Uplink

October 1994

A newsletter for the National Institute for Computer-Assisted Reporting

Uplink update

This month's Uplink offers a potpourri of information that reflects the wide range of uses for computer-assisted reporting in journalism.

In addition to the cover story on aviation data, Gwen Carleton's piece explores the use of CAR in science journalism; Paul D'Ambrosio of the *Asbury Park Press* writes about exposing voter fraud; and David Bloom of the *Los Angeles Daily News* gives advice (in part one of two) about how your mastery of computer lingo allows you to cover government mishandling of computers and computer contracts.

Next month we plan to report on the leading experts at CAR TREK, this year's national conference on computer-assisted reporting.

In the meantime, don't forget to send us your clips and tips.

Inside

FAA database provides quick history

Examining the USAir crash

By Brant Houston
NICAR

When US Air Flight 427 crashed Sept. 8, the new data library at the National Institute for Computer-Assisted Reporting got its first emergency test.

From the results, it appears the institute passed.

More than a dozen news organizations, ranging from small to large, requested and received data from NICAR on mechanical problems that the Boeing 737 had experienced prior to the crash.

While the data did not point to a cause, it provided context for print and broadcast journalists throughout the country.

For several years, the Institute has kept Federal Aviation Administration records on service difficulty reports - reports filed when an airplane experiences mechanical problems or when inspectors find problems.

During the last six months the Institute has reorganized the data, which goes back to 1988, so that it is more easily accessible and can be distributed on different kinds of media. The Institute not only gets monthly updates of the data, but it also is working on a project to include data covering the 15 years prior to 1988.

While some news organizations have purchased the data themselves, and others have accessed the data through commercial vendors online, many journalists still seek the data from NICAR. This includes those who can't afford or do not have access to commercial databases, want the data in a form in which they can do wider searchers or want to work with NICAR to understand the data better.

After the crash, NICAR staffer Drew Sullivan, with the help of graduate student Padraic Cassidy, fielded calls from such organizations as WDIV-TV in Detroit, the *Beaver County Times* in Pennsylvania and the *Chicago Sun-Times*.

Other organizations - *The Boston Globe* and WCCO-TV in Minneapolis - had already purchased a CD-ROM from NICAR that contains the FAA database and did their searches on those CD-ROMs.

The service difficulty reports are not without flaws. After all, every database has imperfections and is often only a starting point for a story. In fact, Cox Newspapers began a database analysis that ended up as a story that revealed shortcomings with the data and the inspection system.

Larry Lipman, a reporter in the Cox Newspapers' Washington, D.C., bureau, worked with Elliot Jaspin (former director of the Institute and now at

Continued on page 5

See related story on page 2.

Coming Events

October 6-9, 1994
CAR Trek Conference
Silicon Valley, California

January 8-13, 1995
NICAR Seminar
Columbia, Missouri

Uplink
October 1994
Volume 6, Number 8
A newsletter for the National Institute for Computer-Assisted Reporting

Editor
Brant Houston
jourjbh@muccmail.missouri.edu
Managing Editor
Matt Reavy
c598895@mizzou1.missouri.edu
Senior Contributing Editors
Richard Mullins
jourram@muccmail.missouri.edu
Jennifer LaFleur
jourjal@muccmail.missouri.edu
Contributing Editors
Rosemary Armao
Tracy Barnett
Staff
Drew Sullivan, Wallace Winfrey, Padraic Cassidy, John Anglim, Rich Deming and Gwen Carleton

Uplink is published every month by the National Institute for Computer-Assisted Reporting,
P.O. Box 838
Columbia, MO 65205.
(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.

When the numbers look too good to be true

FAA reports considered suspect

By Larry Lipman
The Palm Beach Post

Sometimes the numbers look just too good to be true. That's when it pays to be suspicious.

Two days after the mysterious crash of USAir Flight 427 near Pittsburgh, Elliot Jaspin and I were looking for a common thread in the Federal Aviation Administration's service difficulty reports on 737s.

