

Giving credit where credit is due

By DAN WOODS
The (Raleigh, N.C.) News & Observer

Now that the first part of the invasion is complete and computer-assisted reporting has a toehold in newsrooms around the country, a new question is facing editors and reporters: How do we apportion credit for computer-assisted stories?

The taping, datacleaning, reportwriting and spreadsheet analysis done to create computer-assisted stories straddles the technical and journalistic worlds. Some of the work may not be reporting, but most of it clearly is. As of yet, no clear trend for attribution has emerged.

That the problem has arisen at all is a sign of the coming of age of computer-assisted reporting programs at newspapers across the country. In most newsrooms, reporters and editors first worry about getting a few projects into the paper, and any issues of credit are secondary.

But at a growing number of newspapers, including *The Dayton Daily News*, *The Hartford Courant*, *Newsday*, *The News & Observer* in Raleigh, and the *St. Louis Post Dispatch*, computer-assisted reporting is enough of an everyday activity to call for a policy on how to assign credit and tell readers who did the work on a story.

Possible solutions range from no credit lines

would be required for a data analyst, unless he or she also wrote the story. But if bylines are motivated by any of the other reasons, then some sort of credit should appear.

An informal survey of newspapers revealed several approaches. At *USA Today*, no credit is given for routine computer analysis. On big projects or analysis so complex that the method of analysis is an important part of the story, attribution appears as part of a credit box or in the box that explains how the analysis was performed.

USA Today uses this approach for its annual thrift rating story, which first appeared in 1988. Shawn McIntosh, special projects editor at the paper, said the credit boxes serve the purpose intended for them: directing readers to the right source.

"When we first did the bank analysis four years ago, the credit boxes explained who made the critical decisions," McIntosh said.

The flurry of calls from bankers

Political tycoons of '92 campaign
More than \$30 million has poured into GOP and Democratic committees in unlimited "soft money" contributions. Despite limits on giving to candidates, individuals and groups may give any amount to party committees. Some are party stalwarts, some want to be part of the action. Analysis by USA TODAY'S Shawn McIntosh reveals the biggest givers.

CAMPAIGN '92 GROUP CONTRIBUTIONS
Groups take advantage

at all to a special byline for data analysis. In between these two poles are tag lines at the end of stories ("Marie Smith contributed to this story"), separate boxes with attribution for the data analysis and co-bylines.

The issue begs the question of why bylines are put on stories in the first place. By and large, bylines exist to show who wrote the story, who reported the story, who is responsible for its accuracy and whom readers should contact with questions and tips.

If your newspaper includes bylines only to indicate the author, then no credit

who wanted to argue about the formulas used in the story came to the people who did the data analysis, rather than to the reporters, who wrote and reported the story but weren't able to explain the details of how the analysis was designed and performed.

"We're weird here because the data analysts are editors, so typically there is no credit for the editor," said McIntosh. "We get all the credit that editors usually get, which is not much, which is fine by me." She added that the only time any friction over credit for a computer-assisted story had arisen was with respect to awards won for the stories. Because the data analysts weren't mentioned in the bylines, they weren't always recognized when awards were granted, which raised some hackles.

This approach has two problems. The first is that data analysis is reporting. The analyst, in effect, interviews the data using report writers, database managers and spreadsheets and comes up with information

The

Copyright 1992, The Hartford Courant Co.

Home release is b

By BRANT HOUSTON
and LYNNE TUDNEY
General Staff Writers
© 1992, The Hartford Courant

More than 10,000 inmates from the state department of corrections' supervised home release program got out of prison and kept on going.

That's how many criminals escaped from the program during the past five years, and the program is being reviewed by a computer-assisted review of about 10,000 department records.

Those who have disappeared were serving time for sex crimes, robbery, arson, assault, drug crimes, burglary, petty theft, violations, failure to appear for trial, and even escape.

And the number is undoubtedly rising. The figures for escapes could be higher, because the department's official count is not based on accurate records.

