

Uplink

March 2000
.....

INTERNATIONAL CAR

Around the globe

By David Wethe
IRE and NICAR

Although the use of computer-assisted reporting is growing in countries outside of the United States, the issue of access to information – particularly electronic information – remains a concern.

From South America to Scandinavia, journalists say more and more information is becoming available electronically, but it is either aggregated at a municipal or county level or is high-priced from a private company.

To remedy the situation, journalists in

European Union member countries are pushing for the EU to develop a Freedom of Information law similar to that in the United States.

In the meantime, aggressive and dedicated journalists continue to publish and broadcast compelling and powerful stories using computer-assisted reporting.

In this issue, we've included snapshots of some of the work being done around the world. Here's a glimpse:

Netherlands

The origin of CAR in the Netherlands came in 1994 when the *NRC Handelsblad* became the first in its country to publish the results of an election mapping project.

The *Handelsblad* displayed in full color graphics how each party finished according to individual provinces. The idea drew the attention from other media outlets in radio, TV and magazine and is now considered a standard for publishing election results. Yet despite the recent success, the public is still slow to catch on to the importance of CAR projects. Sometimes projects fail to draw any remarks at all from readers, said Dick van Eijk of *NRC Handelsblad*.

A big problem for Dutch journalists is the lack of independent oversight in making sure public data is accessible to everyone. Although most government information is offered in electronic format, it is still too expensive for most newspapers to obtain. The most expensive data is anything on companies or homeowners, van Eijk said.

A recent coup for Dutch journalists, however, was winning a court case that deemed public schools data as public information.

Continued on page five

Inside Uplink

The wonders of computer-assisted reporting have been spreading around the globe the past few years and have yielded compelling and powerful stories in both print and broadcast outlets. This month, *Uplink* has snapshots of some of those efforts in countries ranging from Brazil to the Netherlands.

SEE PAGE TWO.

Nursing homes

Dave Gulliver, of the *Dayton Daily News*, tells how he went beyond OSCAR – the federal database used frequently to report on nursing homes – to show how nursing homes were failing the elderly, and particularly the poor elderly, in Ohio.

SEE PAGE SEVEN.

Mapping

This month, *Uplink* launches the first in a series of columns about using mapping software. Andy Lehen, of Dateline NBC, and Jennifer LaFleur, of the *St. Louis Post-Dispatch*, will alternate writing the column for each issue. The first story, by Lehen, discusses how to make sure you're using the right datum measurements.

SEE PAGE NINE.

SCHOOL TEST SCORES

Lake Wobegon effect

By Drew Sullivan
The Tennessean

Each fall, the state of Tennessee releases its school test scores and every year *The Tennessean* dutifully runs three open pages of scores and graphs showing how local schools are doing. One perennial result is that Tennessee's students score above the national average in most subject areas, a fact trumpeted by the state department of education.

Last fall, we decided to look into that claim. What resulted was a quick one-week story that called into question the basis of this claim. We also learned an important lesson on how information is presented and the importance of questioning basic assumptions.

At the heart of the matter is the way standardized tests are done nationwide.

State assessment programs often include two elements: a norm-referenced test

Continued on page thirteen

Uplink

March 2000

Volume 12, Number 2
A newsletter of the National
Institute for Computer-Assisted
Reporting

EDITOR

Brant Houston

DIRECTOR OF

PUBLICATIONS

Len Bruzzese

MANAGING EDITOR

Mary Jo Sylwester

ASSOCIATE EDITOR

Jason Grotto

ART DIRECTOR

Kerrie Kirtland

SUBSCRIPTION ADMINISTRATOR

John Green

Uplink is published every month by the National Institute for Computer-Assisted Reporting, 138 Neff Hall Annex, Columbia, MO 65211. (573) 882-0684. Subscriptions are \$40 for IRE members, \$60 for nonmembers.

Postmaster: Please send address changes to NICAR. Send e-mail to jgreen@nicar.org

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

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

VENEZUELA

Political Census

By Carlos Subero
El Universal

To know how politicians think at a given moment is valuable to readers in a country like Venezuela, which is going through a process of rapid political changes. But how does a news organization inquire into this?

The solution to this dilemma was an issue raised by the Research department of *El Universal*, a Caracas-based national journal. We designed a 32-question census addressed to a group of candidates to the National Constituent Assembly with chances of being elected in the last July election. This enabled readers to know the opinions of the Constituent Assembly that they had elected, as of the day after the election.

The Constituent Census was an unprecedented journalistic experience in the field of elite studies in Venezuela.

The Constituent Census was an unprecedented journalistic experience in the field of elite studies in Venezuela. It was possible because of the automation of the Venezuelan official ballots by the National Electoral Council. Just as it had been anticipated, the names of the 128-winning candidates were known merely two and a half hours after the balloting centers had closed.

How it started

The questionnaire, handed out to the candidates during the two weeks before the elections, contained topics relating to the national politics, some value features and socio-demographic aspects such as age, state of birth, sex and social origin.

The editor-in-chief's office had expounded to the researchers that its intent was to publish the day after the elections. Therefore, the vital issue that had to be addressed in advance was to determine

who would be the winners amongst more than one thousand candidates.

Such a delicate issue was successfully resolved by resorting to some political reckoning and a little luck.

The decision on how many and what candidates had to be interviewed was made one month before the elections. It was calculated that two candidates for each of the 128 posts would result in 256 interviews, a reasonable amount to be handled by the team of 11 reporters from the Research department and the Fundación Andrés Mata.

The open-list, simple majority electoral system was used. To decide who would be the interviewees, the thesis initially adopted was to choose all the candidates identified with President Hugo Chávez, who had a very high popularity ranking. The 128 candidates were chosen from the opponent and independent party factions.

In Venezuela, candidates are willing to attend to the press on the phone, even to respond to a 32-question query. Consequently, the most difficult step in accomplishing the interviews was to get the phone numbers of the 256 candidates. Only about 10 of the candidates chosen could not be interviewed, whether because they declined to grant the interview or because they could not be contacted.

Results

The electoral result was surprising. President Chávez supporters won 122 of the 128 posts, and the opposition won only four, two were independent: only one of the candidates elected had not been included on the initial list of interviewees. Out of the winners, we interviewed 92 percent.

Two reporters downloaded the questionnaires coded using the SPSS processor. Thus, the database was created with 42 different variables, with data on the answers of 256 candidates to the Constituent Assembly.

Within two hours of the release of partial election results, our final database had already been built in SPSS.

The first day, the results of four questions were published; these questions allowed us to figure out that 97 percent of the candidates elected was thinking of implementing the presidential re-election.

Continued on page six

Making CAR progress

By Nils Mulvad
DICAR Committee

It has been four years since three Danish journalists attended bootcamp at NICAR in Columbia, Mo., where we learned how to use databases and spreadsheets in our work.

