

Uplink

March 1994

A newsletter for the National Institute for Computer-Assisted Reporting

PROPERTY OF
JOURNALISM LIBRARY

JAN 3 1995

Uplink update

We're on our way to a new and improved Uplink, but we'll need our readers help.

We have increased the number of issues to 12 a year. We plan to jack up the number of pages. We want to pack more information and tips into each issue.

It's on the 'more information and tips' part that we'll look for your assistance.

If you do a computer-assisted reporting story that you're happy with, send it to us. Attach a note telling us how you did it.

If you see a story you didn't do but like, call us and tell us about it.

If you find a new piece of software that makes data analysis easier, let us know.

If you find a new online or offline database you think would be valuable, give us the tip.

If you figure out a good way to train journalists in a CAR technique or to bring CAR into the newsroom, tell us.

We're looking forward to hearing from you.

— Uplink's editors

Inside

Dining on eatery data, pages 2-3

Using health inspection reports

EEOC Documents, page 3

WSJ catalogs black job loss

Tech Tips, page 4

ID that record

Bits, Bytes & Barks, page 4

Shaggy dogs, race relations and census

IRE, MU join forces

NICAR on the road

So what is NICAR?

It's the National Institute for Computer-Assisted Reporting.

It's a joint venture of Investigative Reporters & Editors Inc. and the Missouri School of Journalism - started with a generous grant from the Freedom Forum.

It is the logical continuation of the Missouri Institute for Computer-Assisted Reporting, which was begun by Elliot Jaspin five years ago.

And it is the most ambitious project in the country for training journalists and educators, for providing data and data analysis, and for acting as the clearinghouse for information about computer-assisted reporting.

In the coming months, NICAR plans to:

- Increase the number and kind of seminars at Missouri.
- Offer seminars and conferences across the country.
- Build up a data library that can provide information inexpensively to news organizations.
- Help smaller papers deal with large amounts of data.
- Do data analysis on complicated projects.
- Offer services online.

The staff has a varied background, which will enable it to provide a wide range of services.

The managing director of NICAR is Brant Houston, who worked on computer-assisted stories at *The Hartford Courant* since 1988, where he was database editor. The training director is Jennifer LaFleur, who comes from the

San Jose Mercury News and knows computers, statistics and mapping software. NICAR's academic adviser is Richard Mullins, who was at the National Library of Money & Politics.

Three graduate research assistants, Matt Reavy, Drew Sullivan, and Gwen Carleton, will help in the day-to-day operations. The institute also will draw on the resources and experts at the school of journalism.

NICAR's next on-campus, weeklong seminars will be May 15-20 and Aug. 14-19.

The fees charged by the NICAR will be as low as possible, but high enough to support the expansion of its staff, equipment and other resources.

The institute welcomes donations of tapes and diskettes and advice on how to improve its services.

NICAR's phone numbers are 314-882-0684 and 314-882-9491. Its fax number is 314-882-5431. Brant Houston can be reached at Compuserve, 71023,120.

Coming Events

May 15-20

NICAR Training Seminar

Columbia, Missouri

June 16-19

IRE National Conference

St. Louis, Missouri

August 14-19

NICAR Training Seminar

Columbia, Missouri

October 6-9

IRE Computer-Assisted
Reporting Conference

San Jose, California

Uplink

March, 1994

Volume 6, Number 1

A newsletter for the
National Institute
for Computer-
Assisted Reporting

Editor

Brant Houston

Managing Editor

Matt Reavy

Senior Contributing Editor

Richard Mullins

Contributing Editors

Rosemary Arnao

Tracy Barnett

Staff

Drew Sullivan,

Mary Stampley,

Kathy Thomas,

Rich Deming and

Gwen Carleton

Uplink is published
every month by the
National Institute for
Computer-Assisted
Reporting,

13 Neff Hall,

School of Journalism,

University of Missouri,

Columbia, MO 65201.

(314) 882-0684.

Subscription price is

\$30 (U.S.) annually

Postmaster: Please
send address changes
to NICAR.

NICAR is a joint effort
of Investigative
Reporters and Editors
and the University of
Missouri School of
Journalism with the
mission of bringing the
journalism industry
into the 21st Century.

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

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

Post-Dispatch opens inspection larder

Dining on eatery data

By Dan Browning, *St. Louis Post-Dispatch*

You might win public service awards for studies of giant government databases on highway contracts, creative campaign financing or the redlining practices of mortgage lenders, but to twist an old saw, the way to your readers' hearts is through their bellies.

Everybody eats.

The health departments in many metropolitan areas have computerized their restaurant inspections, opening a larder of data to news organizations with the nerve to peek inside the kitchen—or under the dumpster.

