# Missouri School of Journalism
# Washington Program
# Winter 2008

**Professional Skills Component**
Computer-Assisted Reporter
Gannett News Service

**Professional Analysis**
Best Practices and Gannett News Service:
Achieving successful computer-assisted reporting

# Grant Smith

## <u>Project Committee</u>

David Herzog, Chair
Charles Davis
Wes Pippert

## *Table of contents*

## Introduction

I began graduate school without a real clear idea of what I wanted to do. I knew I wanted to further develop my skills I learned as an undergraduate in journalism school and those I used as a reporter for a newspaper in West Virginia, but at first I had trouble narrowing my focus for graduate school. I enrolled in the newspaper reporting and writing sequence, knowing little more about computer-assisted reporting than my desire to be involved in it.

During my first semester I took Mass Media Seminar with professors George Kennedy and Scott Swafford. I learned about journalism theory and the key Elements of Journalism which led me to develop a deeper understanding of journalism and its role in democracy.

At the same time I was a teaching assistant for Dr. Charles Davis for his class teaching undergraduates the Principles of American Journalism. This work further reinforced these notions of journalism and democracy as I helped those undergraduates do the same.

The following semester I took Computer-Assisted Reporting with David Herzog, which opened up a new world of skills and possibilities for reporting in-depth and important stories. I learned to compute and analyze data, look for patterns and illuminate previously hidden stories.

During the same semester, I took Investigative Reporting with Brant Houston, which led me into a practice of journalism in which I was not very adept. Those skills were vital in this project and will be for my career as well.

In the summer of 2007 I took Quantitative Research with Tayo Oyedeji and learned to use statistical analysis as a computer-assisted reporting tool. Thankfully, each

student's interests lay outside academia, and Mr. Oyedeji catered to those interests. Mine lay with using statistics as another tool for computer-assisted reporting.

That same summer I worked as an assistant city editor at The Columbia Missourian, which further developed my awareness of the conflict between in-depth reporting, immediacy and managers' priorities.

In August 2007 I completed the Investigative Reporters and Editors GIS Mapping Boot Camp, which laid the groundwork for furthering these skills in an independent study course on GIS mapping for journalism with David Herzog this semester. This course vastly expanded my understanding of maps and how they can be used to interpret data for news reporting.

Last semester I worked as a data analyst at the National Institute for Computer-Assisted Reporting, under the tutelage of data library director Jeff Porter, who allowed me freedom to develop my abilities while at the same time teaching me new skills on the job.

This project has been a long time coming. I majored in journalism as an undergraduate at West Virginia University, but did much coursework in other areas to better inform my reporting. I minored in international and comparative politics, and took classes in economics and development, including a class that focused on the economics of Cuba as a case study.

This project was the natural conclusion to my graduate work; it has further developed my skills as an investigative and computer-assisted reporter in a competitive, nonacademic atmosphere and helped me learn how to productively apply the skills I've thus far acquired.

My project is divided into two sections: a professional component and a professional analysis.

For my professional component, I worked four days a week for 13 weeks as a computer-assisted reporter for Gannett News Service in Washington, DC, under the supervision of database editor Robert Benincasa. Most of my time was spent working with Projects Team reporter Larry Wheeler compiling U.S. Environmental Protection Agency data from multiple databases for sewage violation enforcement analysis.

This was a greater undertaking than I first imagined and well beyond the scope of anything I have done before. It required an in-depth knowledge of how U.S. EPA prosecutes enforcement in general, and more specifically how the federal agency ensures compliance under the Clean Water Act, through which publicly owned treatment works are regulated and permitted. It also required an airtight understanding of how the EPA records and stores federal and state agency data, as well as an intimate knowledge of each of the myriad quirks which must be accounted for in performing any broad analysis. I read hundreds of pages of manuals and traded many phone calls and e-mails with EPA data wizards to learn more about the structure of the data.

Needless to say, this was a challenge for me and I know I have grown tremendously in my ability to work with data and my analytical and critical thinking skills were challenged each day a bit more than the day before. This project is among the three to four projects of this caliber Gannett News Service does per year, and will eventually culminate in a microsite with a complete story package and searchable database for the Web, for each GNS news outlet to use all or portions of in their own Web, television or print publications. As such, I have had the opportunity to design a

database with searchability in mind, which has been enjoyable and a new experience that

will no doubt serve me in the near future.

I have also worked with other reporters on their data, and built a Intranet

application for searching contacts at all the Gannett newspapers and television stations.

The sewers project, originally slated for an April 15 publication date, has been

pushed back to May 7. At the risk of going overboard, I shall be as thorough as possible

with the materials I include in this report to better illuminate my committee of my

activities at GNS, since the project will not be published until after my defense date.

My professional analysis has been quite an instructive exercise as well. I

interviewed six journalists from four different news organizations that have won the

Philip Meyer Journalism Award for reporting that employs precision journalism,

computer-assisted reporting or social science research methodology. I also used news

reports and official sources to explain the current media environment.

My goal was to drill down into the past to learn how these award-winning

journalists were able to accomplish such investigations through current industry

downturns, and what that might mean for reporters and business management.

This project has supported my ultimate goal to be a reporter at a news outlet that

can offer me the opportunity to do important in-depth investigative reporting. Computer-

assisted reporting is a vital component of both this kind of journalism and my ability to

give the public the tools they need to be self-governing.

I'm a writer at heart, but these skills are essential for adding depth and authority

to my reporting. These skills, properly employed, have the power to inform about the

consequences of policy decisions and empower the public to effect change.

# Field notes from the Washington Program
# January 21, 2008 - April 24, 2008

The notes on the following pages are weekly reports I filed with my

committee regarding my work at Gannett News Service, my professional

analysis and a few of the seminars I attended.

Week 1 — January 28

Hello all.

This is my first set of field notes. We attended seminars each day this week. I enjoyed Brooks Jackson's talk about <u>Factcheck.org</u>. It was pretty interesting to see what they do and hear about the kinds of facts they choose to check. I was also impressed by the organization's circulation -- about 50,000 per day now, with a peak of 400,000 during the last election. Bill Kovach and Geneva Overholser's talk was great too. I enjoyed their thoughts on the future of journalism and engaging the audience. Marisa Katz at was also fascinating. I enjoyed the tour and her discussion about how the business works and the new things they're working on. Charles Lewis at Hearst was excellent -- he offered very practical information about how to get a job. I appreciated his frankness and experience. Terry Bracy was though provoking as well. He wasn't exactly what I was expecting from a lobbyist and I felt he helped put things in perspective as well as give us an excellent breakdown of power in Washington.

I'm looking forward to starting my job at Gannett tomorrow. I'll report on that next week.

I haven't yet begun my research -- still crafting my interview questions. That's my big priority this week.

Week 2 — February 4

Hello everyone.

Last week was great. I'm doing some pretty cool stuff at Gannett, really getting into the thick of things. I'm developing a Web database application for the Intranet, learning PHP and MySQL. I finished two drafts of the searchable Web database last week and am further developing it for when I receive additional data we want to be able to search. I've learned an incredible amount of PHP in the last week -- a skill I'll be able to keep developing and using throughout the semester and beyond.

Also, on Day 1 I began working with another reporter on an EPA dataset. We have a number of tables that aren't designed to talk to each other, and that's exactly what we need to do. I've been learning a lot about how to do that and get the results we desire. We've been looking at hundred of facilities that are administered by different programs. The idea is we want to be able to look at individual facilities and pull data about each facility from a bunch of different tables. I've really been doing most of the data crunching on the tables so far, and I think it's going to be a pretty cool story. I'll continue working on that this week.

My research is a little slow. I'm going to finish up my interview questions tonight and contact my interview subjects this week.

Week 3 — February 10

Hi again!

Things are going well at Gannett. I'm basically working on the same things as I reported last week -- developing a Web database for internal use and analyzing close to 20 different tables that comprise two different EPA databases.

The Web work is going well. I learned quite a bit more PHP last week. I'm beginning to feel more confident in my coding and I understand much better how it works. It really works much like computer code, rather than html code. I'm on my third draft and kind of in a holding pattern now until I receive the second dataset Gannett wants online. While I'm waiting, and when I'm not working on the EPA data, I've been tweaking the code and trying to make the database ready to handle the new data and more useful to the user.

The EPA pollution data is going very well. We (me and the reporter I'm working with on the story) made some great progress last week. One big problem with the tables is that multiple tables don't join well together. One or two join well, but any more than that and records start duplicating and it gets way too jumbled to make any sense and become unusable. So a lot of my time has been spent figuring out what we want to know from the data, and strategizing ways to get the data to talk. We're almost to the point where we can start reporting off the data. We're waiting on some explanations about the data from EPA, and waiting for an update on the data (it only goes to Feb. 2007 right now). One problem we've encountered is that some of the data does not agree with the data EPA has on its searchable Web site, so we've been trying to rectify that to make sure

we're getting the full picture. They're definitely keeping me busy each day.

I'm still behind on my research, but I won't be after this week.

Week 4 — February 18

Greetings!

Last week was great. The EPA data analysis has come a long way from a week ago. I had a problem with combining duplicate records last week, because I had some facilities that had multiple IDs. It wasn't as simple as stripping out the extra IDs, however, because we need those to be able to refer back to to other tables. After struggling with it for a while, I figured it out. I'm working on some other things with the data now, while we wait on a couple answers from the EPA about the data. Then I'll be able to say "these facilities were fined this much by the EPA in the past five years, and here was what enforcement was like, etc." That sounds simple enough, except the data is really complicated, and the databases aren't designed to talk to each other. That's what I've been working on quite a bit -- getting the two databases to talk to each other seamlessly for a thorough analysis. I've just a few steps left on that and we're ready to rock. I've essentially created a new table containing data from a number of different tables from the two databases. We'll be having a meeting about the project today. Also, I've set up a couple interviews for my professional analysis for Friday, and I'm waiting to hear back from a handful of others. I'll have a solid portion of my interviews done by this weekend.

Week 5 — February 25

Good morning!

Last week was pretty productive for me. I found a bunch of new problems with our EPA pollution data, and had to back up a few steps to solve those problems. It's a little frustrating at times, because it seems as soon as one problem has been solved, two more problems surface. But that's all part of thoroughly vetting this data, and each step in the process produces a dataset that much cleaner and reliable. I'm now hand-checking a handful of questionable records. Once that's done, I'll be just a few steps away from having a solid dataset for reporting.

Also, regarding my research, I'm moving right along. I had three interviews Friday -- one from each paper that won the Philip Meyer Journalism Award last year. I am also happy to report that I caught 20 fish Saturday -- 12 walleye, two crappie, a sunfish, a largemouth bass, and four smallmouth bass.

Week 5 Postscript — Feb. 26

Hey ya'll, I've attached a couple screen shots of what I've been working on.

"madmoney.bmp" is a one of the main tables I've been compiling with information from seven (so far) original tables and a number of calculations I did.

"iciscritkey.bmp" shows the SQL for a query I did a couple weeks ago, with a view of a bunch of the queries I've done in the past couple weeks, trying to get the data to cooperate. It joins two tables and narrows the criteria.
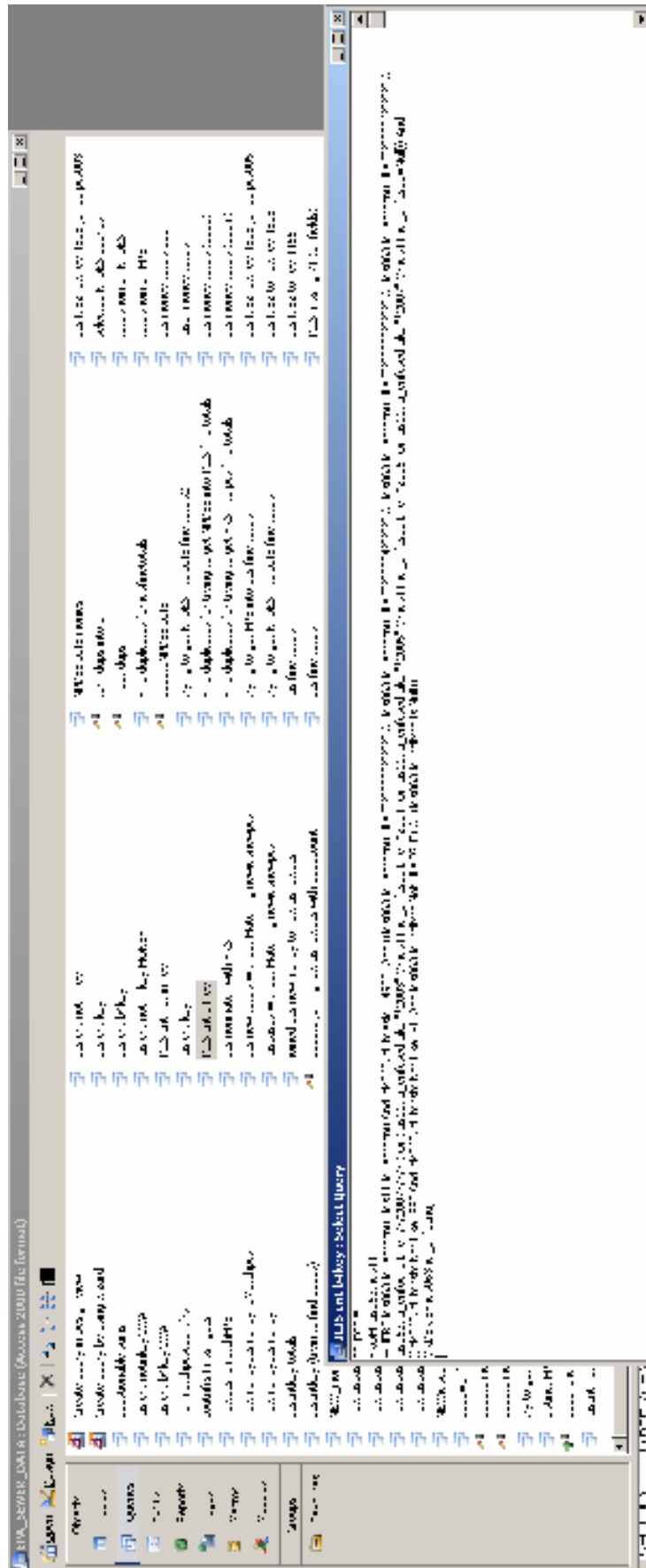
"ifstate4.bmp" shows one of the nested-if statements I came up with to deal with a problem of duplicates when trying to sum.

"ifstate3.bmp" shows another nested-if. Row 78 and 79 show how it handles duplicates. Column P is the sum of two fields for one row, Column Q adds up the rows associated with each record and displays the result only in the first instance of a record, and ignores fields where its indicator field contains an X. Column R adds Columns Q, E and F, so long as none of the indicator fields for E and F have an X.

"ifstate2.bmp" takes the number of instances for each record, and labels one as the first instance, so I can get sums from multiple rows and columns into one cell for each record. Row 22 and 23 shows how it's supposed to work.

