF MAKING THE TOXICS RELEASE INVENTORY ACCESSIBLE.

THE ENVIRONMENTAL WORKING GROUP (HTTP://WWW.EWG.ORG) PUBLISHED A HUGE REPORT ON THE TOXICS-TO-FERTILIZER PHENOMENON.

THE FERTILIZER INSTITUTE (HTTP://WWW.TFI.ORG) IS TRYING TO DENY IT IS A PROBLEM.

I FOUND A CANDID HAZARDOUS WASTE RECYCLER AT HTTP://MEMBERS.AOL.COM/FOURN6/INDEX.HTML.

ALSO SEE PAGE 15 FOR AN EXPLANATION OF THE RECENTLY UPDATED DEPARTMENT OF TRANSPORTATION HAZARDOUS MATERIALS INCIDENT REPORTING SYSTEM DATABASE, AVAILABLE FROM THE DATABASE LIBRARY.

From page four:

Wasting costs away

to about 100 of those people reporting on the actions that followed the articles.

• I published tip sheets for journalists in other states who would like to undertake similar investigations: *http://www.seanet.com/~dufflire/handouts.htm*

• Toward the end of my reporting, I used AskSam, a free text database program, to organize more than 50 *.doc files, 20 *.txt files and 300 e-mails in my PC. AskSam lets you search for a word or phrase in different types of documents and returns a list of the hits.

To some people, e-mail is a less threatening way to answer questions. They are careful in their responses – given a sensitive topic, that is fine with me.

Hidden in plain sight

Back at the beginning: When I heard the incredible claim that manufacturing industries were disposing of toxic wastes by turning them into fertilizers, the first thing I did was to go online.

I typed "fertilizer" and "toxic" in Yahoo, Alta Vista and Hotbot search engines. Nothing came up.

Later I found other, probably better search engines – Northern Lights and Inference – and other Web sites where those two words were discussed together. I found hazardous waste recyclers, conferences on recycling waste to fertilizer, and even the "Industrial Materials Exchange" that serves as a kind of blind-dating service between generators and recyclers. *All hidden in plain sight – on the Web.*

And I discovered Medline (*http://www.ncbi.nlm.nih.gov/PubMed/*), which gets my vote for the best site on the Web. Thank you, Al Gore, for making this free. Medline lets you search 9 million medical and scientific studies.

One of the key breaks in my investigation was the discovery of a company, on its Web site, blasting its competitors' fertilizer products for being recycled from hazardous wastes without having cleaned up the toxic lead and cadmium (*http://www.cozinco.com/*).

Another break was the discovery of Australian and European Web sites where researchers talk about the risks of cadmium in fertilizer.

Authority granted

As the work went on, I kept a spreadsheet of fertilizer test results in Excel, sortable by column headings such as cadmium or manufacturer name.

This enabled me to say more than 100 fertilizer products whose analytical results were obtained by *The Seattle Times* had higher amounts of lead than the average farm soil.

I could also say, with authority, that recycled hazardous-waste fertilizer products contain higher amounts of lead and cadmium than the other products on the market. I was happy to share this spreadsheet with anyone who wanted it.

I used e-mail to file public record requests, troll for experts in farm-related and medical discussion groups, and correspond with dozens of experts, farm families and fertilizer manufacturers. To some people, e-mail is a less threatening way to answer questions. They are careful in their responses – given a sensitive topic, that is fine with me.

I also grabbed all the e-mail of one state official who was in the center of the practice. I used the state public records law that makes it clear electronic records are to be treated the same as paper records.

People like to say computer-assisted reporting doesn't take the place of old-fashioned boot leather. Now isn't that a cowpie? Naturally I pounded a lot of pavement, corridors and dusty farm fields in my months of reporting. But the Internet pulled the whole world to the undersized computer monitor in my overcrowded cubicle in a dark corner of the newsroom.

Come to think of it, maybe I should have spent more time on the farm. Things slow down there, but you get a lot more sunshine and fresh air.

Duff Wilson can be reached at dwilson@seattletimes.com or through the Reporter's Desktop, http://www.seanet.com/~duff/

THE OCTOBER 1998 ISSUE OF UPLINK WILL FOCUS ON THE USE OF COMPUTER-ASSISTED REPORTING IN ENVIRONMENTAL STORIES. IF YOUR ORGANIZATION COMPLETED AN APPLICABLE STORY AND YOU WOULD LIKE TO SHARE YOUR INSIGHTS ON SUCCESSFULLY DOING ONE, PLEASE SEND AN E-MAIL TO BRENT JOHNSON AT BJOHNSON@NICAR.ORG.

OTHER UPCOMING ISSUES OF UPLINK WILL FOCUS ON EDUCATION, MAPPING, THE CENSUS, CAMPAIGN FINANCE, CRIME, TRANSPORTATION, INTERNATIONAL CAR AND CAR AT SMALL NEWS ORGANIZATIONS. IF YOU WOULD LIKE TO CONTRIBUTE AN ARTICLE OR COLUMN TO ONE OF THESE ISSUES, PLEASE SEND AN E-MAIL TO BRENT JOHNSON.