We had many discussions of how these brilliant methods could be used in Denmark, where we do not have the same tradition and access to electronic data. Even though the United States and Denmark are far apart, I must say that these methods have survived the emigration.

In April 1999, we founded DICAR, Danish Institute of Computer Assisted Reporting, which is located at the Danish School of Journalism in Århus. Here we share premises with the journalism students making their own Internet paper, Byline.

A lot has happened during the past year. The number of members of the Foreningen for Computerstøttet Journalistik (Association of Computer Assisted Reporting), which was founded six months ago, has now reached 200.

How far we've come:

- This spring we start teaching senior journalism students a ten week CAR-module.

- We have made an in-service course program in cooperation with DjE, The Danish Institute of Journalistic Supplementary Training. The program includes training in spreadsheets, databases, the use of Internet and other computer-based tools tailored for journalists. This is probably the beginning of a long-term collaboration with DjE, and it has already resulted in the hiring of a part-time course leader.

- Next month, we will gather for the third CAR-day, and we have a longer program than at previous CAR-days. We also arrange conferences and often introductory speakers at other events.

- The past 18 months, we have been running a web-site, www.dicar.dk, which is continuously updated with new services for members of Foreningen for Computerstøttet Journalistik and services which are available to non-members.

- At the DICAR home page we are

building tool boxes for journalists within different beats or subject areas..

- We are starting a bank of databases for our members.

- We have hired five students on a part-time basis to solve different assignments for DICAR.

Daily use of CAR

The number of media using CAR methods is steadily increasing, especially those using spreadsheets and self-made databases. But as always, there is a big step from theoretic lectures to using the methods on a daily basis for stories.

We suffer from the fact that the media managers are very reluctant when it comes to new methods. It hurts CAR, but also many of the other good ideas of how the future journalist should work. It really shows in Denmark at the moment, where the journalism craft is affected by the recession in Denmark and the resulting resignations and changes in management. But some editors are slowly recognizing the advantages of CAR.

**Our experience is that
the authorities actually
listen as we get access to
still more data, and we
seem to have a good
dialog with a number of
authorities.**

In connection with the CAR work we have published, there have been a number of reports to authorities concerning electronic access to databases.

Our experience is that the authorities actually listen as we get access to still more data, and we seem to have a good dialog with a number of authorities.

We are still in good contact with the pioneers from United States and with some of

Continued on page six

Web site for Brazilian
newspaper Folha de Sao
Paulo: www.uol.com.br/fsp/

Web site for NRC
Handelsblad: www.NRC.nl

Web site for DICAR:
www.dicar.dk

SWEDEN

White collar crime

By Helena Bengtsson and
Jenny Nordberg
Sveriges Television

"A stealth plane shooting on doves" – that's how the most expensive Swedish effort to fight financial crime in years turned out to be. Instead of conducting their own investigations to track down the organized criminals, they sat back and dedicated their time to prosecuting small business owners with bookkeeping trouble.

For decades, the Swedish justice system has been struggling to come to terms with organized financial crime: tax fraud, inside information crime, offshore decoy companies etc. Through several police and attorney office reorganizations, the white-collar criminals have been operating without much interference. With an estimated cost to society of more than 100–150 billion Swedish Kronor (\$120 million) a year, the minister of justice decided to create an all new agency: "Ekobrottsmyndigheten."

**When the reporter
interviewed the
responsible officials, they
were stunned by our
conclusions, and
appropriately
embarrassed.**

Four hundred policemen, 90 specially-trained prosecutors, and a number of tax specialists were installed in new offices in Sweden's three largest cities. With higher salaries than the average police officer and attorney, the most competent staff were recruited to the new "super-department," which got an extensive government budget. After operating for a year and a half, our station's current affairs program "Striptease," decided to examine the results of the costly initiative.

Getting started

The investigative team got copies of all the complaints from the three agencies

throughout Sweden. Only available as paper copies, they created large piles at the researcher's desk. An Access database was created in order to get a picture of the most widely prosecuted crimes and individuals. We wanted to know who they were, and if the Swedish authorities finally had managed to track down the most advanced financial criminals.

The essential facts of each defendant was typed into the database, including the name, age, occupation and personal id number. When several suspects were to be tried in the same case, they each got a record in the database. Also, the classification of the crime and a short description of events leading up to the prosecution. We made a note of whether the suspect had been arrested or not, which is a determinant on how severe the crime had been considered by the prosecutor.

Finally, we entered the estimated amount of money withheld and how the crime had come to the police's knowledge – by a bankruptcy report or the tax authorities revisions. In a few cases, it was an independent investigation leading up to the complaint. Some also contained a number (1-3), grading the complexity of the crime by the agency themselves. We adapted their system, grading all the cases accordingly.

The results

Eight-hundred-four suspect profiles and 625 complaints later, we started to analyze the results. By asking questions to the database, we summoned up 18 months' work by the most advanced financial crime unit in Sweden.

First, we got the average criminal; we were able to see that he (very soon, it was clear that he was a man) was a middle-aged Swedish citizen with a small business, having some trouble keeping his books in order. Moreover, he had managed his business badly and gone bankrupt.

Tax authorities discovered some forgotten receipts in a shoebox, and he was taken before the court and accused of tax fraud. Hardly the polished, pinstriped criminal in dark shades taking weekend trips to a tax

Continued on page six

Recent CAR stories:

Copies may also be
available in the IRE

Resource Center. Check
database at

www.ire.resourcecenter.org

"Friends of the Court"

The Miami Herald

Jan. 23, 2000

[www.herald.com/content/
archive/news/browcourt/docs/
020267.htm](http://www.herald.com/content/archive/news/browcourt/docs/020267.htm)

The Herald's analysis of
computerized Special
Public Defender billing
records for 1993 through
October 1999 found that
Broward County judges are
selecting among their
lawyer friends to hire
private lawyers to
represent poor people who
can't afford attorneys.

"Out Of Control: Drunken
Driving in Kansas City"

The Kansas City Star

Dec. 25-26, 1999

www.kcstar.com/projects/dwi/

The Kansas City Star
analyzed 5.7 million
Missouri driver's license
records and 3.6 million
traffic convictions. The
analysis focused on alcohol-
related traffic convictions
and covered municipal and
state offenses.

Continued from page one: Around the globe

Van Eijk has used CAR to look at topics ranging from royal decorations to attendance and voting of European Parliament members. He analyzed where the 2,000 recipients of Queen's Day Royal Decorations lived and used maps that show that the higher decorations were granted in different areas than lower ones.

Prior to the European Parliament elections, Van Eijk analyzed how often members of the parliament had been present at meetings and how they voted during the past five years. National background turned out to be a key factor. The Dutch were present and voted most often while Italians were least often present.

Denmark

Three Danish journalists first learned how to incorporate databases and spreadsheets into their everyday work when they attended a 1996 CAR bootcamp in Columbia, Mo. Last April, the trio started the Danish Institute for Computer-Assisted Reporting. Since then, DICAR has attracted more than 200 members.