Last spring, Jan Paul, a projects editor at the *St. Louis Post-Dispatch*, assigned three reporters to examine restaurant inspections after the public became concerned about increasing reports of E. coli bacteria poisonings in the Pacific Northwest and hepatitis outbreaks in St. Louis.

In March, the newspaper got computerized data from St. Louis County and the city of St. Louis listing 103,493 inspection reports on 8,544 restaurants, delis, snack bars, institutions, corporate cafeterias, mobile kitchens, schools, nursing homes, daycare centers and retail bakeries. (St. Louis County later updated the records, adding 8,524 inspections to the mix.)

In most places that use the FDA's model inspection form, restaurants start an inspection with 100 points (a perfect score). Inspectors look for 29 categories of minor problems, which cost the restaurant either one or two points each, and 13 categories of "critical violations" — those with the greatest potential to have an adverse effect on public health — which cost the restaurant four or five points each.

In St. Louis and St. Louis County, restaurants get letter grades to post in their windows. An "A" means the restaurant scores higher than 84 points. A "B" can go as low as 70 points, and a "C" goes as low as 50.

Restaurants that get Bs or Cs must correct their problems within two days or they can be shut down. Similarly, restaurants that get hit with critical violations must correct them within a couple of days — but the public might never know there

was a problem, because a restaurant could have up to three critical violations and still get to post the "A" rating in its window.

When we dug into the data, it seemed that the obvious way to measure relative performance on inspections was to examine average scores. We did that, and we found that establishments got As 98 percent of the time. In fact, they scored above 90 points in 88 percent of the cases. Ranking restaurants based on their overall scores seemed futile.

We figured that we should look at the frequency of critical violations on the grounds that it might only take one incident to make someone sick. And we decided we should look at the intensity of critical violations — expressed as the average number of critical violations per inspection — on the grounds that the more of them that existed, the greater the chance that someone might get sick.

Unfortunately, the databases we got were so badly designed that the health departments admitted they had never been able to use them. And the data was so "dirty" with entry errors that it took us about six weeks to clean it up so that we could group the restaurants.

Once we could use the data, we decided to limit our examination of the county data to only those inspections since Jan. 1, 1990, so that it would parallel the city data. We purged places that had closed or places that had changed ownership in the study period. And we limited our analysis to establishments that had been inspected at least six times so that the data would not be distorted by anomalous inspections.

We ranked all establishments based on their average inspection scores and the frequency and intensity of their critical violations. We looked at retail eateries as a class (restaurants, delis, snack bars, salad bars), schools and

daycare centers, retail bakeries, any type of operation run by grocery chains, nursing homes and hospitals. We examined scores based on ethnicity. And we examined the scores of all the individual health inspectors to see whether any

**This analysis, based on
44,000 inspection at 3,200
eateries, found
'A' rated restaurants
could still pose a risk.**

Beginning and advanced skills highlighted Computer conference set for Oct.6-9, 1994

By Jonathan Krim, *San Jose Mercury News*

The second annual national IRE conference devoted to computer-assisted journalism will be held in the heart of Silicon Valley this October.

Whether your publication or broadcast outlet is just beginning to use computer-assisted techniques or is ready for more sophisticated approaches, this year's conference will have something for you.

Some of the top practitioners nationwide will serve as instructors, offering classroom training in which each participant will work at his or her own computer.

Conference participants will receive intensive, hands-on training in a variety of computer applications, including:

- Basic PC use, including DOS and Windows
- How to use various spreadsheet programs
- How to use relational database software to further manipulate information
- How to download government tapes
- How to use on-line services, including Internet
- How to map data
- How to use statistical analysis software

The conference, beginning Oct. 6 at 8 a.m. and ending at noon on Oct. 9, also will feature panel discussions on issues affecting computer-assisted journalism, such as legal matters, ethics, and using computers for beat reporting and quick-hit stories.

Building on the strengths of the computer-assisted conference in Raleigh, the planners intend to expand advanced parts of the program while keeping a basic track for those who are just learning these techniques.

"This conference will be a great opportunity to learn to use the most basic and necessary tools for reporting," said Brant Houston, IRE's new computer-journalism director.

"Reporters and editors can no longer afford to ignore this area. After all, most information we will seek is or will be in electronic form."

Hosted by the *San Jose Mercury News*, the conference will also feature displays of products for computer-assisted journalism and a special demonstration of Mercury Center, the *Mercury News*' on-line newspaper service. A luncheon banquet, included with registration, will feature an exciting keynote speaker.

The conference site, the Westin Hotel in Santa Clara, features many amenities, including free parking, courtesy transportation from the San Jose airport and light-rail access to and from downtown San Jose.