THE WEB SITE FOR THE CAMPAIGN FINANCE INFORMATION CENTER – WWW.CAMPAIGNFINANCE.ORG – CONTINUES TO GROW. BESIDES DOWNLOADABLE DATA AND LINKS TO OTHER ONLINE SEARCH ENGINES, THE SITE ALSO INCLUDES:

- A DIRECTORY OF CAMPAIGN FINANCE REPORTERS, SEARCHABLE BY NAME OR STATE. THE DIRECTORY IS UNDER CONSTRUCTION. IF YOU WOULD LIKE A NAME ADDED, SEND AN E-MAIL TO CFIC-COMMENTS@CAMPAIGNFINANCE.ORG
- A LIBRARY OF CAMPAIGN FINANCE STORIES CONDUCTED ON THE FEDERAL, STATE OR LOCAL LEVEL
- THE WINTER 1998 AND SPRING 1998 ISSUES OF TRACKER

TO SUBSCRIBE TO THE CAMPAIGN FINANCE INFORMATION CENTER MAILING LIST, SEND AN E-MAIL TO: MAJORDOMO@CAMPAIGNFINANCE.ORG. IN THE BODY OF THE MESSAGE, TYPE SUBSCRIBE CFIC-L <YOUR E-MAIL ADDRESS>

BOOMING OR BUSTING?

Rising bankruptcies

By Valerie Lilley
Peoria Journal Star

The economy is booming and so are bankruptcy filings. What gives?

In a week-long series, we told readers everything about the Tri-County area's bankruptcy filings. Through computer-assisted reporting, we used numbers to show who the filers were, as well as the causes, losers and winners in the surge of filings.

Surprisingly, we found that most people who filed had been living below or near the poverty level. They weren't the rich – a popular misconception.

The oddest fact we found was that many filers listed the mail-order company Fingerhut, which made it to the top of the list of creditors. Bills were run up using cellular phones and charge cards at Bergners, a local, upscale department store.

Credit cards were the main cause of the bankruptcies. Medical debt trailed far behind.

Incomplete data

To gather these statistics, we had to build our own database because we couldn't find an existing database with complete information.

The U.S. Bankruptcy courthouse had PACER (Public Access to Court Electronic Records), but it didn't list all creditors. The Administrator of the Courts in Washington, D.C., didn't have a complete listing of case contents, either. The courthouse clerks and office manager didn't know of anything else.

We decided to collect the data ourselves. This was our first big dive into computer-assisted reporting, and naiveté drove it from a mini-project of a few months to a year-long study. Collecting the data took the most time in preparing the project.

The first lesson we learned is that mining data is not as easy as it looks. From 2,906 cases, we gleaned data such as: filers' names; ZIP codes; residential market value and money owed on it; car market value and money owed; attorneys; assets and liabilities; government support type and income; child support income and expense; income figures from the two years prior to filing up to the present; monthly expenses; and creditors and amounts owed them. (Some cases listed more than 200 creditors.)

Initially, another reporter and I went to the court house armed with laptops. For a week, case by case, we inputted the data in comma-delimited format. Both of us made multiple mistakes. By the end of the week, we only had 90 cases in our database. We ended up hiring someone to input for us.

Because of the time cases can take to close, we ended up using cases opened and closed between January, 1996, and February 1997.

To check for errors, we ranked by case number and by the minimum and maximum amounts in each column. The cleanliness of the database was also double-checked by a professor at Illinois Central College.

We used Excel to find the averages, sums, minimums, maximums, and the time each case took from open to close. To find how many filers owned a house, we ranked the database by the "residence market value" field and found where the cells dropped off to zero. Only 753 people owned a home. Its average value was $41,701.

Tagging along

We also used Visual FoxPro to find the top creditors, total credit card debt, total unsecured debt and total medical debt. To differentiate between credit cards and regular bank loans, identifiable credit cards were tagged with a "-C." We put a "-M" after doctor's names, physician offices, clinics, hospitals and other medicine-related creditors. We also wanted to know how much wasn't going to get paid back – total unsecured debt, which we clearly listed on a separate schedule in the files and tagged with a "-U."

To find the top creditors, I cut and pasted the creditor columns and the amounts into one super-long, two-column spreadsheet. To double check that I had gotten all of the creditors and amounts in the cut and paste process, I compared the sum to the sum in the original Excel spreadsheet. I imported the two-column spreadsheet into FoxPro and then used the "order by" and "group by" commands.

I used the same table to select the creditors

Continued on page seven

ON THE INTERNET

Estimating relevancy

By Robert Berkman
The Information Advisor

In the information world, units of related information are called "records." A record can be an article, an abstract of an article, a bibliographic citation, a Web page, and so on.

A "field" is a specified portion of a record. For example, for Web pages, fields may include the Web page's title; the date that Web page was last modified; the Web page's URL; or the Web page's meta tags. Meta tags are the textual description – though invisible to users – of the Web page, written by its creator to inform search engines what the page is about.

Some search engines allow you to instruct the engine to look for your keywords *only* in a specified field. When would you want to do this kind of limited field search?

• If you did a keyword for, say, "Vietnam," you'd be swamped with sites. By restricting your search to the title field, you'd help insure that the keyword is the primary focus of the Web pages returned.

• If you want to search for recent events, some search engines, such as Alta Vista, let you search for a date field. You can specify that you only want recently updated Web pages.