Nils Mulvad, who helped start DICAR, said that although most Danish editors are reluctant to change, CAR is starting to catch on. Mulvad said his current goal is

to forge a joint initiative among other journalists to fight for electronic access to documents. (Read more about DICAR on page three.)

Brazil

Reporter Jose Roberto de Toledo of the Brazilian newspaper *Folha de Sao Paulo* said most journalists are still slow to catch on to CAR in his country. In fact, many government agencies are more advanced with electronic files than journalists, he said.

Despite that, de Toledo has published several crime stories with the help of CAR. He constructed a database of more than 100 cases of multiple homicides for 1998 as a way to profile the victims. By using other crime data, de Toledo was able to offer for the first time a list of the Brazilian cities with the most homicides.

Like in other foreign countries, Brazilian readers' interest in CAR stories has been limited to such hot topics as violence and unemployment. "If the article has too many numbers, rates, percentages and little ideas, it tends to be boring," de Toledo said.

David Wethe can be reached by e-mail at davidw@nicar.org

INTERNATIONAL CAR

Finding data

By Noemi Ramirez
IRE Resource Center

One of the reasons to explain why CAR is underdeveloped in countries other than the United States is the absence of computerized electronic databases. At meetings with reporters from all over the world, I have been asked this question on several occasions: "I'm green with envy, we don't have those resources in my home country ... what can we do?"

My answer is twofold. You can reenact the work of medieval monks — or nuns — by manually creating your own databases, or you can search for information about your country somewhere else outside the national borders. There are a number of databases and

online resources kept by American public agencies, companies and institutions that can be helpful when doing international research.

Business

Securities and Exchange Commission Data, Free Edgar: www.freEdgar.com/
10K Wizard: <http://10kwizard.com/>

These web sites offer a free full-text search of SEC filings as well as other useful tools to background every publicly traded company in the United States, no matter whether the company is American or not. Both Free Edgar and 10K Wizard include information — such as the registration statement or the financial annual reports — that is easily

Continued on page ten

International data sources:

• Securities and Exchange Commission Data.

Free Edgar:

www.freEdgar.com/

10K Wizard: <http://10kwizard.com/>

• Center for Responsive Politics:

www.opensecrets.org/

2000elect/lookup/

AllCands.htm

• The International Data Base. Census Bureau's International Programs Center:

[www.census.gov/](http://www.census.gov/ipcl/www/iidbnew.html)

ipcl/www/iidbnew.html

• Private International Law Database. Office of the Assistant Legal Adviser for Private International Law. Department of State:

[www.state.gov/www/global/](http://www.state.gov/www/global/legal_affairs/private_intl_law.html)

[legal_affairs/private_intl_law.html](http://www.state.gov/www/global/legal_affairs/private_intl_law.html)

• World Military Expenditures and Arms Transfers 1997. U.S. Arms Control and Disarmament Agency:

[www.state.gov/www/global/](http://www.state.gov/www/global/arms/wmeat97/)

[arms/wmeat97/](http://www.state.gov/www/global/arms/wmeat97/)

[wmeat97.html](http://www.state.gov/www/global/arms/wmeat97/)

More international data sources:

INS Annual Statistical Reports

www.ins.usdoj.gov/graphics/aboutins/statistics/index.htm

The Immigration and Naturalization Service provides detailed statistical data about immigration trends in the United States. It includes useful information such as refugee-status applications by geographic area and country.

Federation of American Scientists

www.fas.org/

This Web site offers an impressive collection of reports on a variety of issues such as animal health diseases, cyberstrategy, biological and toxin weapons, intelligence and nuclear resources.

Continued from page two:

Venezuela

The SPSS crossing of the variables on family origin and the socio-educational level of the constituents resulted in one transcendent finding. It has been determined that a half of the constituents appointed, the new dominant political class, meets the condition of being a university professional who comes from the low class or the middle low class. This condition was met only by 37 percent of the Congress elected 8 months ago. None came from the high class.

Certainly, the full potential of SPSS was not used. But some indices were created relating to several questions on the military topic and that of the decentralization policy, which permitted us to determine a high tendency to

the political participation of the military, and also towards decentralization.

The Constituent Census was the second study on political elites that *El Universal* conducted during 1999, and one that has been an innovative contribution of precision journalism to the political sciences for consultation at different universities in this Latin American country. A third study is pending with the election of the new Congress, which will provide a more sound knowledge on the quality and thought of the political elite and the change process that is taking place in Venezuela.

Carlos Subero can be reached by e-mail at csubero@callisto.eud.com

Continued from page three:

Denmark

our colleagues in other European countries.

We hope to extend this cooperation, especially with the idea of establishing joint initiatives throughout Europe concerning the right to electronic access to documents. In this connection we have worked out an English-language description of the current status of the right to electronic access of documents in Denmark and a description of the new rules from EU concerning personal data. (If you are in-

terested in the descriptions you can e-mail dicar@dicar.dk.)

Last, but not least, we are having a great time together while working cooperatively on DICAR-projects.

These days we are beginning to take out re-subscriptions for the first time and of course we are anxious to see how that goes.

Nils Mulvad can be reached by e-mail at nmulvad@po.ia.dk

Continued from page four:

Sweden

paradise island to place money in secret bank accounts. Still a definite crime, these were cases that the newly examined attorney at any court could have managed to prosecute.

As it turned out, only six percent of the cases were the top-notch crimes, the bigger part of the administration were simple book-keeping incidents. Taxi drivers and cleaning companies trying to escape some of the value tax were dominating our database. When the reporter interviewed the responsible officials, they were stunned by our conclusions, and appropriately embarrassed. The minister of justice, however, would not grant us an interview, hoping that the storm would blow over soon.

"Striptease" is Sweden's most influential

investigative television program. The title aspires on the fact that power and corruption should be stripped down to the bare truth.

The program has used CAR for a number of stories recently. One investigated the surprisingly large number of Swedes possessing weapon licenses, who are ex-convicts or suffer from severe mental illnesses. In that case, the weapons register was joined with court statistics and hospital records of people receiving treatment for depressions, schizophrenia and alcoholism.

Helena Bengtsson can be reached by e-mail at helena.bengtsson@svt.se

Going beyond OSCAR

By Dave Gulliver
Dayton Daily News

Grace Blankenship was getting some of the best care money could buy, making friends, talking about her favorite players on the Cincinnati Reds. Then her money ran out and everything changed.

She was moved from a facility's assisted living wing to a standard nursing home unit. There, a younger, demented male resident attacked her, threw her against a wall and broke three of her ribs.

Two weeks later, late one night, she began vomiting and went into shock from severe dehydration. By dawn, she died, with her doctor and family not notified until the last hour of her life.