Also, regarding my research, I'll be out of town this Friday, but the following Friday I'm hoping to get three more interviews out of the way.

| name | TOTALSUM | PCSsum | APsum | appa | util | fpax | slpx |
|---|---|---|---|---|---|---|---|
| WARDS ISLAND WATER POLLUTION ( | $2,432,499.00 | 2432499 | 2432499 | | 0 | | |
| NYC-DEP 26TH WARD WPCP | $2,309,999.00 | 2309999 | 2309999 | | 0 | | |
| WASHINGTON SUBURBAN SANITARY | $2,200,000.00 | 0 | 0 | | | | |
| NYC DEP TALLMAN IS WPCP | $2,179,999.00 | 2179999 | 2179999 | | 0 | | |
| MORRIS FORMAN WWTP | $1,900,000.00 | 900000 | 900000 | | 0 | | |
| LOS ANGELES BUREAU OF SANITATI | $1,600,000.00 | 0 | 0 | | | | |
| OKMULGEE, CITY OF WWTP | $1,370,000.00 | 900000 | 900000 | | 0 | | |
| HAMILTON COUNTY (MUDDY CREEK) | $1,200,000.00 | | | | | | |
| NEWTOWN CREEK WPC | $1,029,500.00 | 1029500 | 1029500 | | 0 | | |
| JEFFERSONTOWN WASTE TREATME | $1,000,000.00 | 0 | 0 | | | | |
| MSD WEST COUNTY STP | $1,000,000.00 | 0 | 0 | | 0 | | |
| MSD FLOYDS FORK STP | $1,000,000.00 | 0 | 0 | | 0 | | |
| MSD CEDAR CREEK STP | $1,000,000.00 | 0 | 0 | | 0 | | |
| HARTFORD WPCF | $850,000.00 | 0 | 0 | | 0 | | |
| JAMAICA WPC | $832,100.00 | 832100 | 832100 | | 0 | | |
| HUNT'S POINT WPC | $817,000.00 | 817000 | 817000 | | 0 | | |
| NYCDEP - BOWERY BAY WPC | $800,000.00 | 800000 | 800000 | | 0 | | |
| NYCDEP - ROCKAWAY WPCP | $800,000.00 | 800000 | 800000 | | 0 | | |
| NYCDEP - CONEY ISLAND WPCP | $800,000.00 | 800000 | 800000 | | 0 | | |
| NYCDEP RED HOOK WATER POLLUT | $800,000.00 | 800000 | 800000 | | 0 | | |
| DEPT OF PUBLIC WORKS | $750,000.00 | 0 | | | | | |
| WAYNE CO-DOWNRIVER WWTP | $650,000.00 | 650000 | 650000 | 650000 | 0 * | | |
| KNOXVILLE-LOVES CREEK STP | $612,500.00 | 612500 | 612500 | | 0 * | x | x |
| KNOXVILLE-FOURTH CREEK STP | $612,500.00 | 612500 | 612500 | | 0 * | x | x |
| KNOXVILLE-KUWAHEE STP | $612,500.00 | 612500 | 612500 | | 0 * | x | x |
| SUFFOLK CNTY DEPT OF PUBLIC WC | $600,000.00 | 0 | | | | | |
| ESCAMBIA CNTY-MAIN STREET WTP | $502,700.00 | 502100 | 502100 | | 0 | | |
| WOODCLIFF SEWERAGE TREATMEN | $428,432.00 | 428432 | 428432 | | 0 | | |
| NEW IBERIA, CITY OF | $286,400.00 | 0 | | | 0 | | |
| BAY CTY. - MILITARY PT. AWT | $284,341.00 | 284341 | 284341 | | 0 | | |
| HALLSDALE-POWELL-BEAVER CR. S | $266,600.00 | 266600 | 266600 | | 0 | | |
| CLARKSVILLE STP | $250,000.00 | 250000 | 250000 | | 0 | | |
| NEW IBERIA, CITY OF (ADMIRAL D | $235,000.00 | 0 | | | 0 | | |
| ATLANTA (SOUTH RIVER WRC) | $217,003.00 | 217003 | 217003 | | 0 | | |
| COLLIERVILLE NORTHWEST STP | $180,000.00 | 180000 | 180000 | | 0 | | |
| BROCKTON A W R F | $180,000.00 | 0 | | | 0 | | |
| IDABEL PUBLIC WORKS AUTHORITY | $169,000.00 | 31500 | 31500 | | 0 | | |
| MONROE, CITY OF (WPC CNTR) | $164,500.00 | 164500 | 164500 | 164500 | 0 | | |
| MONTEZUMA WATER POLLUTION CO | $159,924.00 | 159924 | 159924 | | 0 | | |
| LENOIR CITY STP | $156,000.00 | 156000 | 156000 | | 0 | | |
| GARY SANITARY DISTRICT | $150,000.00 | 0 | | | 0 | | |
| BERGEN COUNTY UTILITIES AUTHOR | $144,000.00 | 144000 | 144000 | | 0 | | |

NEWmomoney.xls — NEW_momoney

| frs | enfctpa | enfcslp | enfctsa | enfccaa | pcsnpdes | countFRS | id |
|---|---|---|---|---|---|---|---|
| 110000502386 | 0 | 0 | 0 | 100 | TX0025798 | 1 | 1 |
| 110000503615 | 0 | 0 | 0 | 2500 | TX0025631 | 1 | 1 |
| 110000504044 | 46800 | 0 | 0 | 0 | TX0047180 | 1 | 1 |
| 110000504062 | 0 | 0 | 0 | 100 | TX0032735 | 1 | 1 |
| 110000504721 | 0 | 0 | 0 | 42000 | TX0053112 | 1 | 1 |
| 110000508255 | 600 | 0 | 0 | 0 | FL0021326 | 1 | 1 |
| 110000508987 | 600 | 0 | 0 | 0 | FL0020109 | 1 | 1 |
| 110000510536 | 600 | 0 | 0 | 0 | FL0020303 | 1 | 1 |
| 110000512874 | 250000 | 0 | 2000000 | 550300000 | | 1 | 1 |
| 110000514060 | 0 | 0 | 0 | 150000000 | GA0021482 | 1 | 1 |
| 110000514239 | 600 | 0 | 0 | 0 | FL0021938 | 1 | 1 |
| 110000514578 | 600 | 0 | 0 | 0 | GA0030325 | 1 | 1 |
| 110000519467 | 600 | 0 | 0 | 0 | FL0103349 | 1 | 1 |
| 110000520286 | 600 | 0 | 0 | 0 | FL0027651 | 1 | 1 |
| 110000522275 | 0 | 0 | 0 | 0 | CA0037842 | 1 | 1 |
| 110000522685 | 0 | 0 | 0 | 0 | AR0021750 | 1 | 1 |
| 110000524987 | 0 | 0 | 0 | 60000 | CA0079260 | 1 | 1 |
| 110000525815 | 600 | 0 | 0 | 0 | FL0020532 | 1 | 1 |
| 110000526440 | 0 | 0 | 0 | 100 | AR0021792 | 2 | 2 |
| 110000526440 | 0 | 0 | 0 | 100 | ARL021792 | 2 | 2 |
| 110000538080 | 0 | 0 | 0 | 1515000000 | LA0038091 | 2 | 2 |
| 110000538080 | 0 | 0 | 0 | 1515000000 | LAL038091 | 2 | 2 |
| 110000539631 | 0 | 0 | 0 | 0 | MI0021156 | 1 | 1 |
| 110000540834 | 45000 | 0 | 0 | 0 | | 1 | 1 |
| 110000541735 | 63000 | 0 | 70000 | 35000720 | MA0101630 | 1 | 1 |
| 110000542235 | 0 | 0 | 0 | 700000 | MS0030333 | 1 | 1 |
| 110000542896 | 0 | 0 | 0 | 20000000 | KY0037991 | 1 | 1 |
| 110000543662 | 0 | 0 | 0 | 0 | KY0021466 | 1 | 1 |
| 110000547515 | 0 | 0 | 0 | 100 | LA0053716 | 2 | 2 |
| 110000547515 | 0 | 0 | 0 | 100 | LAL053716 | 2 | 2 |
| 110000551661 | 167000 | 167000 | 2000000 | 529999872 | TN0021822 | 1 | 1 |
| 110000552651 | 0 | 0 | 0 | 10000 | MA0100595 | 1 | 1 |
| 110000555863 | 0 | 0 | 0 | 2000000 | LA0041394 | 1 | 1 |
| 110000557585 | 0 | 0 | 0 | 100 | LA0038741 | 2 | 2 |
| 110000557585 | 0 | 0 | 0 | 100 | LAL038741 | 2 | 2 |
| 110000562409 | 0 | 0 | 0 | 0 | MI0023787 | 1 | 1 |
| 110000565040 | 0 | 0 | 0 | 40000 | SC0033804 | 1 | 1 |
| 110000567636 | 0 | 0 | 0 | 200 | LAL064092 | 2 | 2 |

Formula in cell H20:
=IF(A20=A21,IF(G20>1,IF(A19<>A20,1,G20),G20),G20)

IF(logical_test, [value_if_true], [value_if_false])

Week 6 — March 3

Good morning!

Last week was another excellent week. I wrapped up the most important part of

the EPA data cleaning  -- getting all the data to cooperate and do what I tell it to do --

Monday and Tuesday.

This means that each facility has at least one unique identifier -- a Facility

Registration Number or a National Pollution Discharge Elimination System number. All

records but one have the FRS number, and many facilities have multiple NPDES

numbers under a more-unique FRS number. That was really the crux of the problem --

maintaining the NPDES numbers (in order to pull information from other tables

associated with those NPDES numbers) while still being able to aggregate settlements

and other information without duplication. That's done now.

I spent a good bit of my time the rest of last week assigning latitude and longitude

to each facility. Luckily, there's an EPA geospatial dataset which had information for

most of the facilities. I used a combination of approaches for the missing records. I

looked each of them up in EPA's ECHO system, to see if it had a lat/long for the facility.

Many did, so I plugged those coordinates into my table and added a new field, GEO,

which indicates how the lat/long was derived. I used L for these records, as the lat/long

from ECHO comes from EPA's Locational Reference Table. The rest I used a batch

geocoder, which pulled up another handful of records. I coded these B. For the rest, I

used a combination of Internet research and Google satellite maps eyeballing. I looked

the facilities up to see if I could find a better address than what was included in our data,

or a map, or something to indicate where each facility was actually located. I'd then take

this information to Google Maps, turn on the satellite and see if I could locate the facility

that way. That worked on quite a few records, which I coded as M, for Map. Out of 2441

records, I'm left with 9 still to code. All but one are located in Choctaw, MS, which is not

a town or city but county. It appears many of the facilities are under tribal administration,

and the addresses just don't match up, the satellite images from that area are useless and

there's no information on the Internet that I can find that I don't already have in my own

data. So I'm a bit stuck there, but we'll figure it out.

       Also, I worked with another reporter last week on a project she was doing. I

helped her take a spreadsheet of 20,000 records with tons of fields and boil it down to the

few hundred records and handful of fields she was really looking for, so that was fun.

I'm hoping to do some more interviews this Friday and be well on my way to getting my

research done.

Week 7 — March 10

Good morning!

Last week I wrapped up a bunch of things with this data project.

I used the main table I've put together to pull out information in different formats to support the reporting of the story, and I spent a good bit of time cleaning the address fields in the data. I found better-formatted addresses in another table, and updated my table with those addresses. They still needed cleaned up and structured in a consistent way, which I did last week. I also spoke to the head of EPA's Facility Reporting System about how they came up with addresses for facilities to better understand what I need to do to clean up the data.

I also located latitudes and longitudes for those last few remaining records. I talked to Larry Wheeler, the reporting with whom I'm working on this story, about the various fields we'd like to show the reader for each facility on a data-driven Web site. So I spent some time figuring out how we'd be able to pull those fields from different tables for each record that a reader would click on.

I also updated a database of Gannett contacts from its various papers and TV stations in MySQL, and updated the PHP-driven Intranet site I built earlier this semester. Thursday was a pretty cool day for me too. I didn't have a lot of stuff looming over me, so I spent some time toward the end of the day playing with Google Maps. I managed to turn the main table I put together into a KML file, upload it and get Google Maps to display points for the 2,300 facilities we're looking at. Right now, if you click on a point, it'll only display the amount of money it has been fined in the past five years, but I did

some figuring on how to display ever field we're interested in.

I was a little under the weather Friday, so I didn't get any research done Friday after our seminars, as I'd hoped. Next week.

Week 8 — March 17

Hello all,

I got a good bit done last week. Our EPA data was updated finally, so I was able to redo most of my earlier queries to produce usable updated data for the reporting of the project. I was able to accomplish this pretty quickly because I had taken considerable notes the first time around, basically producing a handbook to the data, despite the fact that the datasets were not completely the same structurewise. I'll spend a good bit of this week pulling additional information from different tables to enhance the monetary data I have so far compiled. That new data will include facility proximity to impaired streams, watershed information, lat/long (though I took care of most of that a couple weeks ago), number of Combined Sewer Outflows (if there are any), the name of the body of water any outflows release into, flow numbers, etc.

Also, last week I compiled Senate votes for a database.

I also completed another interview for my research on Friday. Hopefully I'll finish up interviews this week and get cracking on that write-up.

Week 9 — March 24

Good morning.

I got a lot accomplished last week. I finished building a table out of the data updates, and solved the inevitable problems associated with the new data. There were some discrepancies between what was shown on the EPA Web site and what was contained in the data, and those problems are mostly solved now. I'm still trying to figure out how to indicate whether the total amount of money associated with each facility was a penalty directly on the facility, or a parent utility board, in which case the penalty or settlement is spread out over several facilities. I'm also figuring out what to do with the 1,500 or so facilities that had no enforcement actions in the past five years. I spent some time last week, and will continue this week, figuring out what additional tables and fields will be needed for the Web user interface. It's narrowed down a good bit, but I think I'll be pulling some additional information from the Clean Water Needs Survey that relates to each facility. There is still a good bit of work to finish this week with the data.

Also, I worked with another reporter last week to help him get started building a database of library statistics for each state. There are federal statistics, but have not yet been updated for 2006, so he has obtained most of the information from each state. Now it's a matter of pulling all that data together into one flat table. He's on vacation until Thursday, so I won't be working on that much this week, at least until he gets back. I also helped build a table of motorcycle fatalities for use in a project that should be released this Wednesday.

I also got my Senate Press Gallery credentials last week, in case I need them in the next few weeks as this EPA project winds down.

I had an interview scheduled for Friday, but my source never got back to me before I had to leave town in the afternoon. I'll get that interview done this week and begin writing it all up this weekend.

Week 10 — March 31

Hello all.

Things are really coming to a close soon. Last week I worked on finishing up the data cleaning for the main table I've put together, and adding fields we wish to include in the Web utility. I also spent a good deal of time testing the accuracy of my data against the EPA original data and its ECHO system online. Larry, Robert and I met to go over the project last week, which went well. A couple questions came up about the project, which I'll be working on this week. I also built a metadata table for the main table I put together and three others, which I'll need to update this week as I change a few things. I also need to combine the table with facilities that have had enforcement actions in the past five years with the table of those that don't, and add a field indicating which is the case for each facility. I also need to build another table with case enforcement summaries so people can pull up information about each case their local facility has been involved in.

The symposium went well Friday -- I thought Alan Rushbridger from The Guardian and Vincent Kojo Oppong-Nkrumah from Joy FM were both very interesting. I did another interview for my research last week. I'll hopefully finish up most of the writing this week, though I still need to supplement my reporter interviews with an interview with my boss here at GNS. GNS won the Phillip Meyer Journalism Award a couple years ago, so he'll fit right in with my sample criteria and I'll be able to add my own first-hand observations to his interview.

Week 11 — April 7

Good morning everyone!

Finalized a good bit of data last week. I updated the metadata file as I added a few fields to my set of tables I had created, and I spent a lot of time testing the validity of the data and updated as needed. It's gonna be sweet. There are a few records that the amount of money is unclear because an enforcement action was taken against a utility (which can own multiple facilities) instead of a single facility. This presents a bit of difficulty since our data is structured around individual facilities. I responded by creating a new field to indicate whether the facility was part of a case against a utility, and other fields to contain the best fine amounts we can figure from contacting the utilities themselves, so we can accurately aggregate the data. I retained the old money fields so we can still indicate the quantity of fines associated with each facility, even if the individual facility was not required to pay them. I'll be working on that more this week.

I sent Robert a database with the five tables so he can begin working on some of the program for the Web minisite.

I also spent a good amount of time working with another reporter on his state-by-state library data. He has data from each state with library statistics, but the data is not uniform. The data are in separate Excel spreadsheets, often spread over different sheets in one file. I helped him come up with a game plan for creating one flat table containing all the data, and worked with him on doing that last week. I'll likely continue working on that this week as well.

I made great headway in writing up my professional analysis this weekend and I

hope to send you all a draft later this week. I think it's looking pretty good. I still need to

interview my boss, which will hopefully happen this week.

Week 12 — April 14

Hello!

I accomplished quite a bit last week. Finalized a good bit of data analysis and produced some new tables that aggregate fines and compliance action costs for facilities. I also spent some time updating my "handbook" to the data, which basically explains how I did what I did. I'll include that with my project, as well as the metadata I updated last week for the five main tables I've produced. I also helped another reporter on his library data again. I'm really not working the data so much as working with him so he'll be able to continue working on it when I left in two weeks.

I think the main table, which contains facility information and their fines and compliance action cost is complete as of a few minutes ago. I updated the last two records for which we had questions. The trick for about 50-60 records was the type of case -- if the case was against a utility (which means it owns multiple facilities) we had to figure out how much fine or compliance action cost each facility was individually responsible for so we would not duplicate any of those costs. The EPA data does not do that.

I also made tremendous progress on my professional analysis. I interviewed my boss, Robert Benincasa, last week, and spent the weekend working on my analysis and journal article. I need to add a couple things to the article this evening and will send that tonight before I hit the hay.

It's hard to believe the semester is almost over, but I feel great about what I've accomplished here.

Week 13 — April 21

Hello!