• Sometimes you may only want to see Web sites produced by a specific organization or type of institution. In these cases, you can search the URL field.

Say, for instance, that you're doing a story on Sony, and you want to locate Web pages produced by the company. Zero in on Sony's own sites by searching for "Sony" only in the URL, which normally contains the name of the institution that created the page.

Continued on page eight

From page six:

Filing flurry

and the amounts where the creditor had the "-C" for credit card. The "order by" and "group by" gave the top ten. I repeated this to find top unsecured creditors and medical debt.

The hardest part was finding out the average credit card debts per case because the "-C" tags were included with any "-U" and "-M" tags. I didn't know how to write a loop or what a scan/endscan was. Justin Mayo of NICAR worked out a program for us using a table with the creditors in the original block form.

Handle results with care

We approached the data results in the same manner as a press release. I put the results in sentence format — it was as if we were reading the traditional government and commercial studies we were used to.

During weekly meetings, people threw out questions like "How many filers live with the city limits?" and "What's the median debt per case?" We made cheat sheets and used them as a starting point for reporting.

For example, Sears and CEFCU, a local credit union, were owed the most money — $2.9 million and $2.7 million respectively. Central Illinois Light Co. and Ameritech were also up there. I interviewed company executives and other spokespeople. They provided the story behind the numbers.

Finding real voices

We talked to the people who filed. They were hard to find because most led transient lifestyles. We used the Journal Star's electronic library to get their personal information — we publish bankruptcies in "Matter of Record" along with divorces, marriages, births.

Getting people to talk about their filing was our biggest problem. Some didn't want to rehash old, sour memories. Others didn't want their names published in the newspaper — again. So we ended up with a mix of full names and first names only.

Organizing everything for our readers also challenged us. We had so much data that we could have easily bombarded them with stat-packed stories, covering a broad range. In our numerous weekly meetings, however, we had isolated specific elements of the bankruptcy phenomenon. These elements became the stories and sidebars of our week-long series.

Valerie Lilley can be reached by e-mail at vlilley@pjstar.com or by phone at (309) 686-3260.

THE WEEK-LONG BANKRUPTCY SERIES FOCUSED ON THE FOLLOWING QUESTIONS:

- **WHAT ARE THE ELEMENTS OF THE BANKRUPTCY DILEMMA IN THE PEORIA AREA?**
- **WHAT IS BANKRUPTCY? IT IS FULL OF CONTRADICTIONS.**
- **WHAT'S THE DEBATE? CONGRESS IS GRAPPLING WITH THE TWO SIDES OF BANKRUPTCY.**
- **WHAT CAUSES PEORIA BANKRUPTCIES? CREDIT CARDS GET MUCH OF THE BLAME.**
- **WHO LOSES IN BANKRUPTCIES? WE ALL DO.**
- **WHO WINS? THE BUSINESS OF BANKRUPTCY IS BOOMING TOO.**
- **WHAT CAN BE DONE? A LOOK AT THE POSSIBILITIES.**

BERKMAN WRITES:

ON THE INTERNET, THERE ARE TWO MAJOR CATEGORIES OF WEB RESEARCH TOOLS: RATERS AND ROBOTS. ON YAHOO, THE MOST WELL-KNOWN OF THE HIERARCHICAL INDEXES, SEARCHERS LOOK FOR SITES BY "DRILLING DOWN" -- BY FIRST SELECTING A BROAD CATEGORY, VIEWING THE NARROWER SUBCATEGORIES, THEN CHOOSING UNTIL THE DESIRED SITES ARE LOCATED.

ANOTHER, NEWER SITE THAT WORKS THE SAME WAY IS CALLED LOOKSMART (HTTP://WWW.LOOKSMART.COM).

From page seven:

Flooding engines

Another example is when you want to see sites only from a particular type of institution. If you were looking for statistics on aircraft deaths but only wanted data from official governmental sites, you could instruct the search engine to return only sites that end in ".gov".

Know when to modify

As any librarian knows, getting good results from an online search often means tinkering with the initial search statement. If an initial search turns up no Web pages, too many, or irrelevant ones, it's a good idea to examine and possibly alter that keyword search statement. Try the following:

Check the spelling.

If you are searching unfamiliar territory, make sure the words you've used are standard terminology. For example, if you were searching for Web pages on "static electricity," it would be important to know that the technical term for that phenomenon is "electrostatics."

Double-check your syntax if you used Boolean operators. If you didn't get enough hits, take out any AND operators you entered. If you got too many hits, either add an AND operator to restrict your search, eliminate one or more keywords, or limit your keywords to a title field.

One-way bridges

One caution: even if you've done a perfectly good search, you still can't assume the Web pages that the search engines returned are really the most relevant. Why? Some site creators are manipulating their pages and the keywords on their sites with the specific goal of getting high rankings by the search engines.

According to Danny Sullivan, who publishes an e-mail newsletter called Search Engine Watch (*http://www.searchenginewatch.com*), some sites do this by developing what are called "bridge" pages. A bridge page is created specifically to receive a high ranking from a search engine. When you click on the decoy-like page, you are routed to the site's actual home page.