And some people charged with supervising the program are discouraged.

Some of the supervised release program was created 10 years ago to smooth out the state's prison system.

Webster

15 Lawyers Defended, Sued Same State Fund

By Terry Ganey
Data Analysis by George Landau
Of the Post-Dispatch Staff
© 1992, St. Louis Post-Dispatch
FIRST OF TWO PARTS

WILLIAM ROUSSIN JR. is a lawyer from St. Louis who has made a good living serving at the plea sure of Attorney General William L. Webster. He has been doubly lucrative.

Roussin is one of 15 lawyers around the state who aid to defend an obscure state fund against workers' compensation claims, while at the same time bringing suit against the same fund.

The "Data Analysis" byline seems like a good solution because it describes exactly who did what.

that in many cases makes up the bulk of the story. Generally, the analyst works with reporters to formulate questions and then manipulates the data to answer them. "It's the same as going into documents. It's just the same," said Brant Houston, a reporter at *The Hartford Courant* who does most of that newspaper's computer analysis. Even if an editor is performing an analysis, the task is still reporting and credit should be awarded.

The second problem is that although readers are the most important constituency for any newspaper, other reporters, editors and publishers read and take note of what appears underneath each reporter's byline.

"It's not a matter of vanity; it's not a matter of modesty. The crucial issue is that people are doing work in a new field and if credit isn't given then it will give the impression that they aren't getting journalistic work done," said Houston.

"Here I am in a pilot program really getting things moving, and you can run into editors who aren't giving you any credit. Then upper management says 'What are they doing? We are giving them money and equipment and taking this risk. Where's the product?'"

Houston insists on a byline when he does a substantial amount of work on a story, and that policy has been informally adopted in *The*

Courant's newsroom. Houston said that when he started the program of computer-assisted reporting three years ago, some editors were reluctant to grant credit because they didn't realize what was involved in the analysis.

"Certainly, there are editors who think that you hit a magic button and the story appears with no effort at all," said Houston. "Editors are more sensitive to it now. It's a challenge to make people understand how much work you are really doing."

While granting a co-byline to the data analyst conveys that that person has contributed to the story, it doesn't make clear who did what. The *St. Louis Post-Dispatch* has addressed this issue by granting a "Data Analysis" byline.

George Landau, the investigative reporter who does most of the computer analysis at the *Post-Dispatch*, said that the data analysis byline was aimed at other reporters, as well as readers.

"The 'Data Analysis' byline allows readers with comments to contact the right person (and keeps me from having to pass along dozens of questionable 'tips' to my colleagues)," said Landau. "It also announces to staffers that the story relies at least partly on computer analysis. I mean, reading such stories is one of the best ways for reporters to start thinking about what a computer can do for them."

When working on a computer-assisted story where he isn't involved in the reporting, Landau rarely takes a hand in writing the first draft. He does however insist on reading the story at least once to make sure that the information is being interpreted correctly.

Here is my recommendation: At *The News & Observer*, we have decided to adopt a policy of granting a co-byline to those who perform the data analysis for a story. For the first two years of our program, few bylines were granted. Attribution for the data analyst, if it came at all, usually appeared in a tag line at the end of the story.

The main problem with this approach is that for most computer-assisted stories the data analysis creates the story. It's not fair or accurate to acknowledge that work in a tag line, which is generally present to indicate a minor contribution to a story.

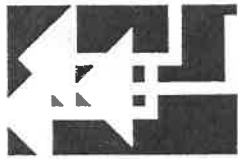
The "Data Analysis" byline seems like a good solution because it describes exactly who did what. The reporter gets the byline for the traditional work of reporting and the analyst gets credit for the analysis.

I am uncomfortable with this solution for the long term, because eventually using computers to help report on stories will be routine and not worthy of special mention. The data analysis should be recognized because it is reporting, not because it involves computers.

Furthermore, as newsrooms attract more and more specialists to cover abstruse fields, it will be common to have highly trained economists or literary critics on staff. Should they get "Economic Analysis" or "Deconstruction" bylines? I think it's better to call us all reporters, give us bylines, and be done with it.