I spent last week wrapping up sewer data updates this week and ensuring there are no questions left unanswered, so the data will be ultraclean and usable when I leave Thursday, which will allow Robert Benincasa to build the Web utility for the project. I'll finish that this week, and I'm also going to continue helping a reporter on his library data this week and that will about round out this semester. It's been a busy one -- I've learned a lot and applied all the skills I learned at Mizzou each day here.

I'm putting the finishing touches on my final project today, and will drop copies off tomorrow at the Washington Program office. I'm looking forward to defending, and I will see you all next Wednesday!

# Supervisor memorandum

April 29, 2008

Wesley G. Pippert
Associate Professor of Journalism
Missouri School of Journalism Washington Program
937 National Press Building
Washington DC 20045

Dear Professor Pippert:

It was my pleasure to host Grant Smith as an intern this spring.

During his time here, Grant helped GNS with two database projects. Both involved organizing data to serve the needs of our reporters and, ultimately, readers of data-driven stories we are producing.

I found him to have a solid work ethic, a team orientation and a commitment to accuracy—qualities that are as critical today to the practice of good journalism as they've ever been.

Grant meshed easily with other staff members here and definitely won some fans on our database/enterprise team.

Thank you for the opportunity to contribute to the Washington Program, and I hope we can keep in touch.

Sincerely,

Robert Benincasa
Database Editor

Copy: Grant Smith, Laura Rehrmann, Val Ellicott

# Professional Project Evaluation

From day one I began working with reporter Larry Wheeler on a project to analyze enforcement data about every active and major sewage treatment facility in the United States.

The idea was to examine these treatment plants and aggregate how much they have been fined, their expected cost to come back into compliance with the Clean Water Act, and the costs of Supplemental Environmental Projects (projects entities agree to complete under a settlement, that are not related to fixing the initial violation).

The data covers the two types of major sewage systems: Combined Sewer Systems and Sanitary Sewer Systems. A Combined Sewer System, or CSS, uses the same pipes to transmit both sewage and stormwater to a treatment plant. These are older designs, mostly located in the East Coast states and Wisconsin. The EPA is working with utilities to phase these out, as they can have huge human health and environmental impacts. In the event of a large rainstorm, the sewer pipes can become overloaded and, in order not to overload the treatment facility, raw sewage is discharged from Combined Sewer Outlets into surface waters.

Sanitary Sewer Systems, on the other hand, have separate pipes for sewage and stormwater. Even so, these sewer systems can discharge raw sewage into surface waters as well. These Sanity Sewer Overflows, or SSOs, are illegal under the Clean Water Act; Sanitary Sewer Systems should be designed to accommodate the expected sewage flow from the area it serves, and not expand its coverage area without upgrading its facilities. SSOs are most often caused by inflow, stormwater entering the system from the surface,

or infiltration, when stormwater enters the system underground through cracks in the pipes.

I ultimately created five tables that incorporate and aggregate information from databases available from the U.S. Environmental Protection Agency, including the Permit Compliance System; Integrated Compliance Information System - National Pollutant Discharge Elimination System;  Integrated Compliance Information System – Federal Enforcement & Compliance; the Clean Watersheds Needs Survey; an EPA database of geographic data; and a spreadsheet created for Gannett News Service by the EPA representing its complete accounting of all active and major sewage treatment facilities.

The first three of these databases are the main permitting and enforcement databases under the Clean Water Act. PCS is an older system that covers only a few states, and is structured very differently than ICIS-NPDES, which covers the remaining states. ICIS-FE&C covers all facilities that have been involved in EPA enforcement cases, so it took a great deal of time getting these three databases to talk to one another at all, and even more time to make sure that data from the databases did not duplicate each other. There was also an incredible amount of dirty data and complicated data recording that was not universal across the three databases, so a large part of my analysis was spent checking the integrity of EPA's records. Because of this likelihood for duplication, working with the data required me to structure my results in a way that could be aggregated for analysis.

Additionally, there was the issue of unique identifiers. One database uses a Facility Registration System for identification of facilities, another uses both this FRS number and a National Pollutant Discharge Elimination System permit number (NPDES)

as unique identifiers, and the third uses both a case number and an FRS number. Because

we wanted to deal with the data on a facility-by-facility basis, the case number could only

be used for pulling case-specific information. A facility can have multiple NPDES

permits, so that does not serve as the best unique identifier outside of the PCS system, so

I chose the FRS number as the unique ID for each facility. This then presented another

problem. Because NPDES numbers are used as the unique identifier under one database,

I needed to retain them to aggregate data linked to those numbers, but two numbers could

belong to the same facility, so linking them to FRS numbers created duplication

problems. I solved these with a series of nested IF statements, which I have since used to

solve other problems with this data.

I encountered numerous other data problems which I ultimately solved to produce

clean, facility based table "muggle." Muggle represents every active, major, publicly

owned treatment works in the United States. The fields in Muggle are either culled from

the numerous databases I've mentioned above, which together account for about 50

different tables, or are fields I calculated or created as indicators of certain characteristics.

Muggle previously had additional fields used to calculate monetary sums or indicate

primary records (secondary records, or those with multiple NPDES numbers, were

stripped out for the final table).

The other four tables that I compiled contain case-level information about each

facility, and are linked to the main table by the FRS number. These are mostly culled

from the three Clean Water Act EPA databases, but also contain some fields I created for

ease of use.

I created a metadata file for these five tables, as well as a rough "handbook" for this type of data harvesting and analysis.

I discovered that about a third of the 4,250 active, major, publicly owned treatment works have had an enforcement action against them in the past five years; nearly 500 have been fined for violations of the Clean Water Act. These enforcement actions against facilities total about $35 million in fines, $15.7 billion to comply with the Clean Water Act and about $40 million for Supplemental Environmental Projects. I also compiled state-by-state, case-by-case, and other specific bits of data to aid Wheeler's reporting for this package.

This project is one of three to four investigations Gannett News Service does each year, that take about two to three months each. Consequently, I feel that my time here has been incredibly well spent on a huge project that required me to learn something new every day.

In addition to the sewers project, I have also worked with other reporters on their data — one on a short-term project to pull specific records from a database of Katrina contracts and with another to help him begin compiling a database of state library data.

The library data has been another good "dirty data" project. Just as I began wrapping up the sewer data, a reporter asked for my assistance standardizing data he had obtained from each state containing library statistics. I helped him formulate a game plan to identify the fields he needed for his analysis and begin compiling those into one flat table for comparison of 2002 data (obtained from the federal government) with 2006 data. While the federal government does compile these statistics from states, the 2006 data will not be out for quite a while, so he is compiling the same data in an easy to use

format. Once he identified the fields he wanted, it was a matter of importing each table into Access and updating his master table. It's hard than it sounds —no two tables were alike in structure or formatting; some did not have unique IDs and others had data spread out across multiple sheets. Many tables required conversion to a comma-delimited format in order to strip out funky Excel formats and be able to specify field formats.

I also helped Mr. Benincasa with a few items on a motorcycle safety project running some formulas for state-by-state analysis.

I updated a database of contacts for each of the Gannett properties and developed a Web interface for internal use at GNS using PHP/MySQL coding. This not only will serve me down the road, as I continue to develop my Web production skills, but it gave me the immediate expertise to be able to design the sewer tables in a way that can be easily adapted for Web programming.

These projects kept me incredibly busy throughout the semester, with very little downtime. The few breaks I did have I spent mostly on the EPA Web site, familiarizing myself with other environmental issues and toying with other databases for future project ideas.

This semester has been everything I had hoped. I was given an important, in-depth project in which I was able to immerse myself and re-emerge with the kinds of problem-solving skills needed to tackle large investigative projects that rely so heavily on data. I'm proud of my work here and am looking forward to its publication in early May. I hope future Washington Program students look to Gannett News Service if they desire an intense and professional learning experience in computer-assisted reporting.

Metadata for SEWERS. Last updated 4/7/08

| TABLE | TABLE EXPLANATION | FIELD | DESCRIPTION | CODES | CODE MEANING |
|---|---|---|---|---|---|
| MUGGLE | This table contains information unique to each facility. A query will return one record for each facility. This information was drawn from PCS, ICIS FE&E, ICP, GEODATA, CWNS, POTWS, ECHO and Census FIPS. This table contains facilities that have had enforcement actions in the past five years AND those that have not. The "enf" field will indicate which is the case for each facility. | frs | The unique identifier for each treatment facility | | |
| | | util | Indicates whether the facility was involved in a case that comprised several facilities under the same utility or authority. If one of these codes is present it would allow us to caveat the displayed record on the Web. | x | The TOTALSUM for this facility may include money that was levied against multiple facilities under the same utility or authority. |
| | | | | y | The CA for this facility may include money that was levied against multiple facilities under the same utility or authority. |
| | | npdes | Another identifier for each treatment facility. These are unique, however, these numbers are only a secondary identifier because some facilities have multiple NPDES numbers. | | |
| | | SEP | Supplemental Environment Project cost - the cost of implementing SEPs as part of some settlement with EPA or state. This is the cost of all SEPs since 2003. | | |
| | | CA | Compliance Action cost - The cost to the facility to bring itself back into compliance. This is the cost of all CA amounts since 2003. | | |
| | | totalsum | The total amount of money a facility has been fined or agreed to pay since 2003. Not aggegable because some TOTALSUMs contain amounts from cases that involved more than one facility. | | |
| | | bestSEP | SEPs corrected for fines agreed to be paid by a utility. Aggregable. | | |
| | | best CA | CAs corrected for fines agreed to be paid by a utility. Aggregable. | | |
| | | bestsum | Totalsums corrected for fines agreed to be paid by a utility. Aggregable. | | |
| | | name | The name of the facility. | | |
| | | street | Street address, or directions to, a facility | | |
| | | city | City where the facility is located | | |
| | | state | State where the facility is located. | | |
| | | statecode | Two-digit state FIPS code | | |
| | | county | Three-digit county FIPS code | | |
| | | fips | Five-digit FIPS code representing a combining of STATECODE and COUNTY | | |

Metadata for five tables I created from databases available from the U.S. Environmental Protection Agency, including PCS, ICIS-NPDES, ICIS FE&C, CWNS and GEODATA.

| enf | frs | util | npdes | SEP | CA | totalsum | bestSEP | bestCA | bestsum | name | street | city | state |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| y | 11000517913 | | CA0107409 | | 0 | 0 | $4,100,000,000.00 | $1,000,000,000.00 | $6,200,000.00 | SD CITY PT LOMA WASTEWATER TREATMEN | 1902 GATCHELL ROA | SAN DIEGO | CA |
| y | 11000526618 | | | | 0 | 0 | $2,500,000,000.00 | $2,100,000,000.00 | $3,000,000.00 | CITY OF NEW YORK DEPARTMENT OF ENVIR | 59-17 JUNCTION BLV | FLUSHING | NY |
| y | 11001085076 | | CA0109991 | | 0 | 0 | $3,000,000,000.00 | $2,000,000,000.00 | $1,600,000.00 | HYPERION TREATMENT PLANT | 12000 VISTA DEL MA | PLAYA DEL REY | CA |
| y | 11001182912 | | PA0025984 | $3,000,000.00 | $400,000,000.00 | $1,200,000.00 | $2,000,000,000.00 | $400,000,000.00 | $1,200,000.00 | ALCOSAN SEW SYSTEM | 3300 PREBLE AVENU | PITTSBURGH | PA |
| y | 11000560081 | | IN0023183 | $2,000,000.00 | $1,860,000,000.00 | $1,117,800.00 | $4,400,000.00 | $1,860,000,000.00 | $1,117,800.00 | INDIANAPOLIS BELMONT WWTP | 2700 S BELMONT | INDIANAPOLIS | IN |
| y | 11000930407 | | MD0021741 | | | $610,848.00 | | $200,000,000.00 | $1,100,000.00 | WASHINGTON SUBURBAN SANITARY COM | 6600 CRAIN HIGHWA | UPPER MARLBORO | MD |
| y | 11000732271 | x | NE0049018 | | $16,004,000.00 | $1,068,600.00 | | $16,004,000.00 | $1,068,600.00 | WAKEFIELD WASTEWATER TREATMENT | HIGHWAY 35 N | WAKEFIELD | NE |
| y | 11000759714 | | KY0022420 | | $500,000,000.00 | $1,100,000.00 | $2,250,000.00 | $500,000,000.00 | $1,000,000.00 | MSD HITE CREEK STP | 5512 HITT LN | LOUISVILLE | KY |
| y | 11000621165 | x | KY0022411 | | | $1,900,000.00 | | | $900,000.00 | LOUISVILLE AND JEFFERSON COUNTY METF | 4522 ALGONQUIN PA | LOUISVILLE | KY |
| y | 11001005290 | | OK0028134 | | $18,500,000.00 | $1,370,000.00 | | $18,500,000.00 | $900,000.00 | CITY OF OKMULGEE | PO BOX 250 | OKMULGEE | OK |
| y | 11000052904 | | CT0100251 | | $120,000,000.00 | $850,000.00 | | $120,000,000.00 | $850,000.00 | M.D.C./HARTFORD WPCF | 240 BRAINARD ROAI | HARTFORD | CT |
| y | 11000539631 | | MI0021156 | | | $0.00 | $4,500,000.00 | $800,000,000.00 | $750,000.00 | BALTIMORE COUNTY BUREAU OF UTILITIES | 11112 GILROY ROAD | HUNT VALLEY | MD |
| y | 11000566487 | | TN0020575 | | | $650,000.00 | | | $650,000.00 | WYANDOTTE WASTEWATER TREATMENT F | 797 CENTRAL AVENI | WYANDOTTE | MI |
| y | 11002395042 | | FL0021440 | | | $0.00 | $2,800,000.00 | $400,000,000.00 | $564,038.00 | CENTRAL WASTEWATER TREATMENT PLAN | 1600 SECOND AVEN | NASHVILLE | TN |
| y | 11000543662 | | KY0021466 | $311,000.00 | $100,000,000.00 | $502,700.00 | | | $502,700.00 | ESCAMBIA COUNTY UTILITIES AUTHORITY | 401 WEST GOVERNN | PENSACOLA | FL |
| y | 11001510348 | | NJ0029084 | | | $476,400.00 | $636,000.00 | $1,100,000,000.00 | $476,400.00 | DRY CREEK WASTEWATER TREATMENT PL | 2999 AMSTERDAM F | ERLANGER | KY |
| y | 11000759705 | | KY0021504 | | | $428,432.00 | | | $428,432.00 | NORTH BERGEN TWP MUA WOODCLIFF STP | 7117 RIVER ROAD | NORTH BERGEN | NJ |
| y | 11000551661 | x | TN0021822 | $2,000,000.00 | $529,999,872.00 | $946,500.00 | $2,730,000.00 | $290,000,000.00 | $425,000.00 | WEST HICKMAN CREEK WWTP | 3174 ASHGROVE PIK | NICHOLASVILLE | KY |
| y | 11001859891 | | TN0024287 | | | $334,000.00 | $2,000,000.00 | $530,000,000.00 | $334,000.00 | LOVES CREEK WASTEWATER TREATMENT | 5760 SANDIS LANE | KNOXVILLE | TN |
| y | 11002930144 | | DC0021199 | $2,000,000.00 | $550,300,000.00 | $266,600.00 | | | $266,600.00 | HALLSDALE POWELL UTILITY DISTRICT | 4301 WEST BEAVER | POWELL | TN |
| y | 11000726251 | | TN0020656 | | | $250,000.00 | $2,000,000.00 | $2,200,000,000.00 | $250,000.00 | DC WATER & SEWER AUTHORITY | 5000 OVERLOOK AVE | WASHINGTON | DC |
| y | 11002028419 | | PA0026549 | $563,000.00 | $53,000,000.00 | $250,000.00 | | | $250,000.00 | CLARKSVILLE WATER TREATMENT PLANT | PUMPING STATION I | CLARKSVILLE | TN |
| y | 11000753364 | | LAL044008 | | | $239,000.00 | $563,000.00 | $53,000,000.00 | $239,000.00 | CITY OF READING | ROUTE 10, MORGAN | READING | PA |
| y | 11000644364 | x y | NC0024112 | | $50,000,000.00 | $235,000.00 | | $50,000,000.00 | $235,000.00 | ADMIRAL DOYLE WASTEWATER TREATMEN | 637 EAST ADMIRAL I | NEW IBERIA | LA |
| y | 11000735232 | | OH0020834 | | | $209,817.00 | | | $209,817.00 | HAMBY CREEK WWTP | 110 OPTIMIST PARK | THOMASVILLE | NC |
| y | 11001649612 | | MA0101010 | | | $190,000.00 | | | $190,000.00 | JACKSON WWTP | 225 WOOD AVENUE | JACKSON | OH |
| y | 11000668907 | | TN0074543 | $180,000.00 | $91,000,000.00 | $180,000.00 | $180,000.00 | $91,000,000.00 | $180,000.00 | BROCKTON AWRF | 303 OAK HILL WAY | BROCKTON | MA |
| y | 11002049792 | | OK0027677 | | | $180,000.00 | | | $180,000.00 | COLLIERVILLE STP | 1500 WOLF RIVER BL | COLLIERVILLE | TN |
| y | 11000557585 | | LA0038741 | | $2,100.00 | $169,000.00 | | $2,100.00 | $169,000.00 | IDABEL PUBLIC WORKS AUTHORITY | 207 S CENTRAL AVE | IDABEL | OK |
| y | 11000984211 | | NC0085359 | | $100.00 | $164,500.00 | | $100.00 | $164,500.00 | CITY OF MONROE (WWTP) | 770 RICHWOOD ROA | MONROE | LA |
| y | 11006521077 | | TN0020494 | | | $161,675.00 | | | $161,675.00 | TWELVE MILE CREEK WWTP | 4104 PROVIDENCE F | MONROE | NC |
| y | 11000928454 | | IN0020977 | $2,500,000.00 | $10,800,000.00 | $156,000.00 | $2,500,000.00 | $10,800,000.00 | $156,000.00 | LENOIR CITY SEWAGE TREATMENT PLANT | 200 DEPOT STREET | LENOIR CITY | TN |
| y | 11017411772 | | NJ0020028 | | | $150,000.00 | | | $150,000.00 | GARY SANITARY DISTRICT WWTP | 3600 W 3RD AVE | GARY | IN |
| y | 11009789050 | | TN0020141 | | | $144,000.00 | | | $144,000.00 | BERGEN CNTY UTIL AUTH WTP & SLF | FOOT OF MEHROF F | LITTLE FERRY | NJ |
| y | 11006521031 | | TN0021865 | | | $140,000.00 | | | $140,000.00 | GALLATIN STP | LOCK FOUR ROAD | GALLATIN | TN |
| y | 11006520951 | | TN0020095 | | $1,700.00 | $138,000.00 | | $1,700.00 | $138,000.00 | PORTLAND SEWAGE TREATMENT PLANT | 122 MORNINGSIDE C | PORTLAND | TN |
| y | 11000915298 | | FL0032808 | | | $138,000.00 | | | $138,000.00 | KINGSPORT SEWAGE TREATMENT PLANT | 700 WEST INDUSTR | KINGSPORT | TN |
| y | 11009725770 | | ID0021491 | | | $134,300.00 | | | $134,300.00 | CITRUS RECREATIONAL MARINA | 3209 PINE VALLEY D | SARASOTA | FL |
| y | 11000736847 | | MA0100447 | | $1,872,348.00 | $134,000.00 | | $1,872,348.00 | $134,000.00 | MOSCOW WASTEWATER TREATMENT PLAN | 2221 WEST PULLMAI | MOSCOW | ID |
| y | 11002152508 | | TX0047911 | | $32,200,000.00 | $127,000.00 | | $32,200,000.00 | $127,000.00 | GREATER LAWRENCE WASTEWATER TREA | 240 CHARLES STREI | NORTH ANDOVER | MA |
| y | 11009706103 | | TN0024180 | $128,748.00 | $32,100.00 | $126,200.00 | $128,748.00 | $32,100.00 | $126,200.00 | WILSON CREEK REGIONAL WASTE WATER T | NORTHEAST OF FOI | MC KINNEY | TX |
| y | 11000732225 | | MA0101508 | $32,100.00 | | $125,000.00 | | | $125,000.00 | SHELBYVILLE STP | SCUDDER DRIVE | SHELBYVILLE | TN |
| y | | | | | $35,000,000.00 | $115,000.00 | | $35,000,000.00 | $115,000.00 | WATER POLLUTION CONTROL FACILITY | 80 MEDINA STREET | CHICOPEE | MA |