In his newsletter, Sullivan found, for example, that State Farm Insurance submitted several pages to a variety of search engines. Each page was different and specifically written to rank well in search engines' ranking algorithms. Several pages were submitted to each search engine and written to "capture" a particular insurance-related keyword like auto insurance or boat insurance.

Cyber consultants

Web site owners also are contracting the services of consulting firms that claim they can help the site get higher rankings.

These companies take their clients' Web pages, run them against the major search engines, analyze placement of the page, and then provide specific strategies on how to improve rankings. One of these firms, called Webposition (*http://www.webposition.com*), also provides free advice on its site on how to improve rankings on a search engine.

Web illusions

Things are not, then, always what they appear to be in the brave new world of the Web. On a traditional online service, like Dialog, you could assume that if you conducted a good search, you'd pretty much get what you expect. But on the Web, it appears the data creator's intentions are another force making an impact on what you'll retrieve.

Although popular search engines like to claim that their robots scour the entire Web, the fact of the matter is no search engine can actually index all the pages out there. The April 2 issue of *Science* magazine reported on an extensive study by Drs. Steve Lawrence and C. Lee Giles of the NEC Research Institute. The researchers posed over 550 scientific search questions to the five largest search engines and discovered that there are about 320 million Web pages – but that search engines index only about 40% of them.

Information overload

If you feel overwhelmed by the enormous number of hits you get back from Alta Vista or one of the other well-known search engines, try a newer search engine called Northern Light (*http://www.nlsearch.com*). Northern Light does an admirable job in making a dent in information overload by grouping the sites it retrieves into logical folders, which contain related information.

Robert Berkman can be reached at (413) 549-9924 or by e-mail at rberkman@javanet.com.

Sampling: Not a dirty word

By Sarah Cohen
NICAR

This is the first installment of a two-part series

In a project on the quality of health care, reporters were interested in the details of malpractice claims. There were something like 50,000 payments to victims of doctors, hospitals and other health care providers in a state database. However, the computer records included only summary information. Paper records were the source for more details, such as the nature of the injury and how it happened.

This problem often arises when working with documents, especially court records. You can find docket information on bulletin boards or other electronic means. But details ranging from previous criminal records to prosecutor recommendations simply don't exist electronically.

These are perfect candidates for sampling in the early stages of a story.

Sampling situations

Most editors and reporters would insist that when the story runs it ought to include all instances. They argue it's simply worth the effort to gain more credibility with readers and viewers. But you can sample early on to find out how many and what kinds of records will – or more importantly won't – help your story.

You may even find early in a story's planning that the effort probably won't result in a credible story. The records are too muddled to be of much help.

Large databases are another candidate for sampling in the newsroom. I think most computer-assisted reporting experts would acknowledge that they've at least once waited hours for a query to run – sometimes overnight – only to find that they left out a field or forgot to include an important case.

Here are some considerations to think about when you begin a project and want to consider sampling as a way to make your work more efficient. Some are geared at figuring out what you need to know before you move further into the deep, dark hole of a long-term story. Others help you figure out whether you need scientific sampling techniques.

Collection difficulty

Sometimes, sampling has little place simply because it doesn't save you much. Collecting all of the records on OSHA violations or fatal accidents, for instance, is probably easier than sampling. There's no reason to revert to a small set of records when you can just as easily test all of them. But if you simply can't get all of the records, such as the details of thousands of court cases, then you'll need scientific sampling methods before you can say anything about what's in them as a group.

Passing acquaintance

When you're dealing with documents you've never worked with before, you'll probably want to start out with a small group of them to test your database. I usually don't recommend bothering with scientific sampling techniques here. Instead, consider typing in the 50 most recent and 50 from the earliest period you're considering. This will give you three legs up on the project.

First, you'll see whether you have the skill (or whether it's even possible) to create queries or other analyses that will address your questions from the database the way you've designed it. Too often, reporters call after spending months typing in records without making this simple test. It makes their work much harder.

Another advantage is that you'll see how the paper records have changed over time. If you're compiling records from many local offices, though, you'll need to take another step: gathering 10 or so from a few different far-flung districts. This will show you how they vary in quality and in content from place to place, as well as over time.

Finally, you'll know what difficulties you will find. In the malpractice example, for instance, reporters who had stopped early in the project and looked at the most recent

Continued on page fifteen

UPCOMING BOOT CAMPS:

OCTOBER 22-25, 1998 – POWER BOOT CAMP IN COLUMBIA, MO. (A BASIC BOOT CAMP SQUEEZING SIX DAYS OF TRAINING INTO FOUR.)

JANUARY 3-8, 1999 – BASIC BOOT CAMP IN COLUMBIA, MO.

MAY 16-21, 1999 – BASIC 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

Clogged artery

By John Kelly
The Shelbyville News

You'd panic, too, if you had never seen nine-track tapes. I didn't let on to the clerk that I had no earthly idea how to get what I wanted from them. I had hired someone to take data from the tapes if necessary. It was necessary. I paid the Indiana state police about $150 for the tapes, which held three years of police reports for every traffic crash in the state, as well as my chance to do our first major computer-assisted reporting project.

The database helped us show readers there were more accidents on three miles of that road than on any other road in Shelby County. Our reporting helped us find reasons and solutions.