Elliot had obtained a five-year file of SDRs by e-mail from the National Institute for Computer-Assisted Reporting. Unzipped, it took up more than 30 megabytes.

Using FoxPro, he sorted the records by level of severity. The FAA ranks its SDRs on a scale of 1 to 5, with 1 having a "seldom relationship" to an accident and 5 having a "frequent relationship" to an accident. We decided to just look at the level 5 reports.

He then sorted the level 5s by the type of problem, looking for some pattern of difficulty with the 737s. By far, the biggest area of difficulty was "Other." We looked at all the other areas, but no significant pattern emerged. We looked closer at "Other," and, again, no pattern emerged.

Then Elliot sorted the data by airline and by unique plane numbers.

Suddenly, it appeared that we had a pattern — and a helluva story. Dividing the number of SDR 5s for each airline into the number of 737s in its fleet over the past five years gave us a percentage.

From this calculation, it appeared that two-thirds of USAir's planes had reported a level 5 SDR. That was among the highest percentage of the major airlines. It looked like we'd made a significant finding. I called USAir for their reaction to our hypothesis that two-thirds of their planes had experienced major service difficulties. They said they weren't sure what the data showed and defended their record as a safe airline.

We decided to run the data past aviation safety experts. I posted an e-mail query on Profnet. Within a few hours I had the numbers of several aviation safety experts affiliated with universities, and they led me to others in the field.

After a day of talking to experts, I was convinced our hypothesis was wrong. The numbers were accurate; they were also misleading. Every expert I spoke with said that publishing a story along the lines I suggested would be unfair. They gave two main reasons:

— Not all airlines report SDRs consistently. At least one expert noted that USAir, which has a relatively young fleet of 737s, was reporting SDRs at twice the rate of two airlines with substantially older 737s. That might indicate that USAir was simply more scrupulous about reporting its SDRs than the other airlines.

— While the number of SDRs compared with the number of planes gave us a percentage, it was meaningless. When you looked at the thousands of times each airline's planes had flown over the past five years, the percentage of SDRs to flights was minuscule. To say that two-thirds of the planes had problems, when far less than 1 percent of the flights had problems, would have been misleading and sensationalist.

Faced with such comments by aviation experts not affiliated with any airline, we decided to scrap the story. But we still had something; we had information that a data system used by the FAA wasn't very reliable or meaningful.

We decided to find out what the FAA did with this data and here, ultimately, is where we found our story.

It turns out that the FAA admits the data from the airlines is neither consistent nor used as well as it might be for statistical analysis of either the airline's performance or safety trends. Virtually all the analysis is done manually because the FAA doesn't have the staff or the computer resources to better handle the data.

That's the story we ran.

— Larry Lipman is the Washington correspondent for *The Palm Beach Post* in the Cox Newspapers Washington Bureau. Elliot Jaspin is Systems Editor for Cox Newspapers Washington Bureau. They can be reached by e-mail at jaspin@access.digex.net.

It turns out that the FAA admits the data from the airlines is neither consistent nor used as well as it might be for statistical analysis of either the airline's performance or safety trends.

Asbury Park Press exposes voting fraud

Voters cast ballots from grave

By Paul D'Ambrosio
Asbury Park (NJ Press

Egads! The dead are still voting in New Jersey.

Since the invention of the ballot box, politicos around the country have always found clever ways of stuffing it with illegal votes. Rumors abound that Kennedy won the 1960 presidential election with votes from the silent majority resting in the graveyards of Hudson County, N.J., and Chicago.

Now, with proposed reforms that will make registering to vote as easy as signing a form, the *Press* wanted to see if the dearly departed were still actively involved in earthly politics. But we encountered one major obstacle: Death records kept by New Jersey, as in some other states, are private.

We had the voter registration tapes. But without death records, no comparison could be made.

The Social Security Administration came to our rescue.

The SSA has developed a "Death Master File" to record everyone with a Social Security number who has died since 1937. That translates into 40 million dead, enough records to fill 19 6,250 bpi 9-track tapes.