Uplink

MISSOURI INSTITUTE FOR
COMPUTER-ASSISTED
REPORTING



We welcome your success stories, your problems, your ideas and insights into computer-assisted reporting. Please write or call.

120 Neff Hall
University of Missouri
Columbia, Mo.
65211
(314)882-0684

Two weapons fight unfair database pricing: old-fashioned haggling and knowing the law

By SANDRA DAVIDSON SCOTT, Ph.D., J.D.

Recently I got a call from a man in Texas who wanted computer tapes of Missouri's drivers license records. He had called the Department of Motor Vehicles. The price quoted to him? \$9,000.

Three years ago, MICAR and the *St. Louis Post-Dispatch* got copies of DMV tapes. When first asked the cost, the DMV said \$7,000. Fortunately for us requesters, Missouri has one of the nation's best statutes for keeping costs low. Missouri Revised Statutes § 610.026(2) limits fees to duplication costs and staff time only: "Fees for providing access to public records maintained on computer facilities...shall include only the cost of copies and staff time required for making copies."

That statute gave MICAR and the *Post-Dispatch* ammunition to take aim at the high price. Quickly the DMV dropped its price to \$1000. After a meeting attended by then-Director of MICAR Elliot Jaspin and myself, the attorney for the *Post-Dispatch*, and a host of persons from the DMV (including the Director of Revenue, who supervises the DMV, and computer experts), the asking price dropped to \$550.

Negotiations on costs in Missouri are fairly straightforward because only two questions on costs arise — 1) duplication costs and 2) staff time required for the requester's project. Period.

Jaspin, ever a prankster, later popped into my office to tell me, "The DMV isn't going to give us the tapes for \$550." He waited until he saw steam and then said, "Nope. \$440." The process had taken less staff time than anticipated.

For nearly 4 million records, \$440 wasn't bad.

So why the \$9,000 asking price to the requester from Texas? I asked that question of the current Director of Revenue.

"Maybe he talked to the wrong person?" came the reply.

Maybe. But the \$7,000 asking price first quoted to MICAR a couple of years ago came to mind.

But even \$7,000 was cheap compared to the \$30,000 price quoted to the Texas requester for some Texas computer records. Texas puts requesters on notice that its costs may be Texas-sized. Under Texas Revised Civil Statutes article 6252-

17a, § 8(b), "The costs of providing the record shall be in an amount that reasonably includes all costs related to providing the record, including costs of materials, labor, and overhead." "Overhead" could include helping to pay for the building and utilities!

Why would state agencies ask such high prices, though, even in states with user-friendly cost statutes?

My observation is this: Agencies sometimes ask a higher price than statutes warrant, and if the requester pays it, that's the end of it. But, if the requester balks, then negotiations may occur, and prices may drop to the maximum amount allowed by statute.

The reason that agencies sometimes quote such high prices is that some requesters pay without even batting an eyelash. To, say, a large insurance company or telemarketer, \$7,000 or even \$30,000 may be no big deal—a simple cost of doing business and not a call to arms.

The information business is highly lucrative. The Information Industry Association (IIA), based in Washington, D.C., represents more than 650 companies. Companies such as Knight-Ridder and Dow Jones are buying government data on tapes, loading it into mainframes and granting subscribers on-line access. That government might want to partake more in this multibillion-dollar industry is not surprising.

Should government have a straightforward policy of charging commercial users more for records than other users? The federal government and some states do. Of course, this entails screening requests on the basis of "purpose," something which some states do not permit.

On the federal side, the Office of Management and Budget published its Freedom of Information Reform Act of 1986; Uniform Freedom of Information Act Fee Schedule and Guidelines on Jan. 18, 1987. The report said that Congress clearly intended to distinguish between commercial and noncommercial users and to shift "some of the burden" of costs onto the commercial users.