The Shelbyville News already had dipped its toes into the CAR pool. We use spreadsheets to analyze government spending. We use the Internet daily. We built small databases for a feature on local inventors and to tabulate a survey about community parks.

But we had never tangled with a 500MB relational database. It was overwhelmingly scary at first, but the computer work turned out to be much easier than I ever dreamed.

Head-first plunge

So we dove into the CAR deep end and gave our readers something special. We delved into every aspect of traffic crashes in Shelby County, empowering readers with information they could use to influence public officials and take action themselves to make the streets safer.

We didn't stop with ranking roads where people crashed most. We used the database to identify where crashes happened, what caused them and how they happened. Reporters talked to countless cops, business owners, road engineers, planners, safety experts and drivers.

We didn't send politicians running. We didn't stun the public. We did what we always do. We took what we discovered and developed it into stories that showed readers what's happening, why it's happening and how problems can be fixed.

Accident prone

We showed the public how the city and state governments made no improvements as East State Road 44 – once a country road lined with corn fields and a drive-in theater – developed over three decades into a busy business corridor. The database helped us show readers there were more accidents on three miles of that road than on any other road in Shelby County. Our reporting helped us find reasons and solutions.

City and state officials told us heavy traffic was to blame. Maybe so, but there was something else, according to police, business owners and drivers. Bad planning and a lack of communication between city and state agencies aggravated the problem, they said. Computer analysis and further reporting showed they were right.

Driveway dilemma

Police, business owners and drivers told us the biggest problem was drivers crossing several lanes of speeding traffic to get in and out of strip-center parking lots. The road is packed with driveways. The state handed out permits for driveways like bank tellers hand out lollipops. In fact, they let new businesses cut more and more driveways until there were places where six driveways sat within a few hundred feet of each other – despite state safety guidelines saying driveways should be at least 150 feet apart.

The database showed us most of the accidents were caused by cars slowing or stopped to turn into the lots or by cars turning onto the highway from them. Transportation department officials said statewide there was too little money to fix too many roads. The engineer responsible for approving driveway permits told us that Shelby County residents needed to improve their driving skills to adapt to changing traffic conditions.

The road affected everyone in our community. In this small Hoosier city, this was a huge story. To buy groceries, grab a fast-food

Continued on page eleven

TO SEE HANDOUTS ADDRESSING THE NEEDS OF SMALL ORGANIZATIONS TO DEVELOP THEIR CAR STATUS, TURN TO PAGE 12.

ALSO, YOU CAN ORDER JOHN KELLY'S HANDOUT – "BUILDING A FOUNDATION: SMALL NEWSPAPERS CAN EXPLOIT CAR TOOLS TO IMPROVE GOVERNMENT COVERAGE, BUILD REPORTERS' SKILLS AND BETTER SERVE THE PUBLIC" FROM IRE'S RESOURCE CENTER. IT'S HANDOUT #769 FROM INDIANA CAR IN MARCH 1998. CALL (573) 882-3364.

From page ten:

Driver education

lunch or go shopping, readers must drive that road. They knew there was a problem. They had theories about what was wrong. Still, most were calling for cliché solutions like more traffic lights, lower speed limits or a center turn lane.

We could have written those things, but the database gave us a chance to show other reasons behind the problem. After reading our report, residents and public officials alike agreed the issue was planning – or lack of it.

Because we asked for help, ... we've given the paper's leaders a taste of what CAR can mean for reporting, and now we'll get more time for CAR projects.

Change of plans

Now city zoning officials are talking about access roads and asking about the impact of new developments on traffic. They are redesigning one intersection before a developer builds a new shopping center and housing addition there. They are pressing the state for input in planning and driveway permits.

Whether the reports will affect state planning is unclear. The state's cooperation with the city on redesigning the one intersection could be a sign of progress. One transportation department official has said the district office pinned our stories on a bulletin board. We haven't found out yet whether they're learning from it or tossing darts at it.

The stories about East State Road 44 anchored the series, but they were just two of a dozen. In other reports, we showed readers how Indiana law allowed teen-agers to get driver's licenses with little behind-the-wheel experience and how teen-agers crashed much more often than any other age group. In addition, we showed how drivers of all ages were their own worst enemies on the road. Three-fourths of all crashes were caused by the fault of a driver, including half because at least one of the drivers was not paying attention.

Small paper, big project

Readers loved it. The series was comprehensive, insightful, sophisticated and fun. In newspaper terms, it was good journalism. Now, we're ready to do more of it.

We learned a great deal from this project. We learned and hopefully showed others that even the smallest newspapers can do big CAR projects. We learned how to do CAR right. Our next projects will be faster and better because of the experience.

We learned what everyone doing CAR – at whatever level – should remember: don't be afraid to ask for help when it's needed. IRE members offered guidance and warnings when I wrote them. Richard Mullins and the NICAR staff taught me how to clean the database at an intermediate boot camp. A data recovery company helped me move the state police data from nine-track tapes onto a CD-ROM, then turn a flat file with seven different types of records into seven comma-delimited files so I could import the tables into Microsoft Access.

Because we asked for help, we were able to overcome what could have been project-ending (and CAR-ending) obstacles for our small newsroom. Instead, we've given the paper's leaders a taste of what CAR can mean for reporting, and now we'll get more time for CAR projects.