While the list is not perfect, it is the best there is if you can't get the full records from your state's vital statistics department. In fact, the U.S. General Accounting Office this summer praised the DMF as the most accurate collection of death records in the United States. Other departments, such as defense, routinely use the DMF to remove the dead from pension records.

I pulled out New Jersey's death records from the DMF using Nine-Track Express and compared the names and dates of birth with our voter registration tapes.

It took a 486/66 an hour to process the information with FoxPro, but it delivered gold: We found more than 400 deceased still on the voting rolls. More importantly, we found some — three in one county alone — who were continuing to vote from the grave. One man's political party even changed from Democrat to Republican after his death.

I examined the findings in conjunction with information on the National Voter Registration Act that will go into effect Jan. 1.

The act, better known as the Motor Voter law, makes registration easier, but mandates that states keep their voting rolls as clean as possible. The law is forcing states to develop statewide, computerized registration lists so they can purge those who have died or moved.

New Jersey is spending $4 million to develop such a database. For now, each county controls its own voter rolls. Checking for deaths is as low-tech as reading the obit page. However, after I told one county how we found their dead voters, officials there said they would begin using the DMF.

While the DMF only has eight fields (SSN, last name, first name, date of death, date of birth, state, zip of last residence and zip of lump sum payment), there were some tricks to massaging the information.

The DOB uses century, a format not known to Nine-Track Express. It looks like 12/12/960. To get around this, I told Nine-Track to set up four separate fields: month, day, century and year.

This meant I had to split the voter's date of birth into two new fields: month and year.

After pulling out names of the deceased by their zip codes, I joined the two files using last name, first name, month and year of birth. I excluded the day of birth since the SSA sometimes used "00" to denote an unknown day. Out of 1.3 million deaths and 450,000 voters, I only got a handful of duplicate matches on common names.

That was the easy part. In one county, about 10 percent of the voters listed as dead by DMF were still alive. Apparently the local SSA office had a habit of killing off living spouses the same day their husband or wife died. This became a sidebar.

To be absolutely certain a voter was dead, I obtained hard copy death certificates from the state, checked voting signature books and called the next of kin before confronting election officials.

Officials tried to claim that votes by the dead were clerical errors until they saw the signature sheets showing that someone had signed in after the voter's death. Investigations are now pending. — Paul D'Ambrosio can be reached at pmd@app.com or by phone at (908)922-6000, ext. 4261.

The SSA has developed a "Death Master File" to record everyone with a Social Security number who has died since 1937. That translates into 40 million dead, enough records to fill 19 6,250 bpi 9-track tapes.

The master set of DMF costs about $1,073 from the National Technical Information Service (703/487-4630). NICAR is in the process of purchasing the data and will offer it to newspapers for approximately $200.

Tapping into the power of databases

Science stories and CAR

By Gwen Carleton
NICAR Staff

Everyone's heard about the computer-assisted story that rocked city hall. But what if you prefer physics to politics? As science writers around the country are discovering, the right computer data can bring perspective, insight and scope to stories large or small.

Toxic chemical stories

Reporters from the *Bee* newspapers in Fresno, Modesto and Sacramento recently used computers to analyze the amount and location of toxic chemicals in California's Central Valley. Their six-month investigation resulted in the series "Dirty Air," which appeared in the *Fresno Bee* in early November, 1993. The series illustrated how tons of agricultural pesticides are released in central California every year, adding to the Valley's already severe air pollution.

Reporters Chris Bowman, Russell Clemings and Alvie Lindsay discovered that more than 130 million pounds of toxic pesticides and industrial chemicals were released in the area in 1990. By analyzing data on 9-track tapes from the California EPA's Pesticide Use Report and the federal EPA's Toxic Release Inventory, they also determined where the largest releases of the most toxic chemicals occurred.

Clemings came up with the idea in the mid '80s after he discovered the California pesticide database. "I thought it would be nice to take that data and sum up the amount of pesticide out there. But it was too big a task for a PC at that time ... it was a database of mind-boggling size."

The reliability of the California records, which included only the most toxic pesticides, also was a barrier. But in the late '80s state legislators increased the number of regulated chemicals, making the records more comprehensive.