As for states that treat commercial users differently, Kentucky § 61.690 says that the fee for a database for commercial users shall be based not only on the government's cost to produce or acquire the database, but also on the "[v]alue of the commercial purpose for which the

database...is to be used."

The Kentucky statute continues: "It shall be unlawful for a person to obtain a copy of all or any part of a database...for a:

- a) commercial purpose without stating the commercial purpose; or
- b) specified commercial purpose, and to use or to knowingly allow the use of the database...for a different commercial purpose; or
- c) noncommercial purpose, and to use or knowingly allow the use of the database...for a commercial purpose. A newspaper, periodical, or radio or television station shall not be held to have used or knowingly allowed the use of the database...for a commercial purpose as a result of its publication or broadcast unless it has given its express permission for such commercial use." (Emphasis added.)

Clearly, in Kentucky the news media are not considered "commercial." (How noncommercial some news outlets truly are, of course, is a regrettable fact.)

Arizona § 39-121.03 is more restrictive than Kentucky's law and so broadly worded that it is not clear whether newspapers would be "commercial users": "[C]ommercial purpose' means the use of a public record...for any purpose in which the purchaser can reasonably anticipate the receipt of monetary gain from the direct or indirect use of such public record." (Emphasis added.)

Arizona also provides for a trebling of costs of records, as well as attorneys' fees, if a requester "obtains public records for a commercial purpose without indicating the commercial purpose or who obtains a public record for a noncommercial purpose and uses or knowingly allows the use of such public record for a commercial purpose..."

New Mexico § 15-1-9(C) restricts all use of government's computer databases. A requester of a database has to agree, among other things, "not to make unauthorized copies of the database"; "not to use the database for any political or commercial purpose unless the purpose and use is approved in writing"; and "not to allow access to the database by any other person unless the use is approved in writing..." The requester also must "pay a royalty or other consideration to the state."

New Mexico law adds:

.....
continued on page four

L.A. Times tackles Department of Transportation data

By HSIAO-YIN HSUEH
MICAR
Missouri School of Journalism

One morning in February 1991, Sacramento, Calif., residents woke to a thunderous blast when a fully loaded gasoline-tanker truck rolled and skidded 100 feet and landed on a parked car at the edge of a tranquil suburban neighborhood. On July 14, 1991, a Southern Pacific freight train filled with toxic chemicals derailed at a bridge near the small town of Dunsmuir, Calif., falling 40 feet into the Sacramento River below.

To many, these might seem to be two incidents with no connection. One damaged property for blocks and burned two homes to the ground. The other spilled tons of weed killer into the Sacramento River, killing virtually every organism along 40 miles of river. Michael Parrish of the *Los Angeles Times* sensed something might be wrong with the hazardous material transportation in this country when he saw the two disasters.

Sparked by the thought, he began researching the transportation of hazardous materials, or "hazmat," by both rail and truck, to put together a picture of the national scene. Parrish's speculation proved to be on target when he obtained computer tapes covering the past ten years from the U.S. Department of Transportation's Hazardous Material Incident Reporting System. A total of 67,657 spills had been reported to the Department of Transportation from January 1982, through December 1991.

"There is a story here," Parrish thought.

Parrish needed someone to transform the data on those computer tapes into comprehensible information. He enlisted the help of Richard O'Reilly, the *Times*' director of computer analysis. For the past three years, O'Reilly has devoted his time to doing data projects for the reporting staff.

O'Reilly took the tapes from Parrish and used an IBM mainframe computer and the data analysis software SAS to conduct the research. Unlike database programs which use structured query language to sort data, SAS needs a tailor-made program to analyze the data. Each time O'Reilly analyzed a new set of data, he needed to write a new program. But SAS helped him do hundreds of cross-tabs and frequency counts on

which areas of the country had the biggest problems with hazardous material and whether the particular kinds of chemicals had problems with particular kinds of containers.