While I did the database work, everyone in our newsroom helped. That is critical for small papers. Two of us did the bulk of the reporting, but every reporter conducted interviews for the stories. Our editors, graphic artist and photographer put together a terrific graphic presentation so writers could leave some numbers out of stories while still giving readers gobs of stats. Copy editors ravaged the early drafts until we got it right. And our publisher laid the groundwork by investing in CAR training for me and CAR resources for our newsroom.

Our next major projects (already in the works) will be more ambitious and hard-hitting.

John Kelly can be reached at
jkelly@shelbynews.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

Spreading CAR at small news organizations

These handouts are excerpts from handouts provided at IndyCAR and the IRE National Conference in New Orleans.

By David Cuillier
Tri-City Herald
Kennewick, Wash.

David Cuillier attended the May 1997 Advanced Boot Camp.

So you are the only one at your small news organization who is using computer-assisted reporting skills, and you want to spread the knowledge. Most everyone seems interested, but few people actually do anything. And those who do take the initiative to learn often end up leaving for bigger ponds.

Many of us have been fishing for a way to lure people into learning computer-assisted reporting at small news organizations. Here's a two-pronged approach to tackling it: Gill-netting along with precision casting.

Gill-netting

Like fishing boats that cast their nets wide, try making basic computer research tools available and easy to use for everyone, even those who don't seem to have any interest. You may find that even the most reluctant reporters will start to use the Internet, online services or CD-ROMs. Here are some ideas:

• **Location, location, location:** If you have a designated CAR computer, put it in a central location in the newsroom on its own desk. Make it clear the computer is available for everyone to use – that it's not the domain of one person or department.

• **Book smart:** Set up a book shelf at the CAR station, filling it with self-help software manuals, back issues of Uplink and useful texts geared toward computer-assisted reporting.

• **Hot off the press:** Produce a monthly in-house CAR newsletter. Give praise for stories that benefited from CAR, especially deadline work. Also, write about what stories are in the works, and offer ideas for CAR stories that might interest reporters. Make sure to route copies to the editor, publisher, advertising director and other company managers. It will help build support for CAR work, future equipment purchases and inter-departmental computer trades.

• **Get linked:** Subscribe to the monthly newsletter *Uplink*, put out by the National Institute for Computer-Assisted Reporting. Show beat reporters how they don't have to be CAR gurus to benefit from a subscription.

• **Salt that data:** Set up brown-bag lunches to show folks some data you acquired, how you crunched it and what you got out of it. Or show useful sites on the Internet or how spreadsheets can help look at numbers.

• **Noteworthy:** Type up cheat sheets for the computer resources. Laminate them or put them in plastic covers. Give each staff member a copy and put them in a binder at the CAR station.

• **Group acquisition:** When you run across an interesting database, get it, even if it's not on your beat. Give it to the appropriate reporter with some ideas for what the data could reveal. If you're an editor, seek out data for every beat to pique the interest of all the reporters.

Precision casting

Some reporters become interested in learning advanced skills, and with a little focused help they may progress fast. If it's unlikely the bosses will send them to training boot camps or classes, the job may rest on you. Seize the interest fast.

You probably don't have a lot of time, so focus your energies on one person at a time. If you give them heavy training over three months with follow-up sessions, then you could train three or four people a year. For most small news organizations, that is exceptional. Even one or two a year would be great.

At least once a week, find a good time for both of you. Stick to it.

Use real data. Start small and work up.

Let them do the work and go to you only for questions when they get stuck or need guidance.

If you attend CAR seminars, boot camps or conferences, take notes on how they teach the sessions. Take home the exercises and try them out on your "students."

Continued on page thirteen

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. HANDOUTS ARE SORTED BY SUBJECT, YEAR AND LOCATION.

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

From page twelve:

Jumpstarting CAR

Maintaining CAR on duct-tape budgets

The following are big resources for small newsrooms to overcome snags and boost skills.

In-house resources:

• Techies: Computer services staff at your organization can help overcome problems or handle nine-track tapes. Bring them donuts.

• Bean counters: Accountants on your business department are probably proficient with spreadsheets. Also, they might have a spare copy of Excel hanging around.

• Marketing: Marketing/advertising might use mapping software for developing demographic reports for readership studies and advertisers. MapInfo is popular in the business world but can work in newsrooms. ArcView is what government tends to use.

Out-house resources:

• Pros: Local computer experts are sometimes willing to help program or crunch data, often for free. Or consider giving them credit in the story.

• Universities: Take courses at universities or community colleges. Not only do you learn, but as a student you can get substantial discounts on software.

Online guides

Here are some places online to learn about CAR, use the Internet and get story ideas:

• CARguides: *http://www.poynter.org/car/cg_chome.htm.* A guide for tapping into online information by Nora Paul of the Poynter Institute. Includes introduction into CAR and explanations of resources.

• Deadline Online: *http://www.deadlineonline.com.* A great breakdown for finding facts, people and sources. By Alan Schlein, who also provides specialized Internet training for journalists.