By 1993, a more advanced computer system and editorial support for an investigative project also were in place. Armed with a new IBM 486, the California team was ready to go.

Instead of summarizing pesticide amounts, the reporters decided to rate each chemical by its LD-50 value, a standard toxicity measure.

"They shovel pesticides into mice until half are dead, and then call that the LD-50," Clemings said. "It wasn't perfect, but toxicologists said that something toxic by an oral route was probably toxic by an inhalation route."

With the help of Foxpro software, they discovered the 10 most common toxic chemicals in the Valley's air, the 10 primary sources of those chemicals and the 10 areas that were most affected. Bowman, Clemings and Lindsay also created a full-page color map, enabling readers to identify the major toxins in their communities.

The study found the farming industry was responsible for many of the toxic hot spots, the majority of which were on the east side of the Central Valley, adjacent to growing cities. Although state legislative reaction to the series was lukewarm, Clemings said he is always ready for similar computer-assisted stories.

"Knowing what I know now, this would take about one month or six weeks — the computer

Some databases produced by the federal government:

Environmental Protection Agency

TRI (Toxic Chemical Release Inventory): Contains information on the annual estimated releases of toxic chemicals to the environment.

Hazardous Waste Data Base: Maintains information by the 5,000 facilities that treat, store, or dispose of hazardous waste and 165,000 handlers who generate or transport hazardous waste.

Asbestos Information System: Assists the EPA in collecting, storing and analyzing data on the commercial and industrial uses of asbestos.

Grants Information and Control System: Tracks the processing of all EPA grant applications and active grant projects. Both Washington, D.C., and regional programs are included.

Department of Health and Human Services

Cancer Surveillance and Epidemiology in the U.S. and Puerto Rico, 1973-77: Contains information and demographics on cancer gathered from hospitals, clinics, private labs, autopsies, death certificates, etc.

Cholesterol and Smoking Data Base: Maintains information on blood cholesterol and smoking geared to the public, health professionals and issues pertaining to the workplace.

Biologics Defect Reporting System: Contains information transcribed from reports of defects, errors and fatalities ascribed to the use of biological products.

Priority-Based Assessment of Food Additives: Contains summaries of food additive safety profiles, ranking of additives based on safety concerns.

— Selected from The Federal Database Finder, Third Edition

analysis part," Clemings said. "There's lots of data on the environmental beat. Those agencies generate data better than they clean up the environment."

Federal and local databases such as the EPA's Toxic Release Inventory can be a gold mine for journalists, often providing information for numerous stories. But federal and state computer records are just the beginning.

University medical school stories

Universities, corporations and non-profits all keep important data in electronic form — the type of thing that can make or break a story. That is, if you can get it.

When Maura Lerner and Joe Rigert of the *Minneapolis Star Tribune* decided to investigate their local university's medical school, they found out just how difficult it can be to take on a local sacred cow.

"In the beginning, they gave us quite a bit of stuff, but after the first article they closed up," Rigert said. "We had to fight for everything we got."

One of the few items they acquired without legal pressure was a pile of computer disks containing information on thousands of university research projects.

"We got disks with all the research grants for five years," Rigert said. "They included dates, money involved, titles of researchers, principal and secondary investigators."

"Using that, we were able to get all sorts of breakdowns on what companies certain professers were getting money from, what companies were doing lots of research at the university," he said. "Every conceivable variation, it was all there."

The disks, which Lerner and Rigert analyzed using Paradox database management software, added significantly to the two-year investigation. "Money vs. Mission at the University of Minnesota," which ran between May 1992 and December 1993, revealed 20 years of wrongdoing in a drug program, misuse of medical-practice funds, get-rich research schemes, kickbacks and research misconduct.

The *Star Tribune* investigation resulted in the resignations of four top university administrators, an FBI investigation, an overhaul of the school's secretive private practice system and one conviction for fraud.

But all computer-assisted science stories need not produce dramatic results. Huge amounts of data that once existed only on paper — or didn't exist at all — now are available in electronic form. For journalists interested in the big picture, the results can be significant.