A number of significant findings emerged. He found that from 1982 to 1991, hazmat incidents rose 37 percent. Incidents involving trucks, which carry most of the hazmat, went up 34 percent. Injuries to people as a direct result of the truck spills soared 374 percent. Incidents on the nation's railroads were up 36 percent. And almost all the deaths, 106 out of 108, involved tanker trucks.

O'Reilly had talked about the structure of the database with people at the Transportation Department earlier to avoid pitfalls. Still, he encountered problems.

There were three sets of data within the database: the incident report, the containers report and the remarks report. In a train accident, each car of the train would be counted as separate spill, but there was actually only one train wreck. In the container database, for example, if there was a glass jar inside a cardboard box and the box was one of four within a larger carton, there would be a separate report on the same spill for each of the containers. Each layer the material leaked through was counted as a different spill.

To solve this problem, O'Reilly spent hundreds of hours producing a summary report of all the different findings and sent it to the Department of Transportation for spot-checks to verify his methodology and findings. The people there got the answers as O'Reilly.

With the data in hand, reporters Parrish and Tom Furlong interviewed police officers, government officials, highway accident investigators, critics, hazmat specialists, engineers, rail inspectors and package delivery workers.

The result was "Danger in Transit," a three-day series in the *Times*. Appreciation letters and phone calls started coming in soon after the series appeared Sept. 20-22, 1992. Readers told the *Times* that they didn't realize the danger of the hazmat transportation until reading those articles. "That (computer-assisted reporting) was terrific research," said Furlong, whose article in this series was seen on Sept. 21. "Using that information, we were able to go out and document the accidents and explore what had happened... That may be the best journalism you can do."

Readers
told the *Times*
that they
didn't realize
the danger of
the "hazmat"
transportation
until reading
those articles.

Unfair database pricing

.....

"[A]ny person who reveals to any unauthorized person information contained in a computer database or who uses or permits the unauthorized use or access of any computer database is guilty of a misdemeanor and upon conviction the court shall sentence such person for a definite term not to exceed one year or to payment of a fine not to exceed \$5,000 or

both. Such person shall not be employed by the state for a period of five years after the date of conviction."

Fortunately for the Texan who wants Missouri DMV records, Missouri won't ask him anything about his purpose or whether he's a commercial user; it won't make him sign a form saying that he won't make any "unauthorized" copies of the information or let any unauthorized person use the information; it won't threaten to make him pay three times the

amount if he makes commercial use of the records but didn't declare his commercial use; and it won't fine him \$5,000 and/or throw him in jail.

But Missouri may make him negotiate for the low price mandated by statute for any requester.

Scott, an attorney with a doctor's philosophy, teaches communications law at the University of Missouri.

Getting rid of those pesky duplicate records

By ELLIOT JASPIN
Cox Newspapers

One of life's more vexing problems — along with faulty plumbing and telephone solicitations — is getting rid of duplicate records. You can, of course, use the **DISTINCT** command, but it has its limitations. If you want to eliminate the second record with a duplicate field in a list, **DISTINCT** has no way of distinguishing between which record came first and which second.

The only alternative, short of manually deleting each record, is to use the programming language that is part of any database software. While the following example is done using the procedural language in XDB, the same approach can be modified to work with PAL in Paradox or any of the xbase products.

Before writing any code to accomplish this task, imagine how you might do this if you were working with a stack of papers instead of electronic records.

You would first sort the records to make sure that the first record you wanted of any duplicates would always appear first. Next you would take the first record off the stack, and compare a key with the second record. If the two fields matched you would throw away the second record. But if they didn't match, you would save the second record and use it to match against the next record. You would then repeat the process until you had gone through all the records.

While the instructions to the computer will appear different from the way it is described above, the method is the same.

Let us assume our list of records looks like the following table, called Doubles:

Key	Company	Last	First
57	Ajax	Smith	John
44	Widget	Clinton	Bill
44	Widget	Roebuck	Bubba
32	Widget	Wheedle	Skeeter
88	Rebar	Lybdenum	Molly
88	Rebar	Stang	Arnold

The first step is to select a stack of records and sort them according to our criteria. In XDB this means running a select statement and assigning the result to a label.