• Reporter's Desktop: *http://www.seanet. com/~dufff.* Compilation of search engines, people finders and other resources for finding information on deadline. By Duff Wilson, investigative reporter at *The Seattle Times.*

• CAR/CARR: *http://www.ryerson.ca/~dtudor/carcarr.htm.* A list of links to CAR sites and other useful places to check out. By Dean Tudor, journalism professor at Ryerson Polytechnic University in Toronto, Canada.

• The Beat Page: *http://reporter.org/beat/* Internet links and CAR resources grouped by beat, from agriculture to transportation. Complied by Shawn McIntosh of *The Clarion-Ledger* in Jackson, Miss.

• WWW Virtual Library: *http:/www.cais.com/makulow/jexercises.html.* Many CAR resources by John Makulowich, including Internet exercises for journalists, Internet tips and links to journalism organization. Hosted by The Writers Alliance Inc.

• CARwash: *http://www.wnpa.com/carwash/carhome.html.* CAR stories done in Washington state, primarily by small newspapers, posted in full text. By David Cuillier, *Tri-City Herald.*

David Cuillier can be reached at (509) 582-1535 or by e-mail at dcuillier@aol.com

Chasing the big story

By Hector San Miguel

Lake Charles American Press

The following Web sites will make your reporting job easier.

• Public Integrity Research Corporation: *www.pihome.com/pirc/pirc11.html.* What a site. It contains great links for general reference, fact searches, statistics, phone directories, search engines, government databases, miscellaneous references, statutes text, and business reference.

• The Internet Sleuth: *www.isleuth.com/arts.html.* Web sites don't get any cooler than this one. It contains more than 3,000 databases you can access.

• Newspapers Online: *www.newspapers.com.* This amazing site gives you access to U.S. and foreign newspapers and other publications on the Web. Many of the newspapers online feature searchable archives.

• Consumerworld: *www.consumerworld.org.* This is a great "public service site" that has access to more than 1,700 consumer resources on the Internet. It has everything from product reviews to consumer rights booklets. This is a definite bookmark for the business reporter.

Hector San Miguel can be reached at (318) 494-4084.

JOBS:

FOR RECENT JOB POSTINGS TO THE IRE WEB SITE, POINT YOUR BROWSER TO WWW.IRE.ORG/JOBS

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

IF YOU WOULD LIKE TO SEE A PARTICLUAR PROBLEM ADDRESSED IN THIS COLUMN OR IF YOU WOULD LIKE TO CONTRIBUTE YOUR OWN SOLUTION TO ONE, PLEASE SEND AN E-MAIL TO BRENT JOHNSON AT BJOHNSON@NICAR.ORG

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 TRI/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.

Tell me Y2K you cry

By Richard Mullins
NICAR/Missouri School of Journalism

Access and FoxPro don't have a year 2000 problem – they can handle dates from 100 A.D. up to 9999 A.D. That should handle the worst jail terms and back-dated purchase orders.

So the real Year 2K problem is yours. You need to understand a little of how Access and FoxPro store dates and know the assumptions that Access and FoxPro use for implying the century when you only enter or import two digits for the calendar year.

FoxPro and Access do not store any dates with only two digits for the year, leaving the century ambiguous or unknown. Both of them, however, give you various ways to control the display format of dates, including whether the year has two or four digits.

Let's deal with FoxPro first. You can toggle the display of the century with the SET CENTURY command. You can test this feature in the command window to see how it works. The lines in bold are the responses to the print-to-screen command. If your computer clock has the current date as July 1, 1998, then:

```
set century ON

? date()
7/1/1998

* back to the default
set century OFF

? date()
7/1/98
```

When converting dates, FoxPro assumes all two-digit years to be 20th century, regardless of the current setting of CENTURY. This default can be overridden.

It is a data-processing standard at NICAR to convert all character dates to the FoxPro date datatype. We also preserve the original character date column. That way, there is no loss of original data and conversions can be checked and re-done. For any character representations of a date that do not represent a valid calendar date, like 19950335, FoxPro deals with this by converting it to a null or empty date. You can do a query for empty dates to see what was in the character original. Here is a query using the FEC data as an example. The column **chardate** is the character original; **cont_date** contains the date datatype.

```
SELECT chardate, cont_date;
  FROM Indc0798 ;
WHERE empty(cont_date)
```

The `empty( )` function in the criteria is the same as:

```
WHERE cont_date = { }
```

Unlike FoxPro, Access makes a conditional assumption about the implied century for a two-digit year. The cutoff year number is 30. Two-digit years of 30 and higher are assumed to be 20th century. Two-digit years from zero to 29 are assumed to be 21st century.

Here is what the test of this looks like in the Access **Debug** window. (Open a new or existing module, then choose **View/Debug Window.**)

```
? cvdate("07/01/98")
7/1/98

? cvdate("07/01/28")
7/1/2028
```

Here is the issue you'll have to consider when looking at new data sets. Do they have date information with two-digit years? If so, can you do integrity and reality checks to determine that some dates should be 21st Century and not 20th. Some obvious examples: an inmate gets out in the year 2015, not 1915; a building was built in 1920, not 2020. If so, you can make the appropriate conversion.

NICAR has been evaluating the data we get for the data library and for custom data jobs with this issue in mind, but so far the century issue has not come up.