This is a sample of the MUGGLE table, which continues on to the following page.

| statecode | county | FIPS | cnty | zip | authority | lat | long | csos | 303a | shedcode | watershed | waters | url | Etotal | ptotal | sneeds | fneeds | needs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 06 | 073 | 06073 | San Die | 92106 | City of San | 32.679167 | -117.245830 |  | Y | 18070304 | San Diego | PACIFIC OCEAN | http://oasp | 184 | 240 |  | 765792 | 765792 |
|  |  | 36081 | Queens | 11373 | CITY OF N | 40.734815 | -73.864954 | 450 |  |  |  |  |  |  |  |  |  |  |
| 06 | 037 | 06037 | Los Ang | 90293 | City of Los | 33.925278 | -118.430550 |  |  | 18070104 | Santa Mon | SANTA MONICA BA | http://oasp | 362 | 450 |  | 236751 | 236751 |
| 42 | 003 | 42003 | Allegher | 15233 | ALLEGHE | 40.476408 | -80.044211 | 21 |  | 05030101 | Upper Ohi | OHIO RIVER | http://oasp | 164 | 200 | 405 | 1654726 | 1655131 |
| 18 | 097 | 18097 | Marion | 46221 | INDIANAP | 39.727916 | -86.192583 | 132 | Y | 05120201 | Upper Wh | W FK WHITE RIVEF | http://oasp | 120 | 120 |  | 1151720 | 1151720 |
| 24 | 033 | 24033 | Prince G | 207724136 | Washingto | 38.791977 | -76.719111 |  |  | 02060006 | Patuxent | W.BR.OF PATUXEN | http://oasp | 17.679 | 30 |  | 21879 | 21879 |
| 31 | 051 | 31051 | Dixon | 68784 | WAKEFIEL | 41.272777 | -96.854444 |  |  | 10200203 | Salt. Nebr | LOGAN CREEK | http://oasp | 0.13 | 0.14 |  | 815 | 815 |
| 21 | 111 | 21111 | Jefferson | 40201 | Louisville | 38.323333 | -85.551666 |  | Y | 5140101 | Silver-Little | HITE CRK | http://oasp | 3.43 | 4.4 |  | 16234 | 16234 |
| 21 | 111 | 21111 | Jefferson | 40211 | LOU-JEFF | 38.230000 | -85.835555 | 115 | Y | 5140101 | Silver-Little | OHIO RIVER | http://oasp | 60.36 | 105 |  | 166736 | 166736 |
| 40 | 111 | 40111 | Okmulge | 74447 | OKMULGE | 35.594370 | -95.979827 |  |  | 11100303 | Deep Fork | 520700 OKMULGEE | http://oasp | 3.5 | 4.15 |  |  | 0 |
| 09 | 003 | 09003 | Hartford | 06114 | METROP | 41.731389 | -72.656389 | 44 | Y | 01080205 | Lower Cor | CONNECTICUT RIV | http://oasp | 53.48 | 70 |  | 644058 | 644058 |
|  |  | 24005 | Baltimor | 21031 | BALTIMO | 39.48286 | -76.661658 | 0 | N | 02060003 | Gunpowde | PATAPSCO RIVER | http://www.epa.gov/compliance/resources/decrees/civ |  |  |  |  |  |
| 26 | 163 | 26163 | Wayne | 48192 | WAYNE C | 42.187794 | -83.162109 |  | Y | 4090004 | Detroit. Mi | TRENTON CHANN | http://oaspub.epa.gov/enviro/fac_gateway.m |  |  |  | 0 | 0 |
| 47 | 037 | 47037 | Davidso | 37208 | Metro Wat | 36.184813 | -86.788322 | 30 | Y | 5130202 | Lower Cun | CUMBERLAND RI | http://oasp | 91.9 | 125 |  | 341468 | 341468 |
| 12 | 033 | 12033 | Escamb | 32501 |  | 30.407666 | -87.231361 |  | Y | 3140105 | Pensacol | PENSACOLA BAY | http://oaspub.epa.gov/enviro/fac_gateway.m |  |  |  |  | 0 |
| 21 | 117 | 21117 | Kenton | 41017 | N KY Sani | 39.059000 | -84.617200 | 74 | Y | 5090203 | Middle Oh | OHIO RIVER | http://oasp | 33.88 | 36 |  | 144933 | 144933 |
| 34 | 017 | 34017 | Hudson | 070470000 | NORTH B | 40.790280 | -74.994045 | 1 | Y | 2040105 | Middle De | HUDSON RIVER | http://oasp | 3.065 | 3.02 |  | 21947 | 21947 |
| 21 | 113 | 21113 | Jessami | 40356 | LEX-FAYE | 37.943472 | -84.499830 |  | Y | 5100205 | Lower Ken | WEST HICKMAN CF | http://oasp | 17.56 | 39.1 |  | 105766 | 105766 |
| 47 | 093 | 47093 | Knox | 37924 | Knoxville U | 36.008981 | -83.830452 |  | Y | 6010104 | Holston. T | HOLSTON RV | http://oasp | 3.081 | 3.3 |  |  | 0 |
| 47 | 093 | 47093 | Knox | 37849 | Hallsdale | 36.010926 | -84.063238 |  | Y | 6010207 | Lower Clin | BEAVER CR | http://oasp | 7.244 | 9 |  | 56413 | 56413 |
| 11 | 001 | 11001 | Arlington | 20032 | DC DEPT | 38.813645 | -77.027176 | 60 | Y | 02070010 | Middle Po | POTOMAC, ANACO | http://oasp | 370 | 370 |  | 160114 | 160114 |
| 47 | 125 | 47125 | Montgon | 370410387 | CLARKSV | 36.545918 | -87.356118 | 2 | Y | 5130205 | Lower Cun | BARKLEY RES @ C | http://oasp | 10.311 | 25 |  | 54294 | 54294 |
| 42 | 011 | 42011 | Berks | 196013690 | READING | 40.303500 | -75.921700 |  | Y | 02040203 | Schuylkill | SCHUYLKILL RIVEF | http://oasp | 15.66 | 28.5 |  | 90163 | 90163 |
| 22 | 045 | 22045 | Iberia | 70560 |  | 29.985556 | -91.639720 |  |  |  | Lower Yac |  | http://oaspub.epa.gov/enviro/fac_gateway.m |  |  |  |  | 0 |
| 37 | 057 | 37057 | Davidso | 27360 | THOMAS | 35.847591 | -80.113671 |  | Y | 3040103 | Lower Yac | Hamby Creek | http://oasp | 3.02 | 4 |  | 26560 | 26560 |
| 39 | 079 | 39079 | Jackson | 45640 | City of Jac | 39.060386 | -82.648915 |  |  | 5060002 | Lower Sci | SALT LICK CREEK | http://oasp | 2.18 | 2.21 |  | 11775 | 11775 |
| 25 | 005 | 25005 | Bristol | 02301 | BROCKTO | 42.046936 | -71.007217 | 12 | Y | 62 | Narragan | SALISBURY PLAIN | http://oasp | 15.73 | 18 |  | 3919 | 3919 |
| 47 | 157 | 47157 | Shelby | 38017 | COLLIER | 35.101778 | -89.714801 |  | Y | 8010210 | Wolf. Miss | WOLF RIVER | http://oasp | 2.047 | 3 |  | 16535 | 16535 |
| 40 | 089 | 40089 | McCurtai | 747454849 | IDABEL P | 33.854160 | -94.789452 |  |  | 11140107 | Upper Littl | 410200 MUD CREEK | http://oasp | 2 | 1.65 |  | 8755 | 8755 |
| 22 | 073 | 22073 | Ouachite | 71202 | MONROE | 32.436560 | -92.102924 |  |  | 8050001 | Boeuf. Ark | OUACHITA RIVER | http://oasp | 17 | 17 |  | 66392 | 66392 |
| 37 | 179 | 37179 | Union | 28110 | UNION CC | 34.950277 | -80.762222 |  |  |  | Lower Cat | Twelvemile Creek | http://oasp | 1.2 | 2.5 |  |  | 0 |
| 47 | 105 | 47105 | Loudon | 37771 | Lenoir City | 35.787320 | -84.262133 |  |  | 6010201 | Watts Bar | TENNESSEE RI | http://oasp | 2.32 | 2 |  | 4666 | 4666 |
| 18 | 089 | 18089 | Lake | 46402 | GARY SA | 41.606527 | -87.384083 | 5 | Y |  | Little Calur | GRAND CALUMET | http://oasp | 41.32 | 60.59 | 8356 | 379024 | 387380 |
| 34 | 003 | 34003 | Bergen | 07643 | BERGEN | 40.836888 | -74.033222 | Yes (Outfall |  | 2030103 | Hackensa | AQUIFER | http://oasp | 3.758 | 5.5 |  | 576808 | 576808 |
| 47 | 165 | 47165 | Sumner | 37066 | GALLATIN | 36.334532 | -86.462756 |  | Y | 5130201 | Lower Cun | CUMBERLAND RV | http://oasp | 1.069 | 1.9 |  |  | 0 |
| 47 | 165 | 47165 | Sumner | 37148 | PORTLAN | 36.584880 | -86.525123 |  | Y | 5130206 | Red. Ky. | TR-SUMMERS BR | http://oasp | 1.9 |  |  |  | 0 |
| 47 | 163 | 47163 | Sullivan | 37660 | KINGSPOI | 36.550082 | -82.574858 |  | Y | 6010102 | South Fork | SO FORK-HOLSTO | http://oasp | 7.74 | 14 |  | 9766 | 9766 |
| 12 | 115 | 12115 | Sarasota | 34239 |  | 27.315416 | -82.506083 |  | Y | 3100201 | Sarasota | PHILLIPPI CR | http://oaspub.epa.gov/enviro/fac_gateway.m |  |  |  |  | 0 |
| 16 | 057 | 16057 | Latah | 83843 | Moscow, | 46.739024 | -117.030765 |  |  | 17060108 | Palouse. I | PARADISE CREEK | http://oaspub.epa.gov/enviro/fac_gateway_g |  |  |  | 3976 | 3976 |
| 25 | 009 | 25009 | Essex | 018451649 | GREATEF | 42.716081 | -71.129461 | 5 | Y | 010700020 | Merrimack | MERRIMACK RIVEF | http://oasp | 31.24 | 52 |  | 168622 | 168622 |
| 48 | 085 | 48085 | Collin | 75069 | PLANO | 33.125980 | -96.561669 |  |  | 12030106 | East Fork | SEG 0821 LAKE LA | http://oaspub.epa.gov/enviro/fac_g |  |  |  | 41768 | 41768 |
| 47 | 003 | 47003 | Bedford | 371620530 | SHELBYV | 35.483716 | -86.466645 |  | Y | 6040002 | Upper Du | DUCK RI MI 221.3 | http://oasp | 3.64 | 5 |  | 584 | 584 |
| 25 | 013 | 25013 | Hampde | 01013 | CHICOPE | 42.152203 | -72.625347 | 31 | Y | 01080201 | Middle Co | CONNECTICUT & C | http://oasp | 10.32 | 15.5 |  | 153808 | 153808 |

At left is a sample of the ENACS table; below is HLRNC and the bottom is ENFCOAC.