Q1: SELECT * FROM DOUBLES ORDER BY KEY, COMPANY, LAST

Just as with our hypothetical stack, you then select the first record and load the comparison value into a variable.

FIRST Q1

< This selects the first record, and

COMPVAL = Q1.KEY < the value in the key field is loaded into compval.

Our next step is to select the second record from the list before we begin repeating the process.

NEXT Q1

To loop through the entire list we can use a **WHILE** statement. In this case the **WHILE** statement is constantly checking to see if the next record grabbed from the list is the last record in the list. In XDB, determining the last record is done by checking the "stat" value. If **STATUS** is "1" then there are still more records. But if the status value changes to "0" we know we have reached the end of the list. What we are saying in the next statement is to keep taking records from the list until there are no more records — until the value of status is "0."

WHILE (STATUS[Q1])>0

We're now ready to make the actual comparison, using an "if-then-else" statement.

If the two values match then delete the second record, grab another record, make it the comparison value and repeat the process. **ELSE** makes the second record the comparison value.

IF COMPVAL = Q1.KEY < If they match,
DELETE Q1 < delete the second record,

NEXT Q1 < get another record
COMPVAL = Q1 KEY < make it the comparison value.
ELSE < Otherwise make the
COMPVAL = Q1 KEY second record the
ENDIF comparison value.

Whether we decide to delete a record, the last thing we need to do is get another record.

NEXT Q1

To repeat the whole process we end this short routine with an **ENDWHILE**. That forces our program to loop back to the **WHILE** statement and repeat the process until, of course, we run out of records.

Note that this method only deletes the first duplicate record. If more than one duplicate exists, another approach must be used.

Now if we could just write a program to prevent salesmen from hawking life insurance by phone...

The only way, short of manually deleting each record, is to use the programming language that is part of any database software.

Bits, bytes and nibbles

The CompuServe Journalism Forum will host a two-week-long online lecture about computer-assisted reporting sometime in January. The exact date for the lecture has not been set.

Chris Feola, of the *Waterbury Republican-American* in Connecticut, will lecture about setting up your own computer-assisted journalism program.

Larry Krumenaker, a freelancer who's written about online research for *Quill* and *PC Today*, will give several days of lectures about online research.

The lecturers will post their lessons on the computer bulletin board during the first part of the week and then devote the end of the week to answering questions.

For those who don't already dial in to CompuServe, call 800-848-8990 for information.

• • • • •

It's virtually impossible to keep track of the growing number of databases out there, much less sample each one for its usefulness.

That's where *The Database Files* can help. This new newsletter includes CD-ROM and software reviews but its focus is online databases.

There are currently more than 5,000 online

databases in existence, said John Ullmann, publisher of *The Database Files*. The newsletter will inform readers about databases which are valuable to journalists but usually unknown because they aren't offered by major vendors, such as Dialog.

The Database Files features an in-depth review of a database every issue, including costs, pitfalls and instructions on how to link up. A regular column on bulletin boards profiles "one or two boards that have content or clientele that would be useful for newsrooms to know about."

The November issue had a meaty article on Westlaw. This isn't the most obscure database, but the authors answered just about every question imaginable. We look forward to more such thoroughly researched reviews with more esoteric, equally valuable databases.

The Database Files is published six times each year. Subscription costs \$100, or \$95 prepaid, and includes one or two special supplements on selected topics in addition to the regular newsletter. Ullmann said he will provide a complimentary copy for review.

For information write to *The Database Files*, 5622 Wood Lane, St. Louis Park, MN 55436, or call (612) 925-6210.

■ CompuServe hosts a January forum on computer-assisted reporting

■ *The Database Files* tracks and reviews valuable but obscure databases

Uplink

MISSOURI INSTITUTE FOR
COMPUTER-ASSISTED

120 Neff Hall : REPORTING
University of :
Missouri :
Columbia, Mo. :
65211 :