There's lots of data on the environmental beat. Those agencies generate data better than they clean up the environment.

Continued from page one

Examining the USAir crash

Cox) on the stories. He said Cox used the FAA data effectively for a story on service difficulty reports. Then, they started an overall story using the data for safety comparisons within the industry.

When they began doing interviews on their findings, however, air safety experts warned them off their conclusion that 737s had a high rate of problems.

"Everybody told us not to use the numbers," said Lipman, who writes for the Palm Beach Post.

Nonetheless, Lipman said the data was useful in the initial story and led Cox to their critical review of the inspection and report system.

"Service Difficulty Reporting is uneven," NICAR's Sullivan said. "Airlines do most of the reporting themselves and government oversight has been diluted. But overall, for a particular airplane, it is a valuable insight into the plane's repair history."

Using the database itself can be tricky. A first search usually is based on the tail number of the plane. But the tail number can change. So, during the first search, you must note the serial number of the plane and do a second search.

Sullivan compares the tail number to a car's license plate and the plane's serial number to a car's vehicle identification number. In the case of flight 427, NICAR's on-the-road trainer Jennifer LaFleur (who happened to be doing a seminar in Harrisburg, Pa.) alerted Sullivan to the crash and later supplied the tail number.

Sullivan turned up 26 service difficulty reports based on the tail number and a 27th report based on the serial number. Each report has a severity rating, ranging from 1 (low) to 5 (high). Of the 27 reports, two were in the serious categories: a faulty fuel valve and a blown tire.

The FAA database has about 30 fields and 10 pages of codes. NICAR charges about $20 for a search and can send data for a particular plane by electronic mail or other online services such as the Internet, by fax, or by diskette. The CD-Rom for 1988 to present is $125 and NICAR hopes to have a file transfer site running soon for updates.

NICAR also has a database on pilot licenses and a database on aircraft registration.

The medium is the message

Covering the computer beat

By David Bloom

The Los Angeles Daily News

The *IRE Journal* and *Uplink* have run a lot of stories in recent years about using computers to improve your journalism, analyzing reams of data, finding trends, and pinpointing problems with computer-assisted techniques.

But we don't often think to write about computers themselves, about the ways they are bought, improved, maintained, developed, used and misused by government agencies, business and others.

We don't often think to write about computers themselves, about the ways they are bought, improved, maintained, developed, used and misused by government agencies, business and others.

It's a fertile field for good investigative pieces, for one simple reason: A lot of times the top decision makers who must pick computer systems aren't particularly good at it, often because they aren't any more comfortable with the arcana of computer systems than any other non-specialist.

Just about any business, including (and sometimes especially) government, has top managers who don't know much about computers but now are making important decisions that may cost millions of dollars and cause uncounted problems if they are wrong.

Luckily, if you've been trying to learn more about computer-assisted reporting, you've probably learned enough of the big-picture details about computing to at least understand what the experts are talking about on these stories. But even if you don't have that background, here are some ideas to get you going.

The Basics

Many of these are basic story approaches that any veteran government reporter will recognize, only with a high-tech spin that often means they've been neglected both within the government agency and without.

Look for the usual sources of dispassionate information, the kind of stuff *The Reporter's Handbook* by IRE talks about, such as audits, contracts, trade press reports, and professors of computer science or management information science and other specialists. Usually there are reports trying to explain what went wrong to governing bodies and investigative agencies such as grand juries. (This depends upon the state. California grand juries issue an annual watchdog report on government operations that often has useful information on these things).

There are also the somewhat less dispassionate sources, such as insiders and former insiders at the agency's data processing operations and competitors' lobbyists and marketing representatives (who are always willing to dish dirt on what's going bad).

Looking to buy a big box

How are contracts structured and awarded? More than with most contracts for tangible goods, a craftily written computer hardware, software or services contract can be set up to nearly guarantee one company will win a lucrative bid.

Why? Well, take one common situation: the county welfare agency needs a new software program to automate the form-bedeviled eligibility determination process.