Mrs. Blankenship's story exemplified two main flaws in Ohio's system of caring for its elderly: If you don't have the money, or if it runs out, your only choice is nursing homes that regularly violate standards, with virtually nothing to fear from the state.

But you need data, not anecdotes, to prove system-wide problems exist, and even more data to explain why.

Starting out

In early 1999, the *Dayton Daily News* decided to look at how Ohio was prepared to handle the booming elderly population. That eventually focused on the available choices for long-term care, and nursing homes are by far the biggest player.

**Inspections,
demographics and
enforcement were our
workhorses, but several
other databases chipped
in to the project.**

Nursing home projects typically use federal OSCAR data, the findings of health and safety inspections, to search for problem nursing homes. For us, that was just the start.

We used the state's own inspection data, which is used to create OSCAR. We wanted to undercut any "one-bad-year" arguments, so we used all the existing data.

Inspection data lets you look at types of problems— from dirty silverware to resident abuse— and whether they're widespread or rare, dangerous or mostly harmless. By tallying the citations, you can develop a crude ranking.

But that alone can be deceptive. For example, a home with one of the highest citations totals was solely for Alzheimer's patients. Advocates argue that such homes should have more staff and better training to ensure better care.

In the end, you need to know more to make valid comparisons. Demographics data was the answer. We used another Health Department data set, the annual survey of long-term care facilities. That allowed us to link citations totals with other key fields. The most important was a nine-category breakdown of ownership type, such as for-profit partnership or church-related non-profit. We also used variables such as age, race and gender breakdowns, dedicated Alzheimer's units, nursing staff per resident, owner and operator, and so on.

We linked the tables in FoxPro, then imported the result into SPSS for the serious number-crunching. We filtered the data for "apples-to-apples" comparisons of similar types of facilities, and used medians and trimmed means where appropriate to account for skewed data. SPSS's crosstabs and nifty scatterplots and histograms made the job infinitely easier, but (as we learned the hard way on a redlining story years ago) you could get to the same results with database and spreadsheet software.

Follow the process

The inspections data — the beginning and end for many projects — was just half the picture. The most surprising findings came from the state's enforcement database.

We found that even when state inspectors found problems, they gave homes little reason to clean up their acts.

The enforcement data was far more complicated to use than the inspections data. It was drawn from a system designed for quick look-ups and case management, not aggregate analysis. That later became apparent when I asked

Continued on page eight

The *Dayton Daily News* series is available online at www.activedayton.com/partners/ddn/projects/1999/Elder_Care/

The story is also available in the IRE Resource Center, story #15917.

Other stories in the IRE Resource Center on nursing homes:

"Warehousing the mentally ill," (story #15173), by Michael Berens of the *Chicago Tribune*, 1998.

"A crisis in care," (story #15500), by Geoff Dutton of *The News-Journal* in Daytona Beach, Fla., 1998.

WDIV-TV (Detroit) investigation on nursing homes, 1996 (story#13696).

To order, call the Resource Center at (573) 882-3364.

Tipsheets in the IRE

Resource Center:

"Money and paper trails of health care," (#1011) by Michael Berens of the Chicago Tribune, for the 1999 IRE Conference.

"Investigating nursing homes," (#1009), by Alison Young of the Detroit Free Press for the 1999 IRE Conference.

To order, call (573) 882-3364.

Continued from page seven:

Nursing homes

the enforcement chief if the basic numbers were correct: 1,261 enforcement cases, 334 penalties imposed. "I have no idea," he replied. The state couldn't readily compile its results, so it just didn't. (They later did their own analysis and confirmed the numbers.)

Because of the database design, the enforcement data had even more pitfalls than the inspections data. There were multiple records for each case, one for each type of penalty proposed, and most homes had multiple cases. Typically a few penalties stuck but most were dropped.

The key was to break the table down and rebuild it several ways. For example, to see how the state handled "repeat offenders":

- Add a field to the original table to flag when a penalty stuck.
- Group the facilities by ID and case number to identify how many cases the state launched against a particular home, and order them by date of case.
- Copy the result to a new table.
- Add and fill an "action number" field. We used a FoxPro SQL program.
- Then join to the original table to add the flag field. Result: How many "strikes" a home received before the state would call it "out."

Dealing with dirt

Like the enforcement data, most of what we used wasn't meant to be used that way.

The Health Department assigns each nursing home an ID number, which is supposed to be the standard way of identifying a home. But the inspections data is tailored for inclusion in OSCAR data, so the quality assurance bureau uses the home's Medicare number as the unique—and only—identifier. Meanwhile, another bureau maintains the demographics table. It uses the department ID field as the identifier. It also has the Medicare ID number, but frequently fills it with a dummy number. We eventually decided to capture as many homes as possible, and to play it safe: We recoded ID numbers by hand.

But simple dirt in the data wasn't the biggest problem. We had to understand how the system had changed over time. For example, in mid-1995, the feds and the state changed what counts as a serious health or safety violation, "substandard care" in health-speak.

If we didn't account for the differences, our analysis would have been garbage. We wrote another SQL program to recode the citations in different ways—to grade them by their respective rulebooks and to grade them by current and former standards.

Use "golden" databases

Inspections, demographics and enforcement were our workhorses, but several other databases chipped in to the project.

Perhaps the best was a little in-house Excel worksheet from the Attorney General's office that summed up abuse prosecutions.

The state's deaths database essentially identified seven cases of nursing home fatalities. Victims' names are redacted from incident reports, but provide dates, locations, and sometimes gender and age—enough to find the death certificate and the name.

We also drew on the state's complaints hotline database; inspections data for assisted living facilities and group homes; and demographic data for home health agencies. Late in the process, when we'd almost given up, the state provided Medicaid Cost Reports data, which we used to check other findings.

In short, getting the data was relatively simple—but understanding it well enough to run the numbers took months. By the end, we could talk Medicaid-speak with the best of the bureaucrats.

So what did we learn?

- Search for data where you don't expect it.

The project started with a Health Department flak saying they didn't keep an inspections database, but could provide paper summaries (sound familiar, anyone?). We went straight to inspectors to see what really existed. We never expected the Attorney General to have abuse cases in electronic form, but there they were. We didn't expect to get complaints data, but after agreeing to redact names, there it was.

But know when to stop searching. We probably gathered more than we need. Several databases made only a cameo appearance in the final project. When you have enough to answer your questions, stop there and dig in.

- Learn everything about the data.

CAR 101: Know the definitions of every field in the database, and every code in the

Continued on page fifteen

The world's not flat

By Andy Lehren

Dateline NBC

It sounds obvious: the world is round.

Aristotle knew that back around 350 B.C. Ptolemy refined it a few hundred years later. Despite fears from the flat earth believers, Columbus made a living off the idea.

The problem is knowing the exact shape of our sphere.

You don't have to be sailing for a new world to face the question. All you have to do is get a few disparate maps and try to make them work in your GIS program.