For those who want to bring their families, the hotel is one block from the Great America theme and amusement park and is close to activities in San Jose, San Francisco and the rest of the Bay area.

Special conference hotel rates are available, but registration will be limited. So, register now.

San Jose Registration (First-come, first served. Limited registration)

Name _____

Affiliation _____

Address _____

☐ work ☐ home

City, State, Zip _____

Phone _____

Please check:

_____ \$150 for entire conference (I am an IRE member.)

_____ \$180 for entire conference (I am not an IRE member, but I understand the fee includes the \$30 membership cost.)

_____ \$165 student fee for entire conference (I am not an IRE member, but I understand the fee includes the \$15 membership cost.)

_____ \$180 for entire conference (I am an IRE member but my membership will expire before the conference. Therefore, I would like to renew my membership at this time.)

_____ TOTAL

As much basic hands-on training as possible will be provided, but we cannot guarantee it to everyone. IRE limits overall registration to the conference.

Do you want hands-on training?

___ Yes ___ No

If you checked yes, now check the areas in which you would be most interested.

- ___ Windows/DOS
- ___ Spreadsheets
- ___ NineTrack Express I: Downloading a flat file
- ___ NineTrack Express II: Downloading redefined databases
- ___ Creating a database, counting, indexing, cutting, copying
- ___ Beginning data analysis
- ___ Advanced data analysis
- ___ Using the Macintosh for computer-assisted reporting
- ___ Internet
- ___ Other on-line databases
- ___ Basic statistics
- ___ Introduction to Paradox
- ___ Introduction to FoxPro
- ___ Introduction to HyperText

Act quickly because registration in the hands-on classes is limited.

Hotel Registration Westin Hotel, Santa Clara

■ When you call, say you are an IRE participant. Call (800) 228-3000 or (408) 986-0700 to make your hotel reservation.

■ Reservations made after Sept. 9 will be on a space-available basis, and rooms may cost more.

■ Hotel rate is \$89, single; \$99 double and \$109 triple.

patterns emerged that might indicate hanky-panky.

Because of the number of ties in critical violations, the range of possible places in our ranking system was dramatically reduced. The critical violations lost their emphasis when we merged them with average inspection scores, which had a greater range of possible rankings. Thom Gross, an investigative reporter, suggested converting all three rankings to percentiles, then averaging them together. We did that and it worked. Several public health officials around the country said our methodology seemed sound.

Two reporters took a list of eateries that scored in the bottom one percent in our lists, and began verifying every single hard copy of their inspections. We found only two errors in several hundred records, and they did not affect the totals. We corrected those, and decided the overall data was statistically reliable.

We came under attack from several quarters. The director of the city's health department said our study was no more valid than ranking restaurants based on their addresses, because no link has been established between foodborne disease and restaurant inspections. But thanks to a search I did on CompuServe before we met with him, we were prepared. We directed his attention to an FDA study in Seattle-King County, which concluded that restaurants with low scores and critical violations were directly correlated to outbreaks of foodborne disease.

The final challenge to our project came from the Missouri Restaurant Association, which argued that the inspection system itself was flawed. We agreed, but for opposite reasons. The industry association argued that any restaurant that was open was safe. Our articles showed that the inspection system failed to communicate anything to the consumer about critical violations, that inspectors often give a break to restaurant owners if they are cooperative and that the letter grade system used in St. Louis can be misleading.

When we finally published the articles, our analysis was based on about 44,000 inspections covering 3,200 eateries.

Our restaurant inspection report was a hit with our readers. Callers are praising us for it. Although it can be time-consuming, it's a project that can be done virtually any place that has a health department. But beware: The project was not a hit with the advertisers. We were told that after the project, an editorial and a cartoon ran, advertisers cancelled a special section on dining out.

Analysis covers 35,242 companies, 40 million workers

EEOC data is goldmine for job stories

By Thomas Monnay

The 1990-91 recession left the impression that many jobs simply disappeared, but a recent computer analysis by the *Wall Street Journal* shows that among ethnic groups only blacks suffered a net job loss.

Journal staffers Ed Souldessy and Rochelle Sharp examined Equal Employment Opportunity Commission (EEOC) reports from 35,242 companies covering more than 40 million workers.

The law requires that businesses with 100 or more employees annually report the number of women and minorities they hire. Businesses with 50 or more employees must also file reports if they have a government contract of \$50,000 or more.

The *Journal's* computer analysis of these records showed that blacks lost a net 59,479 jobs during the study period. Asians and Hispanics gained 55,104 and 60,040 jobs respectively, while whites gained 71,444 jobs.