Richard Mullins can be reached at
(573) 882-2737 or by e-mail at
richard@nicar.org

Hazardous materials

By Seth Hemmelgarn
NICAR

NICAR's Database Library has recently updated the hazardous materials (Hazmat) database. The data are from the Department of Transportation's Hazardous Materials Incident Reporting System. You can find all records dating back to 1971 – when the system was established – and going through early 1998.

Any time a hazardous material is accidentally released while moving from its origin to its destination, a report is made and the data is entered into the system. All modes of transport are included, such as trucks, trains, planes and boats. An accidental release can mean a barrel falling out of the back of a poorly latched truck or a plane crashing.

The data include what the material was, the carrier company's name, where the material was coming from and where it was headed, and where the release happened. There are also fields that indicate if anyone was injured or died, if there was a spill and how much monetary damage was done. These tables can be linked together using unique identifiers.

The database is broken up into three main tables. These tables consist of the details and consequences of the accident, what failed – people or containers – and remarks from the reports. The six lookup tables, which simply provide more details, contain chemical codes and classifications, shipper's names and addresses, and FIPS codes.

Seth Hemmelgarn can be reached at (573) 884-7711 or by e-mail at seth@nicar.org.

NICAR'S DATABASE LIBRARY HAS RECENTLY UPDATED THREE FEDERAL AVIATION ADMINISTRATION DATABASES: THE ENFORCEMENT INFORMATION SYSTEM DATABASE, THE SERVICE DIFFICULTY REPORTING DATABASE, AND THE ACCIDENTS AND INCIDENTS DATABASE. THE LIBRARY HAS ALSO ADDED THE USDA SUMMARY DATA OF MEATPACKING AND SLAUGHTERHOUSE INSPECTIONS FROM 1990 THROUGH THE THIRD QUARTER OF 1997.

FOR MORE INFORMATION ON THE CONTENTS OF THESE DATABASES AND THEIR COSTS, CALL THE LIBRARY AT (573) 884-7332.

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

From page nine:

For efficiency's sake

records would have noticed the fatal flaw. Malpractice lawsuits take years to settle. Even the most recent records were too old to be considered "news." They could have stopped before they even started.

Getting comfortable

How comfortable are you with the analytic techniques you plan to use? We've all had the experience of exploring a new data set, only to wait minutes, hours or even overnight for queries that didn't show us anything of value. Consider testing your queries and analysis on a small set of records.

In this case, you'll usually want something at least moderately scientific. The reason is that you want some hint of what you'll find and, simultaneously, a list of possible problems to check. Do different data entry clerks (or inspectors) enter the data differently? Are fields you care about usually filled out? And what is the scale of the numbers you're expecting to find: one out of a million, or one out of a hundred?

What could compare?

Do you have any basis for comparison? Make sure, early on, that you've thought through a few possible leads for your story. It doesn't matter what your suspicions are, just how you'd like to word them if they prove true. Look for words like "more than" or "compared with" in that short statement.

For example, imagine you're examining trends in people who return to prison over and over again. You want to know if they're more likely to test positive for drugs, be in a gang, or be of a certain ethnic group than other felons.

Think carefully about the words "more likely" in that sentence. The biggest mistake in building databases – whether they're sampled or not – is ignoring this crucial issue for the paper. It implies that you need to test not only the group you care about but everyone else as well. Make sure, early on, that you could write this sentence based on the records you plan to acquire or the slice of a database you plan to focus on. Sampling will help when you know you can't write the sentence based on your test.

Once you've decided to sample, the next step involves figuring out how to sample a set of data. We'll tackle that topic next month.

Sarah Cohen can be reached by e-mail at sarah@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 a downloadable order form, point your browser to *www.ire.org/resources/nicar/conferences/indiana/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. For more information, point your browser to *www.ire.org/resources/conferences/neworleans/audio.html.*

Campaign Finance Information Center

Campaign finance data from 13 states is freely downloadable from the CFIC at *www.campaignfinance.org.* We also have links to thirteen 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*

Database Library welcomes ...

The Database Library recently welcomed two new members to its staff: Jason Grotto and Dawn Fallik, both incoming graduate students at the University of Missouri School of Journalism. Grotto, born and raised in Chicago, started his journalism career at the small independent monthly *Catalyst: Voices of Chicago School Reform.* He helped launch and then directed Catalyst's Web site. Fallik, a 1992 graduate of the University of Wisconsin-Madison, has been a reporter for *The Record* in Troy, N.Y., and the Associated Press in Philadelphia, Omaha and Baltimore. She has won state and national awards for investigative reporting, spot news and column writing.

Risky Behavior Workshops

The next Risky Behavior Workshop will take place Sept. 11-13 in Kansas City. Join IRE and NICAR for a special workshop focusing on reporting about lifestyles and health. For more information, see the Web site at *www.ire.org.*

Relevant health Web site links and sample data sets used at the Jamuary 1998 Risky Behavior Workshop in St. Louis are now available online at *www.ire.org/resources/conferences/training/otr/healthcare-9801/links.html.* The data comes in two formats: *.SAV, which is compatible with SPSS, and *.DBF. The data include demographics on AIDS, health care coverage, and seat belt usage, and survey questions used by the Centers for Disease Control & Prevention.

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

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