There are two fundamental questions to ask when you get a map from anyone: what's the datum, and what's the projection. There are other details that may be helpful, but these are the most important. This column will try to explain datum, and how to convert it when you have to.

Some more history: Isaac Newton wrote in the late 1600s that, well, the earth's a bit flattened on the top and bottom. The French, after measuring their country around the same time, countered that it's more egg shaped. Newton won the debate. The French ate better food. And we live with the question: just how flat on the top and bottom?

Ever since, geodesists have been measuring this. There are hundreds of calculations that have been done figuring out our little ellipsoid. Here are the three that matter:

- Clarke 1866, developed by Alexander Ross Clarke. He was a busy guy, so the year is important.
- Geodetic Reference System 80 (GRS 80), a group effort by an international geodetic organization.
- World Geodetic System 84 (WGS 84), founded by the U.S. military and close to GRS 80, but not exactly the same.

These different measurements of our spinning ellipsoid are the underpinnings for almost all the maps we look at. You can tell from the history that this is still evolving, and you may well face other datums as you do more mapping.

You May Not Be Where Your Map Says You Are

You would think that something like latitude and longitude is absolute. It's not.

First, different geodesists came out with

different calculations on how elliptical is our planet. If you try to draw three circles by hand, you'll have an idea of the differences.

These ellipsoid measurements are the underpinnings for datum. You may be most familiar with datum by seeing latitudes and longitudes, though it's more than that.

If geodesists draw up different spheres, the datum will be different.

It gets murkier. To develop datum, geographers went out and mapped our country. They did not measure every latitude and longitude. They measured a lot, and extrapolated a lot more. They used the ellipsoid calculations as their reference guides for figuring datum. In other words, there's some guess work piled on top of some guess work.

Early this century, geographers used the Clarke 1866 ellipsoid, and result was something called North American Datum 1927 (NAD 27), with the center in Meades Ranch, Kansas.

It had problems. They measured again later that century. This time they assumed the Earth was shaped more like the GRS 80 ellipsoid, and came out with something called North American Datum 1983 (NAD 83).

And now people are running around with GPS units measuring where they are, and assuming the Earth is shaped like the WGS 84 ellipsoid.

That's how we've gotten the three major kinds of datum that are the foundations for the maps we use.

Three different ellipsoids were used to derive three sets of datum. And that means three sets of latitudes and longitudes.

You pull out a GPS, and it can tell you your latitude and longitude. Well, not exactly. It says where you're at based on WGS 84. That's the flavor preferred by most GPS devices. That latitude and longitude may be off by half a meter or more using GRS 80. And it may be up to 300 meters off under Clarke 1866.

If you don't know the background of your map, you could use a GPS, look at a map, and realize you may not be exactly where you think you are.

Converting Your Maps

OK, sometimes your maps will be in

Continued on page twelve

Editor's note: This is the first in a series of columns about mapping that Andy Lehren and Jennifer LaFleur, of the St. Louis Post-Dispatch, will write for Uplink.

Census workshops

Recent CAR stories:

(Stories might also be available in the IRE

Resource Center. Check by searching the database at www.ire.org/resourcecenter)

"Grading Our High Schools"

The Pittsburgh Tribune-Review, Feb. 6, 2000

<http://triblive.com/frames/grades.html>

According to the series' author, this investigation "features a regression analysis using average state assessment test scores and participation in the free and reduced-price lunch program."

"Broken Trust"

The Charlotte Observer Jan. 23-27, 2000

www.charlotte.com/special/mentalhealth/

The Observer identified deaths of the mentally ill and developmentally disabled by searching 370,000 N.C. death records from 1994 through mid-1999, along with hundreds of police and autopsy reports.

IRE and NICAR will hold several two-day workshops on working with Census data, with the first being held in Arizona April 7-9.

The Arizona workshop, which will be held at Arizona State University in Tempe, will have Steve Doig, of ASU, and Paul Overberg, of *USA Today*, as lead instructors.

The second workshop will be held at the Missouri School of Journalism, in Columbia, August 4-6. The third workshop will be at the University of Maryland in December, but a specific date had not been set as of publication.

These Census workshops will be the first in a series and cover getting ready for the Census, what changes there will be from 1990 data, and ideas on the first stories to do and how to do them.

The initial fee for the workshops will be \$75 for two days if you are an IRE member. Membership will be offered at a discounted \$40 for a total of \$115 if you are not an IRE member. (IRE membership fee will go up to \$50 in March.)

More information is available on the IRE and NICAR Web site, www.ire.org

Continued from page five:

Finding data

turned into Excel spreadsheets for further analysis. One extremely useful feature is the electronic alert service that will send e-mail notifications when any of the companies in the Watchlist submits an electronic filing to the SEC. For an explanation on using Free Edgar and 10K Wizard, request tipsheet #1138 from the IRE Resource Center.

Campaign finance

Center for Responsive Politics, www.opensecrets.org/2000elect/lookup/AllCands.htm

The Center for Responsive Politics defines itself as "a non-partisan, non-profit research group based in Washington, D.C. that tracks money in politics, and its effect on elections and public policy." The donor lookup database retrieves individual and soft money contributions that are above \$200.00. To find out how much money employers of foreign companies with offices in the United States are giving to the American presidential candidates, try to enter the name of a company in the Occupation/Employer field.

Demographics

The International Data Base. Census Bureau's International Programs Center, www.census.gov/ipc/www/idbnew.html

The International Data Base (IDB) contains statistical tables of demographic and socioeconomic data for 227 countries and areas of the world. For some countries, the data goes back as far as 1950 and as far ahead as 2050. You can download a zip file that

contains the files and instructions necessary for installing the actual International Data Base in your computer, or you can import specific tables into most spreadsheet programs.

Law

Private International Law Database. Office of the Assistant Legal Adviser for Private International Law. Department of State, www.state.gov/www/global/legal_affairs/private_intl_law.html

This Web site provides information on current negotiations and projects covering the private international law of such areas as trade and commerce, finance and banking, trusts and estates, family and children matters, and international judicial assistance.

Arms

World Military Expenditures and Arms Transfers 1997. U.S. Arms Control and Disarmament Agency, www.state.gov/www/global/arms/wmeat97/wmeat97.html

The ACDA collects data on international arm sales, arm treaties, and military expenditures.

The annual reports include .PDF tables by region, organization, country, major supplier and recipient country. For a quick tip on how to import .PDF tables, check Tom McGinty's article on the July/August 1999 *Uplink* issue.

Noemi Ramirez can be reached by e-mail at noemi@ire.org

Nulls and averages

By Neill A. Borowski
The Philadelphia Inquirer

If you've done data analysis for a while, you've probably fallen victim to the nulls.

You open a new database and discover several fields that could make your analysis richer and deeper. And then you test for nulls, the signs of missing data that plague many databases, particularly databases used by the government for administrative purposes.