| npdes | enac | enac_desc | endt | day | year | frs |
|---|---|---|---|---|---|---|
| AL002382?03 | | LTR OF VIOLATION/WARNING LT  1 | 1 | 12 | 2007 | 110001934042 |
| AL002382?22 | | ADMINISTRATIVE CONSENT ORDER  2 | 9 | 5 | 2007 | 110001934042 |
| AL002382?22 | | ADMINISTRATIVE CONSENT ORDER  2 | 9 | 5 | 2007 | 110001934042 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 10 | 28 | 2005 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 10 | 28 | 2005 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 11 | 23 | 2005 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 11 | 23 | 2005 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 1 | 3 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 1 | 3 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 1 | 17 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 1 | 17 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 3 | 28 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 3 | 28 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 5 | 1 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 5 | 1 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 11 | 1 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 11 | 1 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 12 | 12 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 12 | 12 | 2006 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 3 | 9 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 3 | 9 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 5 | 31 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 5 | 31 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 7 | 5 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 7 | 5 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 8 | 3 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 8 | 3 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 8 | 27 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 8 | 27 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 10 | 11 | 2007 | 110000507078 |
| AL002388?AG | | AO STIPULATED PENALTY      1 | 10 | 11 | 2007 | 110000507078 |
| AL002388?03 | | LTR OF VIOLATION/WARNING LT  1 | 8 | 29 | 2003 | 110000507078 |
| AL002388?03 | | LTR OF VIOLATION/WARNING LT  1 | 8 | 29 | 2003 | 110000507078 |
| AL002388?03 | | LTR OF VIOLATION/WARNING LT  1 | 10 | 3 | 2006 | 110000507078 |
| AL002388?03 | | LTR OF VIOLATION/WARNING LT  1 | 10 | 3 | 2006 | 110000507078 |

| frs | npdes | pcshqrtr | pcshlrnc |
|---|---|---|---|
| 110000502 | TX0024490 | 20051 | E |
| 110000502 | TX0025798 | 20032 | X |
| 110000502 | TX0025798 | 20033 | X |
| 110000502 | TX0025798 | 20072 | X |
| 110000502 | TX0023647 | 20072 | E |
| 110000502 | TX0023647 | 20052 | S |
| 110000502 | TX0023647 | 20053 | S |
| 110000502 | TX0023647 | 20054 | E |
| 110000502 | TX0023647 | 20061 | E |
| 110000502 | TX0023647 | 20062 | E |
| 110000502 | TX0023647 | 20063 | E |
| 110000502 | TX0023647 | 20044 | S |
| 110000502 | TX0023647 | 20071 | E |
| 110000502 | TX0023647 | 20043 | E |
| 110000502 | TX0023647 | 20073 | E |
| 110000502 | TX0023647 | 20074 | S |
| 110000502 | TX0023647 | 20064 | E |
| 110000502 | TX0023647 | 20051 | S |

| frs | enfocnu | enfcoac_text | enfcsld | enfcsed | month | day | year |
|---|---|---|---|---|---|---|---|
| 110000502 | 06-2004-22 | Administrative Compliance Orders | | 5/27/2004 | 5 | 27 | 2004 |
| 110000502 | 06-2008-35 | Administrative Penalty Order With or Without Injunctive Relief | | 12/21/2007 | 12 | 21 | 2007 |
| 110000503 | 06-2005-13 | Administrative Compliance Orders | | 6/16/2005 | 6 | 16 | 2005 |
| 110000503 | 06-2006-70 | Administrative Penalty Order With or Without Injunctive Relief | | 12/11/2006 | 12 | 11 | 2006 |
| 110000503 | 06-2005-18 | Administrative Compliance Orders | | 2/16/2005 | 2 | 16 | 2005 |
| 110000504 | 06-2002-18 | Administrative Penalty Order With or Without Injunctive Relief | 9/19/2002 | 2/13/2003 | 2 | 13 | 2003 |
| 110000504 | 06-2004-18 | Administrative Compliance Orders | | 11/14/2003 | 11 | 14 | 2003 |
| 110000504 | 06-2003-19 | Administrative Compliance Orders | | 6/12/2003 | 6 | 12 | 2003 |
| 110000504 | 06-2004-20 | Administrative Compliance Orders | | 4/5/2004 | 4 | 5 | 2004 |
| 110000507 | 04-2006-80 | Administrative Penalty Order With or Without Injunctive Relief | 2/9/2006 | 2/9/2006 | 2 | 9 | 2006 |
| 110000508 | 04-2002-45 | Administrative Penalty Order With or Without Injunctive Relief | 2/11/2003 | 2/11/2003 | 2 | 11 | 2003 |
| 110000508 | GA-N0000 | Administrative Compliance Orders | | 7/3/2003 | 7 | 3 | 2003 |
| 110000508 | 04-2006-45 | Administrative Penalty Order With or Without Injunctive Relief | | 9/15/2006 | 9 | 15 | 2006 |
| 110000509 | GA-N0000 | Administrative Compliance Orders | | 12/19/2003 | 12 | 19 | 2003 |
| 110000509 | GA-N0000 | Administrative Compliance Orders | | 6/19/2003 | 6 | 19 | 2003 |
| 110000509 | GA-N0000 | Administrative Compliance Orders | | 10/17/2003 | 10 | 17 | 2003 |
| 110000509 | GA-N0000 | Administrative Compliance Orders | | 8/18/2004 | 8 | 18 | 2004 |
| 110000509 | GA-N0000 | Administrative Compliance Orders | | 12/9/2004 | 12 | 9 | 2004 |
| 110000509 | GA-N0000 | Administrative Compliance Orders | | 2/24/2006 | 2 | 24 | 2006 |
| 110000509 | 06-2005-35 | Administrative Penalty Order With or Without Injunctive Relief | 10/19/2005 | 10/19/2005 | 10 | 19 | 2005 |
| 110000510 | 04-2005-45 | Administrative Penalty Order With or Without Injunctive Relief | 9/28/2005 | 9/28/2005 | 9 | 28 | 2005 |
| 110000510 | GA-N0000 | Administrative Compliance Orders | | 3/13/2006 | 3 | 13 | 2006 |
| 110000510 | GA-01/22/ | Administrative Compliance Orders | | 1/22/2007 | 1 | 22 | 2007 |
| 110000510 | GA-N0000 | Stipulated Penalty Assessed Against Previous Action | | 3/31/2006 | 3 | 31 | 2006 |
| 110000510 | GA-N0000 | Administrative Compliance Orders | | 3/13/2006 | 3 | 13 | 2006 |
| 110000510 | GA-N0000 | Stipulated Penalty Assessed Against Previous Action | | 10/10/2005 | 10 | 10 | 2005 |
| 110000510 | GA-01/22/ | Administrative Compliance Orders | | 1/22/2007 | 1 | 22 | 2007 |
| 110000510 | GA-N0000 | Administrative Compliance Orders | | 11/30/2005 | 11 | 30 | 2005 |
| 110000510 | GA-N0000 | Stipulated Penalty Assessed Against Previous Action | | 3/31/2006 | 3 | 31 | 2006 |
| 110000514 | GA-N0000 | Stipulated Penalty Assessed Against Previous Action | | 7/20/2006 | 7 | 20 | 2006 |

A sample from the ENFOSUM table.

*Handbook for preparing ICIS FE&E, ICIS-NPDES and PCS for analysis:*

Updated 4/8/08
▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪▪

Alter the KEY: (this is not as necessary with the new data and the EPA's POTWs table they sent us, because it has pretty thorough NPDES numbers)
      -Kill all columns but FRS and NPDES.
      -Save as "KEY2"

PCS likes to cooperate:
      -filter by State(not GU, VI, PR), Active and Major, and SIC 4952. (pcscritb4key)
      -Join to Key2.
      -Join to PCS005, narrowing by Date.
      -Call this table "pcstokey" (pcscrit key)
      -Find those that did not match the key, save for later.  (pcs crit not inkey)

ICP:
      − find whatever records are in ICP that are not also in ICIS. Save these.
      If the key is insufficient, use ICP later to get a NPDES number for records not in POTWs.

Getting ICIS to cooperate:

1)      Find and fix the records that are should have a 4952 SIC code, but don't:
      a.  Select the specific fields from ICIS06 into new table ICIS062, joining by FRS numbers in icis06 and in the "pcstokey" table
      b.  Delete the records that have an SIC number of 4952. Keep the rest and make them all 4952.
      c.  Append those altered records to a copy of ICIS06
2)      Narrow by 4952 and by state (not like (not GU, VI or PR) or ( is like Null))
3)      Join to KEY2.  (or in this second time around, use the POTWs that EPA gave us to determine if it is a major facility. What otherwise meets our criteria, but is lost when matched with POTWs should be looked up in ECHO to determine if it merely has a expired permit and is otherwise active. Keep these. From ECHO: "If the CWA permit is past its expiration date, this normally means that the permitting authority has not yet issued a new permit. In these situations, the expired permit is normally administratively extended and kept in effect until the new permit is issued.")
4)      Join to ICIS11, and narrow by both case number date and date settled (include null dates), to include old cases that have been settled in the past 5 years. Join also with ICI08, filtering out by milestone date and excluding "Enforcement action data entered." (This may still capture some we're not interested in — we'll filter them out later with a back-up check through ECHO.)
5)      Make a table of that. Call it "iciscases"
6)      Join to ICIS05, but only include one extra field, the one that has the Law in it. Include only those with "CWA" in the field.

—We want to link FRS and NPDES and aggregate all monies associated with a certain facility. With the data being so dirty, we must first isolate ICIS and PCS from each other and get their totals.

-Two steps: ICIS we select out ici_fcltuin and the five fields we're interested in: enfocnu, enfcfpa, enfcslp, enfctsa and enfccaa, and then sum the federal and state money fields for each FRS number.

-Do the same for PCS: PCSNPDES and apam and appa. (See NOTE below under "To Find These Total Sums")

Joining them to get all the PCS money into ICIS causes a problem, because many FRS numbers are associated with multiple NPDES numbers, and because our key and our other linked tables has multiple entries, we'll get FRS numbers linked to only one NPDES number most of the time, and but sometimes FRS numbers will be linked to two or more NPDES numbers.

Solution:

-In ICISmoney, add NPDES numbers from PCS, NPDES and key.

-Find duplicates, with a count (FRS). — will indicate how many times a record appears.

- SELECT ici_fcltuin, count(ici_fcltuin) AS countFRS
FROM joiningnewicisnewpcs
GROUP BY ici_fcltuin
ORDER BY 1;
Then in your table, make a new column, update the results of this query into that CountFRS column. Or, join countFRS query to ICIS Money, only include count(frs) field

-Take table to Excel.

-Make new column to indicate if the record is unique or the first FRS in a set of duplicates. DUPS must be grouped together for this to work.

=IF(C2=C3,IF(H3>1,IF(C1<>C2,1,H2),H2),H2)

-Where dupindicater is not 1, turn to zeros all FRS money fields with update query.

-Find unmatched ICIS records.

-Append.

OR

-do momoney left join pcsmoney, returns all of momoney plus npdes numbers if they match, blanks if there is no match.

-Add PCS money fields to momoney

-Find unmatched PCS records.

-Match NPDES numbers with FRS numbers by joining 'pcs crit key' to npdes numbers in 'momoney' (Don't forget to Group BY)

-Append.

{{Get all ICIS and PCS records that meet our criteria but do not match key. Put them in table "momoney."

-join icis063 with icis11, narrowing by criteria.
-Join that query to ICIS05, filtering out all that are not CWA
-Match to Key.
-Find those that do not match.
-for this query, run through ICIS sums and ICIS total sums, just like you did earlier with the ones that did match the key. Do a CountFRS and ID here. Put in a new table with structure like momoney, add field to indicate "inkey"
-append those records to "momoney"

-for records with no FRS number, but have a case number:
      -look up on ECHO, find FRS and NPDES.
      -update in icis063, update in "momoney" with field indicating "handmatch"

-Grab the records from pcsmoney that did not match the key.
-Put them in a new table with structure like momoney, add field to indicate "inkey"
-append those records to "momoney"}

Did not need to do this with the updated data.}

So now momoney should include FRS numbers with all NPDES numbers, and individual sums of ICIS totals (showing only in the first instance of a record) and individual PCS totals.

**TO FIND THESE TOTAL SUMS:**
-Add six columns to "momoney" in Excel.
-sort by FRS number then ID, both ascending.

NOTE: We need to check and make sure sums in APPA or APAM are not exactly the same as in enfcfpa. Pull out all records where this is the case and look up in ECHO to see what you should do. The first three new fields will be indicators for the other money fields. If you put an X in the field, the formula will know not to add it.

-In the next new column, sum PCS records for each NPDES number
    =IF(C2=D2,D2,C2+D2)
    -compares to see if they're the same. If they are, we don't want to add them together, if they are different, we do.

    **Second time around, with "X" in field to indicate not to add it:
    =IF(I2=J2,I2,IF(N2="x",J2,IF(O2="x",I2,I2+J2)))

-In the fifth column, we want to add up all the SUMS for each NPDES number, and put them in this column, for only the first instance of a unique FRS number. The other instances should have a zero. We want this to happen so we can just add up across the row to get a total. Our ICIS totals are already in only the first instance of a record.

=IF(L2=1,IF(F2=F3,IF(F2<>F4,M2+M3,M2+M3+M4),M2),IF(L2>1,0,M2))

**Second time, checking four FRS fields for matches instead of just three.
=IF(A2=1,IF(C2=C3,IF(C2<>C4,P2+P3,IF(C2<>C5,P2+P3+P4,P2+P3+P4+P5)),
P2),IF(L2>1,0,P2))

Now all we have to do is add up our two ICIS columns with the last column we just
made, but we need to make sure not to add fields that are marked with an X.
=IF(L2<>"x",IF(M2<>"x",E2+F2+Q2,E2+Q2),IF(M2="x",Q2,F2+Q2))

-Update the address fields off of GEODATA and POTWS. This will take care of most of
the addresses and names. Then update by NPDES number.

-Still may be some not filled in, so update fields based on NPDES number, where the
address isn't already filled in:
UPDATE madmoney, pcs_crit_b4key
SET madmoney.name = pcs_crit_b4key.[nam1+nam2],
madmoney.city = pcs_crit_b4key.cynm,
madmoney.state = pcs_crit_b4key.state
WHERE madmoney.npdes=pcs_crit_b4key.pcs001_npdes And (madmoney.name
Is Null And madmoney.city Is Null And madmoney.state Is Null);

Update Mugglemoney with lat/long on the FRS field from PCS, ICP, EPA_GEODATA
or CWNS.

Also from GEODATA, grab fac_url.

Update Mugglemoney with FIPS codes

Check TotalSums against ECHO for discrepancies.

Clean Watershed Needs Survey:
Where POTWS cannot help us, update Mugglemoney from CWNS to include:

—State and county fips codes (combine later; use these codes to fill in the missing county
fields)
—Authority name
—existing flow and present design max capacity flow
—CSOs
—303a
—watersheds and receiving waters
—state and federal needs (create a field indicating the sum of these two fields)

Cleaning Muggle:

—In the name field, change format of "Charleston, City of" to City of Charleston." Same with towns, and generally go through and clean this field pretty good. Change to all the same case, delete duplicate naming and extra spaces.
—Delete all records without a 1 in the key field once you're confident you won't need those other NPDES records.
—Delete all the fields that aren't necessary any more, which includes all the fields that fed into TotalSum, and delete appa and apam.

Almost last step:
Run every record with Money fields through ECHO to determine if they were part of cases that were against a utility with more than facility. Indicate this in the "util" field and create three fields (bestsep, bestca, bestsum) that will show how much the facility was on the hook for. Do this by in-depth ECHO searches by case number, press releases, and contacting the facility directly.

Last step:
Run each and every record through ECHO to make sure that our records indicate whether or not the facility had an enforcement action in the past five years. Fix those that do not.

**Additional tables**:

HLRNC
Join on NPDES pcs03 with Muggle.frs, pulling out hqrtr and hlrnc, where records have a 2003 or greater date in the HQRTR field and either an 'S' or 'E' or 'X' in the HLRNC field. Make a separate table.

ENACS
Use the NPDES numbers from the final Muggle draft and join with pcs07, filtering by our date criteria on endt. To get the ENAC_Description, you must first join the old version of this table with the new version, because only the old has the description. Make a separate table when done.

ENFCOAC
Join ici06.fcltuin with ici10 to get the FRS numbers in ici10, sorting by our date criteria. Join this query then to muggle on the FRS number.
Make this a separate searchable table.

ENFOSUM
Get all records with a CWA lawsesc from ICI05.
Join this to ici06 to get the frs number.
Then join this to ici08 to filter out by milestone date and type. Exclude ""Enforcement Action Data Entered".
Join this to muggle to sort out only those that have had enforcement actions in the past five years.
Join this to ic04 to get the enforcement summary.
Take this table to Excel, add three fields.

Because we want to find the last line of a multiple-line enforcement summary, we need to find the last line to know when to stop adding lines. The first field will do this with this: =IF(C3=1,1,2). That will put a 2 in every field that is not the last enfoln. However, we don't want to include the last line, because it's only code. We want to pull the record on the second to last line, which will hold the whole enforcement summary excluding the code. So in the second new field, do this: =IF(E2=1,2,IF(E3=1,1,2)).
The third new field contains the text of the enforcement summary. It compiles all the enfosum lines with this : =IF(C2=1,D2,G1&" "&D2)
Keep in mind that it only does this when the table is sorted by enfocnu, then frs, then enfoln all ascending, and we need to add a "1" in the enfoln field following the last record in the table, or the last formula above will not work. (Delete this record when you get the table back into Access.)

A list of some of the queries used for the sewers project.