But say the welfare agency already is using an aging IBM mainframe. Because the agency already has IBM, the contract could specify only a program that runs with that company's mainframe and the operating system it is using.

While this seems rational enough, does it mean the welfare agency is locking itself into an old technology and approach that ultimately will be more expensive and less effective?

Specifying one computer and one kind of operating system for it generally means a lot of other companies that could provide workable alternatives for less cost are being shut out of the process from the beginning.

The state of California, currently trying to develop a new welfare computer system, has trashed one of two approaches after spending millions on it. It is choosing to go with the other system, though it uses quite dated technology and is already behind schedule.

Contract Wars

Look at the lobbying wars for computer contracts. What are companies doing to influence the writing of these requests?

In poking around Los Angeles County government, I noticed in lobbying reports that the computer companies were spending hundreds of thousands of dollars on all kinds of county officials. Those reports detailed how IBM, EDS, Unisys, Lockheed Information Management Systems and other big data companies were jockeying for millions of dollars in contracts of all kinds.

The companies didn't even focus their efforts on department heads and elected officials, though those top decision makers weren't ignored. Instead, the companies were targeting the mid-level bureaucrats who typically have to work with such systems and often are charged with the grunt work of acquiring new systems.

Not everyone has such lobbying reports to tap for information. But maybe the agency requires officials to file campaign contribution reports, or annual conflict-of-interest reports, or gift reports that a reporter can look at.

One of the things I do is keep copies of all the county's contracts when they're approved. That's a great place to look when you're trying to figure out what is at stake for these companies. Follow these contracts over a few years and you'll find some interesting patterns of renewals and expanded payments.

Look outward, angels

Many governments are now looking at "outsourcing" their data processing operations, contracting virtually everything with one big provider, like EDS or Lockheed. Outsourcing is a hot trend in both the private and public sectors, but one laden with land mines. If your government agency is considering doing it, watch it closely, and watch who gets the contract.

These outsourcing contracts are notorious for their problems, particularly when the agency has unanticipated new computer demands. At this point, they often are vulnerable to exorbitant charges from their new "partner." Having jettisoned their in-house specialists, the agency often has little choice but to go along, at a premium price.

Outsourcing also raises issues of data integrity and security that just aren't present in the private sector.

For instance, while it's darned problematic for some private company if its financial system gets scrambled, it's not a community-wide data disaster like having birth certificate information, criminal or judicial information or property records fouled up, tampered with or destroyed. How secure is this information?

The Long Run

Look at the maintenance contracts for these systems. Often they are more lucrative than the original purchase, and get less oversight. These contracts can run into the millions of dollars, yet are often routinely, even automatically awarded to the same vendor year after year without bids or contract reviews.

Look at what decisions are being made about technology. Mainframes are no longer the sole answer and often are the worst one. Distributed systems, such as networks of PCs or work stations, sometimes connected to mini computers, often are cheaper, more flexible and can have new programs developed in substantially less time. Again, are government agencies making the best decisions, or even considering all the options?

The Diaspora

As systems have decentralized, so has the oversight. This trend drives central data processing operations bonkers. It also has implications for the agency in terms of controlling its spending, making sure systems are compatible and making sure they all can talk together.

Los Angeles County, the nation's largest, spends hundreds of millions of dollars a year on computer-related expenses, but it can't tell you how many hundreds of millions of dollars a year. That's virtually a story in itself.

A rule of thumb on computer-related costs is 3 percent of sales in the private sector, 3 percent of expenditures in the public. That indicates a lot of money, but if they aren't spending in that ballpark, are they too slow in automating?

The county is now doing a study to find out what it's spending on computers; so far they know they have more than 700 different computer systems. Ask your agency what they spend. They probably can't tell you.

By the way, if they *have* done such a study, get it. It's a good resource for computer-assisted reporting projects. You'll have some idea of the databases the agency tracks, and who's in charge of them.

As systems have decentralized, so has the oversight. This trend drives central data processing operations bonkers. It also has implications for the agency in terms of controlling its spending, making sure systems are compatible and making sure they all can talk together.