Those great fields are missing thousands of records. Can you do anything to salvage them and come up with summary statistics, such as averages and medians? Or, do you have to throw out the field and, with it, the chance for some insightful analysis?

Trying to find an average for a variable when one-third of the records are missing can be misleading and dishonest. The tendency might be to report the summary statistics, with a footnote disclosing that these are the averages for cases in which the data were reported.

Trying to find an average for a variable when one-third of the records are missing can be misleading and dishonest.

However, any suggestion that the average or median or other summary statistic represents all of the data is misleading. We are working with what should be the universe of the data – all of it. But some is missing. We could try some statistical tests that would draw a random sample of the data we have and shed some light on the probability that the true mean would fall within a range. Yet, the random sample of the incomplete data wouldn't necessarily represent a random sample of all of the data.

Possible effects

Consider a database with 500,000 records. The field of interest is missing 20 percent, or 100,000, of the records. You average the

400,000 records you have and come up with a mean of 5.0. Is this "good enough" to use?

The missing 100,000 records suddenly are found (they slipped under a dusty desk somewhere). The average for the missing 100,000 is 2.0 and, when combined with the 400,000, yields a mean of 4.4. If the average of the missing 100,000 were closer to the average of the 500,000 records – say 4.0 – then the overall average with the missing records replaced would be 4.8 – much closer to 5.0.

Does a run of missing data occur often? It has with us.

We recently worked on a one-million-plus record database that is a collection of data submitted by hundreds of different entities. The database, which spans two years, has several fields with thousands of records of missing data. Some of the entities didn't file data with the collecting agency for the fields that we found most interesting.

Are there sweeping averages or summary statistics (race, age, etc.) that might be calculated from the fields with missing data? Yes. But I'm not comfortable that they would necessarily be representative of the "whole."

Checking for nulls

This is why it's so important to check for nulls, or missing data. One might run the averages or medians and never discover that some records are missing. SPSS will tell you how many records were included in the calculation and how many were missing. But basic programs such as MS Access will simply calculate the average with no alarm for nulls. The "Is Not Null" condition should be used in the average calculation in Access, but this still will not report to you how many nulls exist. Before you calculate the average, count for nulls.

And, heaven forbid, the nulls somehow were transformed into "0's" in the database. Most programs then will calculate them into the average or the median and – depending on how many there are – come up with a useless number.

What can you do if you come up with a lot of missing data?

One way is to see if a subset of the database has fewer records missing (or none at all).

If the database is for two years, test for nulls

Continued on page twelve

If you've discovered some interesting methods of overcoming missing data, e-mail Neill Borowski at nborowski@phillynews.com and he will share the tricks and tips with everyone.

Some recent CAR stories:
(Stories might also be
available in the IRE
Resource Center. Check by
searching the database at
www.ire.org/resourcecenter)

**"Understaffed or Poorly
Managed?"**

St. Petersburg Times
Jan. 23, 2000

[www.sptimes.com/Archive/
012300/Pasco.shtml](http://www.sptimes.com/Archive/012300/Pasco.shtml)

The Times' analysis of
sheriff's staffing and
response times showed the
office had deputies
available to answer 80
percent of its priority one
calls. The average delay at
the dispatch center was
nearly four minutes.

"Discarded and Decayed"

The Journal Gazette
Dec. 19, 1999-Jan 2, 2000
www.jg.net/jg/landlords/

A six-month investigation
of Fort Wayne housing
that included computer-
assisted research of health,
housing code and property
found that the county was
one of the biggest violators
of housing ordinances.

Continued from page eleven:

Nulls

in the field in question for each year separately. In the example above, the nulls might have come from a reporting problem with one or two entities in one year. The other year might have 100 percent of the data.

Now you can calculate the summary statistics and report them with confidence for the one year you selected.

Another possible way around the missing data would be to group the data by the entities (towns, police departments, etc.). Are the null fields present for only one or two towns, for example? If the absence of information on this town isn't troubling, calculate the summary

statistics for the field and report in the footnote that Smithville isn't included in the data.

Of course you can select other filtering methods, depending on the database — type of crime, type of property, etc.

The goal is to be able to spot the nulls early on in your analysis. Before you "give in" and calculate summary statistics for interesting fields, test the field for nulls. Also review the field for nonsensical data, such as "0's" when a zero is impossible.

Neill Borowski may be reached by e-mail
at nborowski@phillynews.com

Continued from page nine:

Mapping

NAD 27. And sometimes they are in NAD 83. And when you use your GPS, you're in WGS 84.

In a perfect world, you would try to work in the datum used to make your maps. But that doesn't always happen. You get a map from one agency in one datum. You want to match it against a map in a different datum. Now you need to run a conversion so they line up better.

Here are a couple solutions. These examples are for converting datum only. (Just so you know, the ArcView utility mentioned here can handle projection conversions. There is also a script in older versions of ArcView to switch projections.)

1. ArcView's Projection Utility

The new version of ArcView comes with something called the Projection Utility. You can run it outside the program. Just click on the icon on your desktop.

Once you're in the utility, point and click your way to the shape file you want to convert. Hit the next button. That ends step one. For step two, first check the box that says "Show Advanced Options." We're assuming the data is not projected, so keep the coordinate system at "Geogálphic." Choose the datum your map came in, say North American 1927. Hit the next key. For step three, select the datum you want to convert to, like North American 1983. Hit next. Step four is where you navigate to where you want your converted files.

Now you can hit the next key and then finish. The utility will create a reference file that keeps track of the datum for your new file.

The new ArcView program is based on code written by a group of oil and gas companies. In its manuals, ESRI, the maker of ArcView, does not talk about errors and estimations in running conversions. But the program is not perfect. The petroleum consortium whose work is behind the ArcView utility acknowledges that fact. Your end result may be off by a meter or more.

2. Wessex' TransNAD

If you have an older version of ArcView, or use MapInfo, here's another solution that's free. Go to www.wessex.com. It's the home page for a company that sells street map data and other things.

- From the home page, navigate to utilities, and download TransNAD. It's GUI application that lets you point and click your way quickly to a solution. It only lets you convert between NAD 27 and NAD 83. Unlike the new ArcView utility, this program will not make a new file, but alter your original file. So it's best to work off a copy.

Once you've installed the program, navigate to the file you want to convert. Choose which conversion you need to run, either from NAD 27 to NAD 83, or visa versa. Hit next to run the conversion. You're done.

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

Continued from page one: Lake Wobegon

and a criterion referenced test. The first allows the state to look at how the kids in the state match up to national standards. The second is how kids match up to state standards. It was the norm-referenced test we were interested in.

Tennessee students ranked between the 49th and the 59th national percentile rank depending on the grade level and the subject area of the test. To check whether these numbers were inflated, we asked around about other tests that might corroborate these results. This is always a good technique to check the results of any study or statistic. What we found was the National Assessment of Educational Progress or NAEP test.