Sample SQL from the sewers project:

```
SELECT icicrit06_10.[ici06.enfocnu], icicrit06_10.fcltuin, icicrit06_10.enfcsld,
icicrit06_10.enfcsed, icicrit06_10.enfcfpa, icicrit06_10.enfcslp, icicrit06_10.enfctsa,
icicrit06_10.enfccaa, ici05.lawsesc
FROM icicrit06_10, ici05
WHERE (icicrit06_10.[ici06.enfocnu]=ici05.enfocnu) And ici05.lawsesc="CWA"
GROUP BY icicrit06_10.[ici06.enfocnu], icicrit06_10.fcltuin, icicrit06_10.enfcsld,
icicrit06_10.enfcsed, icicrit06_10.enfcfpa, icicrit06_10.enfcslp, icicrit06_10.enfctsa,
icicrit06_10.enfccaa, ici05.lawsesc;


SELECT money.*, potws.[major /minor]
FROM [money], potws
WHERE money.npdes = potws.npdes
ORDER BY money.frs;


SELECT frs, count(frs) AS countfrs
FROM moneymajorpcs
GROUP BY frs
ORDER BY 2 DESC;


SELECT *
FROM icp01, icp02
WHERE icp01.perexno=icp02.perexno And ((icp01.fclstcd<>"GU" And
icp01.fclstcd<>"VI" And icp01.fclstcd<>"PR") Or (icp01.fclstcd Is Null)) And
(icp02.sic="4952" Or icp02.naics="221320") And icp01.permmsf="M";


SELECT frs, [pcs01.npdes], Sum(nz([apam],0)) AS sum_apam, Sum(nz([appa],0)) AS
sum_appa
FROM pcscritkeymonies
GROUP BY frs, [pcs01.npdes]
ORDER BY 1 DESC;


SELECT fcltuin, sum(nz([enfcfpa],0)) AS fed, sum(nz([enfcslp],0)) AS state,
sum(nz([enfctsa],0)) AS SEP, sum(nz([enfccaa],0)) AS CA
FROM icicritlawdated
GROUP BY fcltuin;


SELECT [icicritsums with npdes].*, pcscritkeysums.sum_apam,
pcscritkeysums.sum_appa
FROM [icicritsums with npdes] LEFT JOIN pcscritkeysums ON ([icicritsums with
npdes].npdes=pcscritkeysums.[pcs01.npdes]) AND ([icicritsums with
npdes].fcltuin=pcscritkeysums.frs);
```

SELECT pcscritkeySUMS.frs, pcscritkeySUMS.pcs01.npdes,
pcscritkeySUMS.sum_apam, pcscritkeySUMS.sum_appa
FROM pcscritkeySUMS LEFT JOIN somemoney ON pcscritkeySUMS.pcs01.npdes =
somemoney.npdes
WHERE (((somemoney.npdes) Is Null));

UPDATE moneycountid SET appa = 0
WHERE appa is null;

SELECT badmoney.*, countfrs.countfrs
FROM badmoney, countfrs
WHERE badmoney.frs=countfrs.frs
ORDER BY badmoney.frs;

UPDATE doughycount SET sep = 0, CA = 0, fed = 0, state = 0
WHERE id>1;

SELECT somemoney.fcltuin, somemoney.fed, somemoney.state, somemoney.sep,
somemoney.ca, [missingpcs sums].npdes, [missingpcs sums].sum_apam, [missingpcs
sums].sum_appa
FROM somemoney, [missingpcs sums]
WHERE somemoney.fcltuin=[missingpcs sums].frs
ORDER BY somemoney.fcltuin;

SELECT *
FROM ici06, ici10
WHERE ici06.enfocnu=ici10.enfocnu And (ici06.fclasic Like "*4952*" Or ici06.fanaics
Like "*221320*") And ((ici06.fcltstc<>"PR" And ici06.fcltstc<>"VI" And
ici06.fcltstc<>"GU") or (ici06.fcltstc is null)) And ((ici06.enfocnu Like "???2003?????"
Or ici06.enfocnu Like "???2004?????" Or ici06.enfocnu Like "???2005?????" Or
ici06.enfocnu Like "???2006?????" Or ici06.enfocnu Like "???2007?????" Or
ici06.enfocnu Like "???2008?????") Or (ici10.enfcsld Like "*2003" Or ici10.enfcsld Like
"*2004" Or ici10.enfcsld Like "*2005" Or ici10.enfcsld Like "*2006" Or ici10.enfcsld
Like "*2007" Or ici10.enfcsld Like "*2008" Or ici10.enfcsed Like "*2003" Or
ici10.enfcsed Like "*2004" Or ici10.enfcsed Like "*2005" Or ici10.enfcsed Like "*2006"
Or ici10.enfcsed Like "*2007" Or ici10.enfcsed Like "*2008") Or (ici10.enfcsed Is Null
And ici10.enfcsld Is Null))
ORDER BY ici06.fcltuin;

UPDATE mugglemoney, mugglegeoPCS SET mugglemoney.lat = mugglegeopcs.flat,
mugglemoney.[long] = mugglegeopcs.flon, mugglemoney.county = mugglegeopcs.cnty
WHERE mugglemoney.npdes=mugglegeopcs.npdes;

```
SELECT *
FROM pcsbase, pcs07
WHERE pcsbase.npdes=pcs07.npdes And (pcs07.endt Like "*2003" Or pcs07.endt Like
"*2004" Or pcs07.endt Like "*2005" Or pcs07.endt Like "*2006" Or pcs07.endt Like
"*2007" Or pcs07.endt Like "*2008") And ((pcs01.state Not Like "GU" And pcs01.state
Not Like "VI" And pcs01.state Not Like "PR") Or pcs01.state Is Null)
ORDER BY pcsbase.npdes;

SELECT enf, exceeding, count(exceeding) AS [count]
FROM mugglestats
WHERE exceeding = "y" or exceeding = "e"
GROUP BY enf, exceeding
ORDER BY 3 DESC;

SELECT pcs07.npdes, pcs07.enac, pcs07.eatp, pcs07.endt, pcs07.apam, pcs07.appa,
enackey.enac_desc
FROM pcs07 LEFT JOIN enackey ON pcs07.enac = enackey.enac;

SELECT baddataicibase10.frs, baddataicibase10.enfocnu, baddataicibase10.enfcfpa,
baddataicibase10.enfcslp, baddataicibase10.enfctsa, baddataicibase10.enfccaa,
ici08.subacad, ici08.subacty_text
FROM baddataicibase10, ici08
WHERE baddataicibase10.enfocnu=ici08.enfocnu And (ici08.subacad Like "*2003" Or
ici08.subacad Like "*2004" Or ici08.subacad Like "*2005" Or ici08.subacad Like
"*2006" Or ici08.subacad Like "*2007" Or ici08.subacad Like "*2008") And
(ici08.subacty_text Not Like "Enforcement Action Data Entered")
GROUP BY baddataicibase10.frs, baddataicibase10.enfocnu, baddataicibase10.enfcfpa,
baddataicibase10.enfcslp, baddataicibase10.enfctsa, baddataicibase10.enfccaa,
ici08.subacad, ici08.subacty_text
ORDER BY baddataicibase10.enfcslp;

UPDATE mugglemoney, potws SET mugglemoney.name = potws.name,
mugglemoney.street = potws.street, mugglemoney.city = potws.city,
mugglemoney.statecode = potws.state, mugglemoney.zip = potws.zip,
mugglemoney.csos = potws.cso, mugglemoney.303a = potws.[303],
mugglemoney.shedcode = potws.shedcode, mugglemoney.watershed = potws.watershed,
mugglemoney.waters = potws.water
WHERE mugglemoney.frs=potws.frs;

SELECT getenfosum.[ici06.enfocnu] AS enfocnu, getenfosum.fcltuin AS frs,
ici04.enfosum, ici04.enfoln
FROM getenfosum, ici04
WHERE getenfosum.[ici06.enfocnu]=ici04.enfocnu
GROUP BY getenfosum.[ici06.enfocnu], getenfosum.fcltuin, ici04.enfosum, ici04.enfoln
ORDER BY getenfosum.[ici06.enfocnu];
```
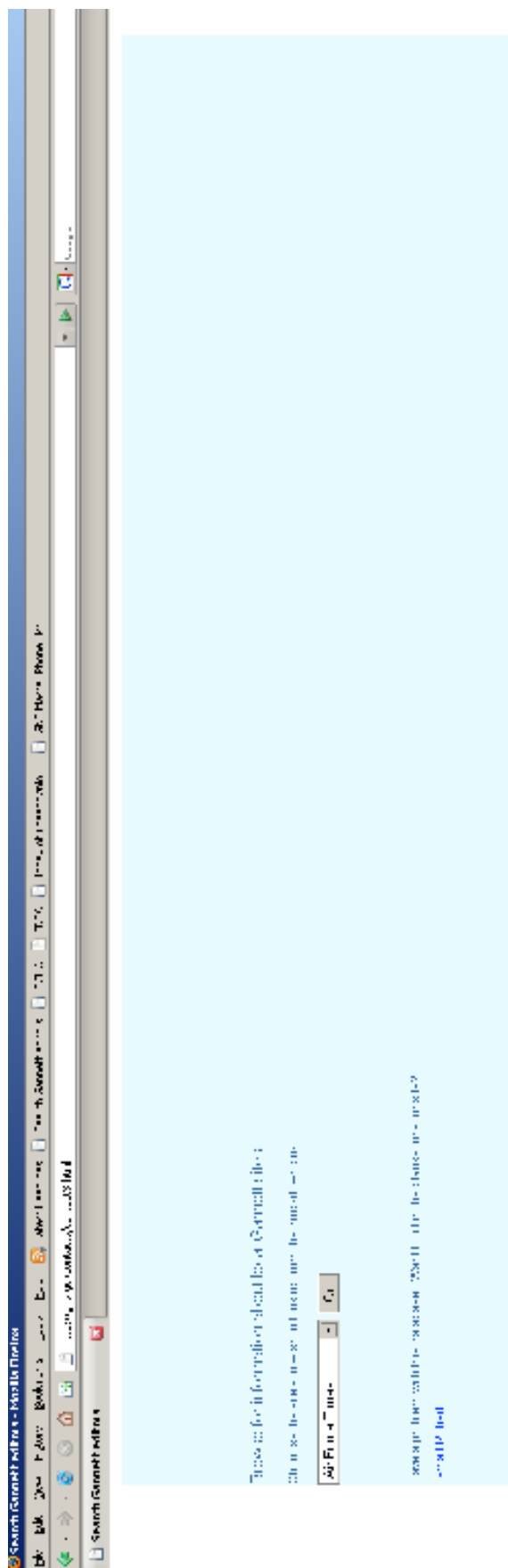
A list of some of the tables used for the sewers project.

| COUNTY_NM | AUTHORITY_NM | F_NEED | |
|---|---|---|---|
| Hale | AKRON, TOWN OF | 218 | 218 |
| Shelby | CITY OF ALABASTER | | |
| Tallapoosa | ALEXANDER, CTY OF, SW DPT | 8399 | 467 |
| Tallapoosa | ALEXANDER CITY SEWER DEPT | | |
| Etowah | ALTOONA, TOWN OF, WW&SB | 2802 | |
| Covington | ANDALUSIA, CITY OF | 6529 | |
| Calhoun | CITY OF ANNISTON WW & SB | 40688 | 1 |
| Calhoun | CITY OF ANNISTON WW & SB | 22746 | |
| Marshall | ARAB SEWER BOARD | 2154 | |
| Marshall | ARAB SEWER BOARD | 2839 | |
| Limestone | CITY OF ARDMORE | | |
| Houston | CITY OF ASHFORD | 4342 | |
| Limestone | CITY OF ATHENS | 262 | |
| Etowah | CITY OF ATTALLA | 11905 | 696 |
| Lee | MERSCOT-AUBURN LIMITED PARTNER | 4701 | |
| Lee | MERSCOT AUBURN LIMITED PARTNER | | |

A sample of two tables in the CWNS database.

Screenshots of the searchable Gannett contacts utility I designed for the company Intranet. Screenshots continue onto the next page.
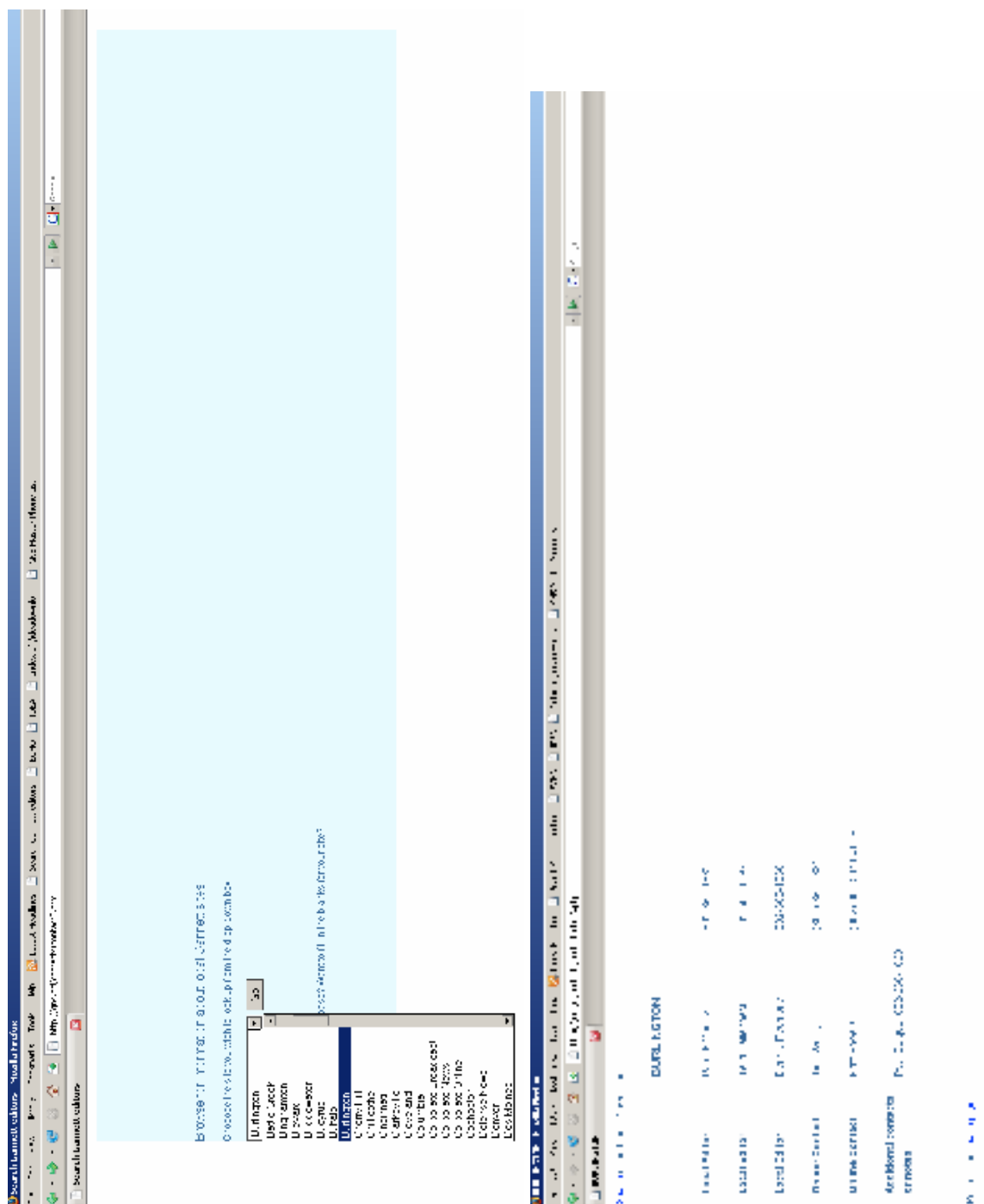
A sample of Katrina contracts.