Check the priesthood

You should also look at what's happened to the agency's information services or data processing department. Frequently, it gets left behind, as other departments decide they can do without mainframes and the high internal charges the DP guys bill them for their services. But they are the organization's internal experts. Are they effective and underutilized, or are they out-of-touch dinosaurs wedded to old technology?

It's all timing. And money.

When it comes to developing a new computerized system, you can almost guarantee any large project will come in late and over budget. An expert I interviewed routinely doubled all estimates of delivery time and price to get a realistic view of a project's development schedule.

Part two will appear next month.

Tech Tips

Calculating medians in Paradox

By Ray Robinson
The Press of Atlantic City

Any database software can easily calculate the average of a numeric field. But sometimes a few extremely high values skew the average.

We ran into that with the income field of the Home Mortgage Disclosure Act data for New Jersey. Our MSA, due to some extremely high incomes, was showing the average loan applicant with an income of more than $84,000 per year.

When the average is skewed, you can often get a more representative number by calculating the median. That's a little more difficult than calculating an average. Here's how it's done in Paradox 4.0:

1) Query your table, marking the following fields: the field you want to group by; the field you want to calculate as a median (be sure to mark this field with Alt-F6, rather than just F6, so it will include duplicate values in the answer). Rename the answer table to save it.

2) Sort the table in ascending order, making the median field the last sort field.

3) Restructure the table, adding a numeric field called RECNO.

4) Insert the record number Paradox has assigned to each record into the new field. Here's the code:

```
EDIT "Tablename"
MOVETO [RECNO]
SCAN
      [] = RECNO()
ENDSCAN
DO_IT!
```

5) Query the resulting table to mark the grouping field and do a CALC AVERAGE on the RECNO field. Rename the answer table to save it.

6) The resulting table will have each value in the grouping field with an average in the RECNO field. Now you need to strip the decimal places off the averages in RECNO. Here's the code:

```
EDIT "Tablename"
SCAN
 IF ([Average of Recno] - Int([Average of Recno])) > 0
  THEN [Average of Recno] = Int([Average of Recno])
   COPYTOARRAY arecord
   DOWN
   INS
   COPYFROMARRAY arecord
   [Average of Recno] = [Average of Recno] + 1
 ENDIF
ENDSCAN
DO_IT!
```

7) Now set up a multi-line query as follows: In the new table, put an example element in the Average of Recno field. In the earlier table, put the same example element in the Recno field. Mark the grouping field in that table.

And in the field you want to calculate as a median, type CALC AVERAGE AS MEDIAN.

The resulting table should display each value in the grouping field with a median value.

Yes, it's a headache. And if you only have one median to figure, it's obviously easier to sort the original table and scroll down to the record number that falls in the middle. But if you want to calculate the median for a number of different groups, it's worth it. —Ray Robinson, (609) 272-7273; Primary: 73203.3644@compuserve.com; Secondary: acpress@acy.digex.net

Bits, Bytes and Barks

D'Ambrosio, Hill win NJPA Award

Asbury Park (NJ) Press staff writers Paul D'Ambrosio and Michelle Hill have won the 1994 New Jersey Press Association's Business Writing Award.

It is the group's top business award, which comes with a $1,000 check.D'Ambrosio and Hill won the award for their four-part series on redlining in New Jersey, entitled, "The Mortgage Game."

The computer and statistical work, as well as mapping, was done by D'Ambrosio, the paper's CAR coordinator. The series was researched and written by D'Ambrosio and Hill.

The series was also named runner-up for the NJPA's Enterprise Writing Award.

Send us your stories

Looking for someone to listen to your nightmare experiences in negotiating for data? Interested in sharing what you did well, and perhaps not so well, while pursuing your latest computer-assisted reporting venture? Drop us a line.

NICAR is interested in publishing your stories in *Uplink* and sharing them in other forms through our on-the-road seminars.

If you'd like to submit an article for publication, simply e-mail it to Matt Reavy (c598895@showme.missouri.edu). Articles specifically not for publication should be sent to Jennifer LaFleur (jenster@aol.com). Either can be reached by phone at (314) 882-0684.