The NAEP test, which is administered by the Department of Education, is given in one subject area every couple of years to a sample of students across the country. In 1998, students were tested in reading and in 1996 and 1992 students were tested in math.

The comparison surprised us. Tennessee showed up below the national average in both tests. The discrepancies between the two tests were large. We checked another test. We found Tennessee students scored below the national average in every subject area of the American College Test or ACT given to high school students.

Checking inflation

To verify the extent of the inflation, *The Tennessean* built a spreadsheet of all states that used the same test we did, the TerraNova published by CTB/McGraw Hill. We found 17 states that used the TerraNova but only 11 states used it to calculate national norms. We then got the national percentile rank for all of those states. In 80 percent of them, the median student in a particular grade level and subject area was above the national average. We then used the NAEP for comparison and found of those same states, only 50 percent of the median student was above average.

There were limitations of our methodology (i.e. states tested at different grade levels, they used different forms of the TerraNova, NAEP years didn't correspond to some TerraNova years). To lessen these effects, we aggregated the data and didn't try to make too specific of a comparison.

The discrepancies didn't surprise testing

experts. In fact, they expected them. When we asked why, the answer was uniform: the Lake Wobegon Effect, named after the mythical town in Minnesota Public Radio's "A Prairie Home Companion" where "all the children are above average".

The Lake Wobegon Effect is an inflation of the national percentile rank due to discrepancies between the current test scores and an outdated, faulty or limited national average. All our states soon report their students are above average.

Most national norm-referenced standardized tests are normed every seven or eight years. The norming process involved testing a stratified sample of hundreds of thousands of students across the country to derive the national average. The norming process is rigorous, expensive and very time consuming for test publishers.

Limitations

But the process has some serious limitations.

- Age: If test scores rise over time like they have over the last decade, more and more kids suddenly appear above average.

- Teaching the test: States that use the same or similar tests over time show a rapid rise in test scores. When those tests change, the scores drop dramatically. Teachers learn what is on the test and may teach types of problems or even specific problems.

- Environment: Students who help norm a standardized test take the test under a vastly different environment than students who are taking it later as part of an operational assessment program. In a norming test, students are completely unfamiliar with the type of test. Imagine taking the SAT having never seen the types of questions on it. In an operational environment, students may have been taking the same test for years and may have a much higher comfort level. Also, in an operational environment, there is enormous pressure on students and teachers for the scores to be high whereas students in the norming process know that how well they do doesn't matter to them. To test this effect, the University of Iowa re-normed a norm-referenced test in an operational environment and found large differences in what the national average would be.

- Cheating: It's more common than people

Continued on page fourteen

Here are some Web sites
Drew Sullivan found
invaluable:

1. NAEP results: <http://nces.ed.gov/nationsreportcard/site/home.asp>

2. A great survey of state
assessment programs
from the Council of Chief
State School Officers:
[http://www.ccsso.org/
introprofile.html](http://www.ccsso.org/introprofile.html)

3. An invaluable report
with contact names and
numbers of all the
assessment people who
deal with the numbers
and Web sites of state
data: [www.ccsso.org/pdfs/
status98.pdf](http://www.ccsso.org/pdfs/status98.pdf)

4. A list of state
assessment directors:
www.ccsso.org/asap.html

The Tennessean's story:
[www.tennessean.com/siil
99/12/12/test12.shtml](http://www.tennessean.com/siil99/12/12/test12.shtml)

Dangerous devices

By Jessica Larson
IRE and NICAR

The Manufacturer and User Facility Device Experience Database (MAUDE), which contains information about medical devices that have caused problems for patients, has a number of interesting facets - particularly for reporters on the health beat.

MAUDE replaced the older MDR (Medical Device Reporting) dataset, which did not require user facilities to report their information about defective medical devices. NICAR recently updated this dataset and now has data from August 1996 through March 1999, which includes voluntary reports, user facility reports, distributor and manufacturer reports.

Some of the fields included in the database:

- The problem with the device.
- The outcome of the event (patient injury, death, etc.).
- The brand name and type of device, its age, expiration date and shelf life.
- Name and addresses for the device manufacturer and distributor.
- The location the incident occurred (patient's home, health clinic, etc.).

In 1992, U.S. News and World Report did a story on dangerous implants (breast, penile, testicular) and other medical devices of concern to the FDA. The MAUDE dataset can provide information for similar stories - although no identifying patient information is included, the frequencies of device failure and resulting injury offer a good starting point.

Which manufacturers produce the highest number of defective devices could prove to be another interesting topic. Old or expired devices, particularly when compared with the dates that authorities became aware of the problem, may be worth looking into, as well.

Because the format has changed, NICAR is providing the 1996-1998 data and the older information separately. Costs for the 1996-1998 MAUDE is \$100 for under 50,000 circulation (50-200 market), \$125 for 50-100,000 circulation (25-50 market) and \$150 for above 100,000. The older MDR data, 1984-1995, is \$75, \$100 and \$125 respectively.

To order this dataset or others available from NICAR, call (573) 884-7711.

Jessica Larson can be reached by e-mail at jessica@nicar.org

NICAR databases:

(As of Feb. 1, 2000)

- ATF Firearms dealers, 1993-1998
- FEC Contributions, 1991-1998 cycles
- Federal Contracts, 1992-1998
- Consolidated Federal Funds, 1983-1997
- National Bridge Inventory, 1994-1998
- National Inventory of Dams, 1993-1998
- OSHA, 1974-1998
- Social Security Master Death, to 1998
- Home Mortgage Disclosure data, 1992-1998
- IRS Exempt Organizations, through June 1999
- FBI Uniform Crime reports 1980, 1990, 1993-1997
- SBA 7a Business loans, 1980-1998
- SBA Disaster loans, 1980-1998
- Medical Device reports, 1974-1998
- CDC HIV Surveillance data, 1997

More information is

available at:

www.nicar.org/data or by calling (573) 884-7711.

Continued from page thirteen: Lake Wobegon

think and can skew results upward.

While the Lake Wobegon Effect is old, it continues to skew state results, sometimes very seriously. Be aware that every norm-referenced test is a victim of this effect. The result in your state may be an apparent upward trend in scores that just isn't real.

Questions to ask

Here are some questions you should ask your state assessment officers:

- Ask your state's test authorities how they calculate the national average, what test they use and when it was last normed. It's a complicated picture because states may use an off-the-shelf standardized test, part of a test, or just some test questions from a testing company. These questions are often mixed with criterion-referenced test questions.

- Ask how often your state changes test forms and how many questions are repeated every year.

- Ask about cases of cheating. Often these are called 'anomalies' or some other innocuous term. The state testing office collects the information because they must remove these cases from their statistical analyses. Often it may be a matter of a teacher giving students too much or too little time, but sometimes it is far more egregious.