| | signed | value | compete | why | placeofperform | vendorname | vendorcity | vendorstate |
|---|---|---|---|---|---|---|---|---|
| 1 | 10-Sep-05 | $1,139,695.00 | Not Comp | Urgency | BATON ROUGE | BOLEK INC | SPRINGDALE | AR |
| 2 | 13-Sep-05 | $22,440.00 | Not Comp | Urgency | BATON ROUGE | BEN'S RV CENTER | YORK | PA |
| 3 | 13-Mar-06 | $6,451,500.00 | Not Comp | Urgency | METAIRIE | BENTS RV RENDEZVOUS LIMITED LIABILITY COMPANY | METAIRIE | LA |
| 4 | 7-Dec-05 | $3,177,865.00 | Full and Open Compe | | METAIRIE | BENTS RV RENDEZVOUS LIMITED LIABILITY COMPANY | METAIRIE | LA |
| 5 | 16-Feb-06 | $5,098,500.00 | Not Comp | Urgency | METAIRIE | BENTS RV RENDEZVOUS LIMITED LIABILITY COMPANY | METAIRIE | LA |
| 6 | 19-Nov-05 | $14,108,079.65 | Not Comp | Urgency | PONCHATOULA | BERRYLAND MOTORS LIMITED LIABILITY COMPANY | PONCHATOULA | LA |
| 7 | 12-Dec-05 | $4,999,500.00 | Not Comp | Urgency | PONCHATOULA | BERRYLAND MOTORS LIMITED LIABILITY COMPANY | PONCHATOULA | LA |
| 8 | 10-Sep-05 | $711,159.00 | Full and Open Compe | | NEW ORLEANS | BEST BUY RV"S, INC | RICHMOND | IN |
| 9 | 9-Sep-05 | $224,400.00 | Not Comp | Urgency | BATON ROUGE | BEN'S RV CENTER | YORK | PA |
| 10 | 28-Nov-05 | $432,700.00 | Not Comp | Urgency | BATON ROUGE | BLANCHARD TRAILER SALES | BATON ROUGE | |
| 11 | 11-Sep-05 | $3,021,075.00 | Not Comp | Urgency | BATON ROUGE | BATES SHOW SALES STAFF INCORPORATED | DOVER | FL |
| 12 | 22-Sep-05 | $0.00 | Not Comp | Urgency | BATON ROUGE | BOLEK INC | SPRINGDALE | AR |
| 13 | 17-Sep-05 | $98,100,000.00 | Competed under SAT | | NORCO | BOURGET'S OF THE SOUTH | NORCO | LA |
| 14 | 17-Sep-05 | $4,059,490.00 | Not Comp | Urgency | NORCO | BOURGET'S OF THE SOUTH | NORCO | LA |
| 15 | 8-Sep-05 | $2,402,346.00 | Not Comp | Urgency | METAIRIE | BOURGET'S OF THE SOUTH, LLC | NORCO | |
| 16 | 7-Sep-05 | $851,095.00 | Not Comp | Urgency | METAIRIE | BIG COUNTRY RV CENTER | BEND | |
| 17 | 9-Sep-05 | $1,110,895.00 | Not Comp | Urgency | NEW ORLEANS | ART'S RV | GLENELLYN | |
| 18 | 11-Sep-05 | $448,723.00 | Not Comp | Unique Sc | ELKHORN | ANNIE RAE CHEVROLET GEO INC | DEWITT | MI |
| 19 | 5-Oct-05 | $448,723.00 | Not Comp | Unique Sc | ELKHORN | ANNIE RAE CHEVROLET GEO INC | DEWITT | MI |
| 20 | 8-Sep-05 | $1,936,393.00 | Full and Open Compe | | NEW ORLEANS | ANNIE RAE CHEVROLET GEO INC | DEWITT | MI |
| 21 | 24-Sep-05 | $-21,584.00 | Full and Open Compe | | NEW ORLEANS | ANNIE RAE CHEVROLET GEO INC | DEWITT | MI |
| 22 | 2-Jun-06 | $49,379.57 | Full and Open Compe | | SELMA | ARROWHEAD R V SALES INCORPORATED | MARIANNA | FL |
| 23 | 16-Feb-06 | $2,249,700.00 | Not Comp | Urgency | BOSSIER CITY | BAYOU OUTDOOR SUPERCENTER | BOSSIER CITY | LA |
| 24 | 10-Sep-05 | $1,485,297.00 | Full and Open Compe | | NEW ORLEANS | ARROWHEAD R.V. SALES, INC. | MARIANNA | FL |
| 25 | 11-Sep-05 | $124,698.00 | Full and Open Compe | | NEW ORLEANS | BRADFORD RV CENTERS | BROCKTON | |
| 26 | 16-Feb-06 | $8,440,000.00 | Not Comp | Urgency | CARENCRO | ATCHAFALAYA RV CENTER LIMITED LIABILITY COMPAN | CARENCRO | LA |
| 27 | 22-Sep-05 | $141,885.00 | Not Comp | Urgency | BATON ROUGE | AUTO READY, INC. | UPLAND | CA |
| 28 | 6-Sep-05 | $90,699.00 | Not Comp | Urgency | METAIRIE | BALZANNAS RV | DELTA | |
| 29 | 11-Sep-05 | $617,700.00 | Not Comp | Urgency | BATON ROUGE | BATES RV | DOVER | |
| 30 | 8-Sep-05 | $276,371.00 | Not Comp | Urgency | METAIRIE | CAHABA RV INC. | ALABASTER | |
| 31 | 23-Nov-05 | $1,646,506.00 | Not Comp | Only One S | BOSSIER CITY | BAYOU OUTDOOR SUPERCENTER | BOSSIER CITY | LA |

| datesigned | currentcontractvalue | extentcom | reasonnot | placeofperformance | vendorname | vendorcity | vendorstate |
|---|---|---|---|---|---|---|---|
| 30-Sep-05 | $4,737,500.00 | Full and Open Compe | | NEW ORLEANS | AMERICAN HOMESTAR CORPORATION | LEAGUE CITY | TX |
| 23-Sep-05 | $80,800,000.00 | Not Comp | Public Inte | NEW ORLEANS | CHAMPION HOME BUILDERS COMPANY | AUBURN HILLS | MI |
| 27-Sep-05 | $69,790,000.00 | Not Comp | Public Inte | NEW ORLEANS | CMH MANUFACTURING INCORPORATED | MARYVILLE | TN |
| 23-Sep-05 | $9,582,876.00 | Not Comp | Public Inte | NEW ORLEANS | DYLAN HOMES LIMITED LIABILITY COMPANY | LISBON | OH |
| 26-Sep-05 | $4,270,000.00 | Full and Open Compe | | NEW ORLEANS | FALL CREEK HOMES LIMITED LIABILITY COMPANY | ELKHART | IN |
| 23-Sep-05 | $4,294,440.00 | Not Comp | Public Inte | NEW ORLEANS | FUQUA HOMES INCORPORATED | ARLINGTON | TX |
| 26-Sep-05 | $113,958,000.00 | Not Comp | Public Inte | NEW ORLEANS | MORGAN BUILDING TRANSPORT | GARLAND | TX |
| 23-Sep-05 | $51,090,095.00 | Not Comp | Public Inte | NEW ORLEANS | NORTH AMERICAN CATASTROPHE SERVICES INCORPORATED | MELBOURNE | FL |
| 4-Sep-05 | $3,658,191.00 | Not Comp | Urgency | BATON ROUGE | PALM HARBOR HOMES INCORPORATED (6634) | ADDISON | TX |
| 30-Sep-05 | $4,559,400.00 | Full and Open Compe | | NEW ORLEANS | SILVER CREEK HOMES INCORPORATED | HENRIETTA | TX |
| 16-Sep-05 | $5,269,814.00 | Not Comp | Urgency | BATON ROUGE | SOUTHERN ENERGY HOMES INCORPORATED | ADDISON | AL |
| 29-Sep-05 | $30,917,100.00 | Full and Open Compe | | NEW ORLEANS | SOUTHERN ENERGY HOMES INCORPORATED | ADDISON | AL |

Katrina contracts for trailers and mobile homes.

| STATE | stateab | REST_USE | CountOfGNSID | |
|---|---|---|---|---|
| 01 | AL | 0 | 8 | 7.62% |
| 01 | AL | 5 | 95 | 90.48% |
| 01 | AL | 15 | 2 | 1.90% |
| | | | 105 | 100.00% |
| 02 | AK | 0 | 2 | 22.22% |
| 02 | AK | 5 | 7 | 77.78% |
| | | | 9 | 100.00% |
| 04 | AZ | 0 | 87 | 61.27% |
| 04 | AZ | 5 | 45 | 31.69% |
| 04 | AZ | 15 | 1 | 0.70% |
| 04 | AZ | 99 | 9 | 6.34% |
| | | | 142 | 100.00% |
| 05 | AR | 0 | 54 | 71.05% |
| 05 | AR | 5 | 18 | 23.68% |
| 05 | AR | 15 | 2 | 2.63% |
| 05 | AR | 99 | 2 | 2.63% |
| | | | 76 | 100.00% |
| 06 | CA | 0 | 53 | 10.47% |
| 06 | CA | 5 | 429 | 84.78% |
| 06 | CA | 8 | 1 | 0.20% |
| 06 | CA | 15 | 14 | 2.77% |
| 06 | CA | 99 | 9 | 1.78% |
| | | | 506 | 100.00% |
| 08 | CO | 0 | 60 | 81.08% |
| 08 | CO | 5 | 13 | 17.57% |
| 08 | CO | 15 | 1 | 1.35% |
| | | | 74 | 100.00% |
| 09 | CT | 0 | 34 | 64.15% |
| 09 | CT | 5 | 18 | 33.96% |
| 09 | CT | 99 | 1 | 1.89% |
| | | | 53 | 100.00% |
| 10 | DE | 0 | 7 | 58.33% |
| 10 | DE | 5 | 4 | 33.33% |
| 10 | DE | 15 | 1 | 8.33% |
| | | | 12 | 100.00% |
| 11 | DC | 5 | 1 | 100.00% |
| | | | 1 | 100.00% |
| 12 | FL | 0 | 244 | 43.42% |
| 12 | FL | 5 | 278 | 49.47% |
| 12 | FL | 99 | 40 | 7.12% |
| | | | 562 | 100.00% |
| 13 | GA | 0 | 19 | 12.34% |
| 13 | GA | 5 | 125 | 81.17% |
| 13 | GA | 15 | 2 | 1.30% |
| 13 | GA | 99 | 8 | 5.19% |
| | | | 154 | 100.00% |

A sample of percent calculation for the motorcycles project.

*Professional Analysis*

# Literature Review

Computer-assisted reporting involves using computers to pinpoint and expose problems that were otherwise unclear or hidden. It is used as a tool to further inform computer assisted reporting with databases analysis, geographic information and mathematics. Computer-assisted reporting often relies on government-produced, tax-payer funded data, and has the capability of adding tremendous depth to journalism.

The power of computers first hit the newsrooms in 1952 when an old-school mainframe computer was used in the presidential election.

As Steve Doig, a Knight Chair professor at Arizona State University and a former Miami Herald research editor, says about precision journalism – a term coined by Philip Meyer – computer-assisted reporting turns "anecdotal evident into statistical evidence. Instead of saying, 'We found a home that's all beat up,' and then talking about it, which is the approach you always had to use before, now we can say, 'Here's the horrible case of a drunk driver and there are 172 cases like it.' You couldn't do that kind of thing before." (Garrison, 1997)

Common stories that employ computer-assisted reporting involve analyzing school testing scores, vehicle registration, traffic ticket and crime statistics, campaign contributions and public employee salaries. (Garrison, 1997)

That was 10 years ago, when computer-assisted reporting was just becoming more prevalent in news-making routines.

Reporters have taken federal Home Mortgage Disclosure Act data and mapped it out according to census tract, to see where sub-prime lending foreclosures are happening. Reporters tell local residents which of their bridges are in danger of collapsing. Computer-assisted reporting exposes waste in distribution of Federal Emergency Management Agency funds after hurricanes, and it shows how many sex offenders live within 1,000 feet of schools. And it shows how many felons (who are not permitted to have firearms) have hunting permits.

Computer-assisted reporting every day is becoming even more widespread, and even expected, but there are challenges for investigative reporters using computer-assisted techniques. This type of reporting takes time and resources, both to train journalists and execute projects.

Brant Houston, then-executive director of Investigative Reporters & Editors, states this conflict bluntly, and in no uncertain terms: "When it comes to quality public service journalism or profits at a corporation, profits come first." It's something "some newsroom managers hate to say these days but what everyone in the newsroom knows." (Houston, 2006)

Yet, just a couple months earlier, Houston reported on the most recent round of newsroom budget cuts, stating that 2006 was thus far the "worse than ever" because it was not believed funding would be restored for Meyer's brand of journalism without a new "business model." Yet Houston said journalists seem to still be producing good investigative stories; the same quantity – and of the same quality – of stories were submitted for consideration for an IRE Award in 2006 as in past years. (Houston, 2006)

This perception is backed up by data. In his column for The IRE Journal, Houston reports on a survey conducted at Arizona State University which found that IRE members "think newspapers care about investigative stories but 'frequently don't back that up with resources that reporters say they need to do in-depth work.'" (Houston, 2006)

The survey questioned journalists at the 100 largest newspapers, who all responded that they "are seeing resources diminish for investigative reporting as corporations slash budgets to maintain or increase profits … reporters and editors agree that the desire to do investigative journalism is there, but that *money for staff and training often isn't*." (Houston, 2006)

Houston illuminates two problems here. In addition to newsroom staff cuts, money for training is disappearing, or, in many cases already gone or never existed in the first place. In fact, the 2006 conference of the American Society of Newspaper Editors, at which watchdog journalism was the theme, newsroom budget cuts prevented many editors from attending. (Houston, 2006)

Meyer considers this a big problem. In 2001, Meyer, often considered the father of precision journalism, lamented the lack of widespread use of computer-assisted reporting: "It's become a specialty where one person in the newsroom does all the heavy-duty computing. I think journalism deserves better than that. I think we need to raise the ante on what it means to be a journalist." (Meyer, 2001)

But computer-assisted journalism *is* a specialty due to lack of widespread training.

Jane Singer (2004) discusses a longitudinal study of computer usage that reports that  computers were "entrenched as newsgathering resources" by the late 1990s, but another relevant study addressed the "diffusion of computer-assisted reporting in

newspaper newsrooms (and) found complexity to be a key factor and emphasized the importance of peer communication." (Singer, 2004) Communication in newsrooms is no doubt a factor in the success rate of computer-assisted journalists, but good computer-assisted reporting is complex.

Meyer would take the definition of good investigative journalism and computer-assisted journalism beyond just illuminating trends or patterns, and "looking at structural problems in society that public policy isn't dealing with effectively." (Meyer, 2001)

Clearly these are important stories. But do newsroom cuts threaten these types of stories, and if so, are they worth the cost?

They do a number of things with which our society cannot (or should not wish to) go without. Serving as an independent monitor of power is one of the nine elements of journalism, as defined in Bill Kovach and Tom Rosenstiel's "Elements of Journalism." Investigative journalism, and computer-assisted reporting by extension, does this very thing by exposing phenomenon in society that weren't readily understood or visible. Indeed, "the primary purpose of journalism is to provide citizens with the information they need to be free and self-governing." (Kovach and Rosenstiel, 2001)

Computer-assisted reporting gives the investigative reporter the tool to give the citizens this information. Additionally, "investigative journalists … can issue a compelling call for public moral indignation. Their particular sort of reporting yields stories that are carefully verified and skillfully narrated accounts of specific injury and injustice but stories with a meaning that always transcends the facts of the particular case. … In this way investigative journalists are custodians of public conscience." (Ettema and Glasser, 1998).

But economic interests are constraints on journalists' ability to do computer-assisted reporting, investigative journalism, or quality journalism of any sort. An assignment editor at KGO-TV in San Francisco had this to say after its budget was cut by 20 percent:

I have to laugh when I hear executives say cutbacks haven't affected quality. A producer doesn't have the time he did in the past to carefully consider a story. There's not sufficient planning – not sufficient time to do stories. Too often people aren't getting that that when they say "Hey, I have a great story. I need two days to report it and two days to shoot it." (Shoemaker and Reese, 1996)

These two authors conclude that in then-recent years the view that quality journalism should come before profit is waning in a large way, as Houston illustrates continues today. "We should ask to what extent these economic 'constraints,' as they become more severe, affect content." (Shoemaker and Reese, 1996)

Newspapers and other media basically rely on advertising; subscription fees make up a minimal, but not insignificant, portion of revenues.

"In the rush of daily journalism, most stories cannot be weighed on the basis of their economic payoff. Many are clearly evaluated for their audience appeal, which translates into higher circulation and ratings, producing greater advertising revenue." (Shoemaker and Reese, 1996)

Yet stories are chosen on their economic cost. While in-depth reporting is expensive, much news which appeals to audiences isn't. Take sex, violence and celebrities, for example.

Given that computer-assisted reporting is such a vital tool for investigative reporting and journalism, how to successfully employ these techniques in a environment of economic conservatism in newsrooms is worth further examination. That is the purpose of this professional analysis.

## *Article*

### Finding time: Persuading editors to fund a project

By Grant Smith

Disparities in capital punishment sentencing. Fraudulent insurance agents. Institutionalized cheating on standardized tests for school children. Discriminatory medical care.

These are all corrupt practices or injustices uncovered by enterprising journalists using time-consuming computer-assisted reporting techniques.

And it's these techniques that some fear may be affected by dropping circulation numbers and plummeting advertising revenue.

Gannett News Service won second place in the Philip Meyer Journalism Award for computer-assisted reporting in 2006 for its project that rated more than 3,000 hospitals on the care they give their patients.