When the news hit Washington, angered civil rights leaders and members of the Black Caucus pressed for an immediate investigation into the companies involved.

The article itself capped hundreds of hours spent locating the tapes, analyzing the data and verifying reports.

The *Journal* originally obtained 9-track computer tapes containing the EEOC data, but found that the fields containing company names had been deleted. Department of Labor officials claimed the data could not be released without permission from the companies involved.

When contacted by the Department, most companies objected. The *Journal* appealed the decision, but still no companies responded.

After filing three separate Freedom of Information Act (FOIA) requests, the *Journal* succeeded in obtaining the names, addresses and summary employment filings of the nation's 400 largest federal contractors in terms of market capitalization. The appeal process on these filings alone lasted six months.

"I certainly didn't expect that," said Sharpe. "But it was worth sticking with."

After Souldessy and Sharpe analyzed the data, a team of reporters began the time-consuming process of calling all 400 companies to verify the data.

During verification Sear, Roebuck & Co. indicated the firm had provided incorrect data to the government. Although the company said the mistake exaggerated blacks' proportion of job loss, the revised data still showed blacks losing jobs at a disproportionate rate.

As a result of the story and the resulting response among minority activists, Congressman Julien Dixie of California has vowed to launch an investigation into the matter.

Tech Tips

How to keep data straight by adding an ID

By Richard Mullins, University of Missouri

Every data table you work with should have rows that are unique. Duplicates can be confusing, for one thing, because you never know if the duplicate row belongs there. (Did Richard Mullins bounce a check twice, or did a data clerk just type it in twice?)

This uniqueness rule is not just a good idea, it's the law, at least according to relational theory. Most often this is a number, in a column called ID, (or Casenum, Case# or CaseNo).

But we all know that there are lots of databases and data creators that overlook this. You may have done it too, in your young and foolish days.

Here's how to right the wrong and be sure every row in a table is uniquely identified. The commands are for FoxPro or dBase.

After you've created a new field to hold the number (make sure it's NUMERIC and big enough to hold the largest number that might be put there), type this command

in the Command Window: (The field or column name here is ID)

REPLACE ALL id WITH RECNO()

That's it. This one-line command works because Fox stores internally a number for every row, or record, in the table. The REPLACE... WITH command puts that number (supplied by RECNO(), a built-in function) in a field where you can see it. The command word ALL makes the replacement on every row.

Here's one more twist on this command, if you want to start the ID numbers with a larger number. You can add another number, say 1000, to the internal record number. Then your ID numbers would start at 1000, instead of one. Here's the command:

REPLACE ALL id WITH RECNO() + 1000

Bits, Bytes & Barks

Akron completes yearlong CAR series

The Akron Beacon Journal recently wound up its year-long analysis of blacks and whites in the 90s with a four-part series about crime and punishment.

Other series during the course of the investigation dealt with race relations, living conditions, health and education, and economic progress.

Bob Paynter, assistant metro editor-projects, indicated that the staff used computers to examine census data, efficiency test scores, police reports, and a variety of county and school district databases.

Reprints of the story are available from the *Beacon Journal*.

— Matt Reavy

A shaggy dog story?

Imitation is the best form of flattery. *The Eagle-Tribune*, Lawrence, Mass., wanted a quick computer assisted story that would give a reporter a chance to learn databases as well as grab the reader. So we borrowed a story that was done by Steve Doig at the *Miami Herald*.

We requested dog registrations for 20 communities in southern New Hampshire and northern Massachusetts. Some towns used FoxPro and provided dog registrations, including name, breed, owners, address, color and sex, on floppy diskettes. Others gave us the data on nine-track

tapes. Some towns required the reporter to input data. In the end, we had 18,000 dogs. We crunched them down by most common breed per town, per state, and most common name per breed. The winner: Max, the mutt.

The payoff. A Plaistow, N.H. woman who had 33 Chihuahuas — each with its own bed, living in her home. They all ate at 3 p.m., and she never let them out of her home. Oh yes, the most common Chihuahua's name? Chico, of course.

— Brad Goldstein

Census Bureau phasing out mainframes

The U.S. Census Bureau has announced plans to phase out its mainframe system by 1996, just in time for the 2000 decennial census.

The bureau plans to utilize its 18 Digital Equipment Corporation VAX minicomputers to pick up the bulk of the load. DEC terminals or PCs emulating DEC terminals are used to connect to the VAX.

Plans also call for construction of a new computer center at the University of Maryland's Technology Center in Bowie, Md. by 1996.

The bureau is reportedly examining what will become of its archive, which currently reside on 9-track computer tape, as well as the possibility of getting into digital imaging and moving onto Internet.

— Matt Reavy