- Compare your schools test results to the NAEP, ACT, SAT or other standardized tests. Also look at trends. If your students are improving on a state test and not on NAEP, you could have a problem.

- Talk to state assessment coordinators in other states. They often know what manipulations other states use to make themselves look good.

- Find out how your state deals with students who need special accommodations for the test.

Drew Sullivan can be reached by e-mail at drew@tennessean.com

NEW YORK

IRE Conference 2000

Investigative Reporters and Editors will celebrate its 25th anniversary with its national conference in New York, June 1-4.

The conference, which will be co-hosted by *The New York Times* and held at the Hilton Waldorf-Astoria, will feature an optional computer-assisted reporting day on Thursday, June 1. Although there will be no hands-on classes, there will be demonstrations and one-on-one instruction in the CAR Demo area.

The optional CAR day will include panel discussions on how to access and work with data on topics ranging from Census 2000 and campaign finance to transportation, environment and covering local and state government.

The main days of the conference will offer more than 60 panels on investigative reporting on such topics as covering bad doctors, newly discovered perils in the nuclear industry, nursing homes, day care, juvenile boot camps, wrongful convictions, corrupt city hall,

and transportation hazards.

The invited speakers include Bill Keller, managing editor of *The New York Times*, Mike Wallace of 60 Minutes, Stone Phillips of Dateline, Diane Sawyer of ABC News, and the *Time* investigative team Donald Barlett and James Steele.

In addition, the demo area will be available the entire time of the conference as will reprints of top investigative work, a TV Show and Tell Room and a variety of vendors.

Registration forms and more information on the conference are available on the IRE Web site, www.ire.org/training/ny00/

To make hotel reservations, contact the Hilton Waldorf-Astoria at (212) 355-3000 or toll-free at (800) WALDORF. Reservation cut-off is May 16. Room cost is \$215 single/double occupancy plus 13.25 percent state and local taxes and \$2 occupancy tax.

Continued from page eight:

Beyond OSCAR

fields. One relatively rare code in the inspections data meant a home's fire code violation actually wasn't a violation. Miss it, and your numbers for total citations would be off.

But learning the data isn't enough.

For the numbers to mean anything, we also had to learn federal law on inspections and enforcement, and how it evolved over time. We consumed hundreds of pages of manuals and studies to know what to ask, what data mattered and what had changed.

- Keep your data organized.

We've lost count of how many data tables we've built, but it's got to be around 1,000. But most were one-time knockoffs to test a particular question. The base tables and the key ones for analysis number about two dozen, all in one folder, and listed on paper and in Excel, with the name, location, and a description of how we built them. All the queries are saved by date and name so we can retrace any steps.

- Ask why.

Don't just settle for a list of the worst homes. See what they have in common. Are they for-profits, cutting corners on quality to save a buck? Are they non-profits that lack trained staff and modern facilities? Do they have a problem popula-

tion? Why are some problem homes still around?

And find a new question to ask.

In Chicago, Mike Berens unearthed the warehousing of the mentally ill in nursing homes. In Tampa, Doug Stanley looked at how the profit motive affects care. We looked at the big picture: What types of care are out there, who has access, and how well the state watches over those in the system.

- Use data to narrow your paper chase.

Once we had identified the problem homes, we requested the detailed, paper reports. That painted pictures on top of the numbers, unearthed more people and stories, and confirmed another tip: Homes routinely appeal inspections, getting some citations dropped from the report-often enough to drop below a penalty threshold.

But forget the data. It's about people.

A bit of an exaggeration, but it will keep you focused on your readers. Even with all our findings, with the directors of Health and Aging saying they'll make changes based on our reporting, what people will remember is Grace Blankenship.

Dave Gulliver can be reached by e-mail at dgulliver@coxohio.com

NICAR databases:
(as of Feb. 1, 2000)

- NEA grants, 1987-1997
- INS Legal Residency, 1980-1996
- FAA Enforcements, 1967-1999
- FAA Service Difficulty Reports, 1974-March 1999
- Airmen Directory and Aircraft Registry 1998
- FAA Accidents and Incidents, 1973-1999
- NASA Air Safety Reporting system, 1988-1998
- NTSB Investigation of Aircraft accidents, 1985-1998
- DOT Fatal Accidents, 1975-1998
- NHTSA Vehicle Recalls and complaints, 1966-1997
- Hazardous Materials, 1971-1998
- Railroad Accidents, 1991-1997
- Boating Accidents, 1969-1998
- Boat Registration, through July 1999

More information is available at: www.nicar.org/ data or by calling (573) 884-7711.

Bits, Bytes and Barks

Driver's license data

In January, the U.S. Supreme Court upheld the constitutionality of a 1994 law that restricts access to drivers' license data kept by the states.

For journalists, this will mean continued difficulty in obtaining the records, but this varies from state to state. Some have established an "opt-out" policy that allows the release of the records to the public, except for those individuals who voluntarily choose to have their information withheld. States that don't have such a policy can't release any of the records.

The Society of Professional Journalists posted a list on their Web site three years ago, listing each state's policy on the issue. However, this is out-of-date in some cases. The Web site is located at: www.spj.org/foia/drivers/drivside1.htm

Upcoming bootcamps

Now is the time to sign up for bootcamps that will be held this year. Sessions will be held May 7-12, July 16-21 and August 6-11.

More information, including registration forms, costs, a typical schedule and travel and lodging options, is available at the NICAR Web site, www.nicar.org. The boot camps are held at the Missouri School of Journalism in Columbia, Mo., with hands-on training from NICAR staff members.

There are minority fellowships available for bootcamps. Check out NICAR's Web site for more information. Applications can be obtained from John Green, membership coordinator, jgreen@ire.org, (573) 882-2772.

NICAR data update

Several datasets have recently been updated at the NICAR data library, including significant improvements to the documentation that is included with each dataset.

Documentation now includes references to print and broadcast stories using that data and, in some cases, tipsheets from IRE and NICAR conferences about that topic or beat. All items referenced in the documentation are available in the IRE Resource Center.

Other improvements have been made to make it easier for journalists to work with the data, including better record layouts and codesheets.

Here are the most recent datasets updated:

- Home Mortgage Disclosure Act (HMDA), 1992-1998.
- Federal Aviation Administration Enforcements, 1962 through Sept. 1999.
- Federal Aviation Administration Accidents and Incidents, 1973 through Dec. 1999.
- NTSB Investigations of Aircraft Accidents, 1985-1998.
- National Inventory of Dams, 1993-1998. There are four states that didn't submit new data: Delaware, Alabama, Iowa and Rhode Island.
- Recreational Boat Accidents, 1969-1998.

Several other dataset updates will be coming soon. Stay tuned to NICAR-L for announcements.

More information about these datasets and others – including the costs – are available online at www.nicar.org/data or by calling (573) 884-7711.

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