The Dallas Morning News won first place in the Philip Meyer award last year for a three-day series that exposed cheating by more than 50,000 students on standardized tests in Texas schools.

The Atlanta Journal-Constitution revealed how the state of Georgia failed to deliver death sentences equitably, and won a second place Philip Meyer award last year.

Despite their accolades, these three news outlets all experienced dramatic drops in stock prices or announced staff cuts in the past few months.

"I have a big concern this is going to be something that gets cut," says Kansas City Star reporter David Klepper. "If you gotta squeeze the budget and let people ego, this might be an area, that you know, bean counters might look at when they're trying to cut budgets and I would consider that a real mistake for journalism and the role we play."

All of these stories took a lot of time, and time equals money. Klepper's concern that computer-assisted reporting and investigative reporting could get the corporate axe is not entirely unfounded, as new organizations struggle to adapt to the Internet while their virtual oligarchy on advertising slips a little further each day.

It's also possible that data-heavy reporting could serve both newspapers' business interests and the public need for watchdog reporting.

Robert Benincasa, database editor for Gannett News Service, says that newspapers are generally appreciative of data analysis because of the traffic it drives to their Web sites. Anecdotally, he says he doesn't believe computer-assisted reporters and Web gurus are

being cut by upper management because searchable databases seem popular among readers.

Most Gannett newspapers have a Data Center, with their own locally produced databases and links to databases created by Gannett News Service that can be searched locally.

The Asbury Park Press, a Gannett-owned newspaper in New Jersey, launched its own database Web site in late 2006 and saw its Web traffic skyrocket a half a million page views in under two weeks, according to its executive editor, Skip Hidlay. That's gold to advertisers.

These award-winning, data-heavy investigations offer myriad lessons for reporters wishing to follow in their footsteps, especially in light of industry developments.

Klepper, the Kansas City Star's Topeka correspondent and a state government reporter, worked with reporters Mike Casey and Mark Morris for six months analyzing insurance data and reporting on the widespread insurance fraud they uncovered.

"I began hunting around a bit and found that there's perhaps no greater betrayal of trust that will touch an individual reader than perhaps an insurance agent taking your money from premiums and not giving you coverage," Morris says.

They also learned that some insurance regulators let companies get away with that fraud, and patients had virtually no recourse.

"You're telling people about important and significant things that affect them in very real ways – how government regulations and profit motive really hurt consumers," Klepper says. "The data really was just a way of getting at that."

With the vastness of their data, there'd have been no practical way to do the project if the reporters were still writing daily stories, Klepper says.

Similarly, at the Dallas Morning News, Josh Benton and Holly Hacker spent months analyzing scores on standardized tests in more than 50,000 Texas schools to look for evidence of cheating.

But Benton started by writing a few stories about one specific district.

"It was a really, really bad, corrupt, awful school district in the Dallas suburbs," Benton says. "It was just the worst school district in Texas."

Hacker joined Benton's quest, and they discovered widespread cheating on standardized tests, undoubtedly made possible by assistance from educators.

Following those first few stories, the Texas Board of Education shut down the school district.

"The first round of cheating stories really made an impact, so when it came to the second round, (editors) were pretty supportive," Hacker says. "They were willing to put the money out for data … and things like that.

All of these award winners work at news organizations with a deep commitment to investigative projects using computer-assisted reporting.

"Most newspapers have never thrown all the resources at this that they probably should, which is why you see the same newspapers winning the Pulitzers every year," Morris says. "Newspapers that want to do it will continue to do it. They may not cut people loose for three years to work on projects like they used to … but they'll find ways to do it."

Because of that commitment, Morris says, the reporting trio had to do little to convince its editors to cut them loose to work on the project.

"I think it works well at large papers," Benincasa says. "Once you get smaller than big metro papers it gets more difficult."

Benincasa and reporter Jennifer Brooks spent three months analyzing hospital data from the Centers for Medicare and Medicaid Services to look for patterns in the quality of care available in different demographics and under different ownership structures.

They found that poor areas were underrepresented in rankings that indicate whether or not hospitals followed recommended treatment guidelines. Patients in more affluent areas were more likely to receive better care, and some regions of the country ranked better than others.

Gannett News Service conducts three to four major investigations each year, each taking two to three months, in addition to 10 to 12 less ambitions projects each year, Benincasa said.

But Gannett News Service organizational structure lends itself to in-depth investigations and long-term computer-assisted reporting. It has a six-member projects team that is rarely responsible for turning over daily copy. Regional reporters do the vast majority of the daily reporting, freeing up the projects team to dive into data projects. Those projects can be picked up by any of Gannett's 85 daily newspapers, and news stations add the GNS-produced content to their on Web presence.

But before joining GNS, Benincasa worked for the Burlington Free Press as a health care beat reporter. He produced 30 projects in 36 months, in addition to two or three daily stories a week.

"It was very difficult – I put in extra time, basically," Benincasa says. "You couldn't work 40 hours a week and do that."

Bill Rankin ended up bringing work home every day and every weekend for the Atlanta Journal-Constitution's investigation of inequity of capital punishment cases in Georgia.

"I've never worked harder on anything," Rankin says. Rankin, Heather Vogell, Sonji Jacobs and Megan Clarke discovered that Georgia had not made promised reforms to its capital punishment sentencing practices, and that being sentenced to death still had quite a bit to do with demographics. The project took nearly two and half years.

Rankin says he thinks a lot of reporters stumble over getting approval for their projects.

 "I think a lot of projects don't get started because they think they would never get approved, but you never get to do it if you don't try," Rankin says. "You just got to study up on it and write the best proposal you can and make your best pitch."

Benincasa recalls a strategy he would have employed with his editors in Burlington, had he not already secured a green light for a project on state salaries. He found that the highest-paid state employee was a guy whose job it was to relight gas furnaces for residents. He was on call 24 hours a day, made money while he slept and consequently, made more money than the governor.

"Give them something," Benincasa says. "Do part of the story on your own without asking permission and give them something juicy and say, 'I can write a story on this.' Do part of the analysis already."

Most journalists will not get the opportunity to work for a metropolitan newspaper or a news service with resources enough to release reporters on crusades for months at a time. But the experiences of these award winners are the Rosetta Stone for the rest of us at smaller news organizations.

Even at Gannett News Service, where large-scale investigations are the lifeblood of its projects team, reporters still must propose projects and seek approval.

At smaller organizations, it's just a matter of thoroughly researching your subject and crafting a proposal to which no editor could say no.

And doing your homework means not just researching your subject, but spending time looking at the investigative and computer-assisted reporting work journalists have done in other regions.

"There are a number of papers that have followed in our steps," Benton says. "If they were starting from scratch they probably wouldn't have been able to do it, but they can say 'This is what I want to do and the Dallas Morning News already did this and they won all these awards. Isn't this great?'"

## *Conclusion*

It may indeed be as simple as these award winners make it out to be — that obtaining permission to produce an in-depth, data-heavy report merely takes an incredible pitch to editors and the willingness to work longer hours in the name of solid watchdog reporting. These kinds of projects certainly lend themselves to tremendous interactivity on the Web, a venue in which most newspapers are struggling; even as online advertising revenue grows few — if any — newspapers' Web editions could exist as solvent news organizations without subsidizations from their print products (whose advertising revenue is certainly plummeting).

And it very well may be that searchable databases coupled with in-depth investigative reports online could be the magic potion that newspapers seem to be struggling to find. This would mean that while newspapers continue to slash budgets in order to show fewer losses, the old axiom that you have to spend money to make money is not really too far off the mark.

More research is certainly needed. Data on cuts in newsroom positions would need to be analyzed across the board. Are data and Web production positions being cut? My inclination, and the inclination of Gannett News Service's database editor Robert Benincasa, is that they are not, at least, not at the rate of the rest of the positions in newsrooms.

Additionally, my analysis includes sources that are award winners from news organizations with much larger budgets than the vast majority of newspapers in the United States. What are smaller newspapers doing with data? Is the process the same at small newspapers, or are reporters there required to juggle much more, in terms of daily

stories and long-term projects? It would certainly be important to look into the good work

small papers are doing to understand their newsroom routines and obligations to

management regarding turnaround on stories and investigations.

It would also be interesting to analyze the Web traffic such reports generate

versus other types of coverage. This research could take into account budget allocated to

each type of reporting and the advertising revenue generated by these different genres, as

a way to use correlation for its predictive powers. Since, in the grand scheme of things,

interactive databases on news sites is a relatively new phenomenon, a longitudinal study

of one or more newspapers' usage of such projects and their Web traffic would certainly

be interesting. Perhaps as more databases are offered over time, readers would become

more accustomed to them and usage would grow. A study like this would necessitate an

understanding of each newspaper's commitment to marketing these database offerings as

well, since they can remain in place on the Web for as long as the data is relevant, unlike

your average daily story.

So many questions, so little time.

## Citations

Ettema, James S. and Theodore L. Glasser "Custodians of Conscience," Columbia University Press 1998

Garrison, Bruce "Computer-assisted reporting," Editor & Publisher; 06/21/97, Vol. 130 Issue 25, pp40-43

Garrison, Bruce "Newsgathering tools of the 1990s," Editor & Publisher; 6/24/95, Vol. 128 Issue 25, pp. 16-17

Houston, Brant "Service journalism vs. corporate profits," The IRE Journal Vol. 29, No. 4, p. 4, 2006

Houston, Brant "Cuts in resources don't deter members from doing good stories," The IRE Journal Vol. 29 No. 2, p.4, 2006

Johnson, Michelle "Keeping the newsroom current," Quill; May 2001, Vol. 89 Issue 4, p. 25-28

Kovach, Bill and Tom Rosenstiel "Elements of Journalism," Three Rivers Press New York, 2001

Le, Nghiem Thi Xuan, "Best Practices of Successful Computer Assisted Journalists," Master's Project, University of Missouri School of Journalism 2007

Lindlof, Thomas R. "Qualitative Communication Research Methods," Sage Publications 1995

Meyer, Phillip "Q&A: Reporting with Computers: Some Doubts from a Founder," Columbia Journalism Review, May/June 2001

Shoemaker, Pamela J. and Stephen D. Reese "Mediating the Message: Theories of Influences on Mass Media Content," Longman Publishers USA 1996

Singer, Jane "Strange Bedfellows? The diffusion of convergence in four news organizations," Journalism Studies, 2004, Vol. 5, No. 1, pp. 3-18

Grant Smith
5021 S. Providence Road
Apt. C
Columbia, MO 65203

Megan Means
IRE Journal
138 Neff Annex
Missouri School of Journalism
Columbia, MO 65211

Dear Ms. Means,

As the newspaper industry struggles to recoup lost advertising revenue, adapt to the Internet and respond to Wall Street's demand for increasing profit margins each quarter, some journalists fear computer-assisted and in-depth reporter — the kind of journalism that fulfills our roles as watchdogs — may feel the heat.

As newspapers were slow to adapt to the Web, they have been left behind in online advertising and have been struggling to catch up for some time. While newspapers try new approaches to offering online content, business managers and online programmers work to develop some kind of reliable metric for Web traffic.

At the same time, the Web offers tremendous opportunities for interactive content and data produced by computer-assisted reporters. These products are not inexpensive to develop, but have been shown to drive Web traffic upwards.

Could investigative and computer-assisted reporting be the golden goose that saves journalism?

In conjunction with my master's project at the Missouri School of Journalism I have written an approximately 1,500-word article that explores these issues. My sources are winners of the Philip Meyer Journalism Award from the past two years, and the article draws upon my experience as a computer-assisted reporter for Gannett News Service during the winter 2008 semester as well.

They offer helpful lessons regarding their experiences that could be translated into success across the board, from the giant news service to the community newspaper.

Thank you for your consideration.

Sincerely,


Grant Smith
304-319-0083

## *Appendix*

## I. Project Proposal

**Professional Skills Component**

I will work in Washington, D.C., for Gannett News Service as my professional skills component. I will be a member of a six-person team engaged in computer-assisted reporting work.

My duties will include managing and analyzing news databases, and researching and reporting stories in addition to other newsroom tasks. My immediate supervisor will be Robert Benincasa, database editor for Gannett News Service, and I will work with the enterprise/database team, led by Val Ellicot.

The work I do will ultimately be disseminated to the numerous newspapers Gannett owns to be either used as is or localized for their respective audiences.

In my conversations with Mr. Benincasa and Mr. Ellicot, they explained the work Gannett also does transferring databases to the Web, for ease of use by the public. This will be something new and exciting for me, as I have no experience in this area.

In light of the fact Gannett News Service has never hosted a graduate student from the Missouri School of Journalism during his professional project, I have explained the goals of this project to Mr. Benincasa and Mr. Ellicot, and both are aware of the focus of my professional analysis. I am tremendously excited to be the lead the charge for future Missouri journalism school graduate students, so to speak.

I will work Monday through Thursday January 28 through April 24, 2008, and attend Washington Program courses and seminars and work on my professional analysis on Fridays. As my professional component is only 13 weeks, and the minimum work

requirement is 30 hours a week for 14 weeks (420 hours) I will work at least an average of 33.3 hours a week for those 13 weeks to meet this requirement – though I doubt I'll have trouble finding work enough to keep me busy.

**Professional Analysis Component**

As I will be in Washington next semester working on a project centered on computer-assisted reporting, my professional analysis questions ask how computer-assisted reporting fits into reporters' news-making routines, how best practices in computer-assisted reporting jibe with management priorities, and how computer-assisted reporting fits into the broader scheme of watchdog roles for journalists.

More specifically, how do computer-assisted reporters juggle their responsibilities to produce in-depth investigative reporting with their responsibilities to management? These responsibilities to management may include, but are not limited to, consistent production of computer-assisted journalism that meets budgetary expectations and expectations of a quick turnaround.

My method for compiling information for analysis will be first-hand observation and interviews. My first-hand observation will obviously come from my experience working in Washington at Gannett News Service next semester. Gannett News Service is an excellent choice because it won second place for IRE's Phillip Meyer Award last year.

I will survey a group of journalists who have experience in computer-assisted reporting. I have not yet selected my sample, but I will choose journalists who have been successful, and judged by their peers as so, at answering my over-arching question to achieve their career goals. My sample will not be representative of the industry as a

whole. As I am not trying to survey the industry on computer-assisted reporting practices,

I will choose these successful journalists because I can then take the lessons they've

learned and share them with the broader audience of journalists and newsroom managers.

I will choose about 15 journalists who have won IRE Awards for computer-assisted

reporting in the past five years. I hope to ultimately end up with six to eight journalists

willing to participate. This methodology is based on the idea that if you want to learn best

practices on a given topic, talk to those who are recognized as being the best.


*Theoretical Framework*

Of course, this analysis is also grounded in theory and – quite a few, actually, but I've

chosen to focus on just two or three.

The first theory I will lean on is one that aims to explain journalists' watchdog

role as a morally engaged voice (Ettema and Glasser, 1998). The two authors explain this

important role and the function it plays in democracy in such a way it's best to quote it

directly:


> Their stories call attention to the breakdown of social systems and the disorder
> within public institutions that cause injury and injustice; in turn, their stories
> implicitly demand the response of public officials – and the public itself – to that
> breakdown and disorder. Thus the work of these reporters calls us, as a society,
> to decide what is, and what is not, an outrage to our sense of moral order and to
> consider our expectations for our officials, our institutions, and ultimately
> ourselves. In this way investigative journalists are custodians of public
> conscience. (Ettema and Glasser, 1998)

Clearly, then, investigative journalism is a priority, and by extension so is the investigative tool of computer-assisted reporting.

Economic viability is also a priority, more immediately so perhaps for managers, but they're the ones who cut the checks, so they must be satisfied.

These two competing priorities can be explained by the public interest model and the market model – the two models on either end of the spectrum.

The public interest model seems to suggest that "so long as revenues are sufficient to ensure organizational survival, professional and social objectives take precedence over profits." (Shoemaker and Reese, 1996)

But the public interest model seems to be a rare dinosaur, and research suggests the profit-driven model is the driving force behind journalism today.

In fact, "from all accounts the profit motive has become more important … rendering economic constraints into dictates and weakening the insulation of the news department from the larger firm." (Shoemaker and Reese, 1996)

In addition, diffusion of innovations theory suggests "four key elements of social change." (Singer, 2004) As computer-assisted reporting has only really become commonplace since the 1990s, this is especially relevant when coming to terms with expensive reporting tactics in a profit-driven world. These elements are the idea itself; the channel through which the idea is communicated; time it takes to become aware of the idea and then adopt the idea; and the social system in which the idea exists (Singer, 2004).

The social system is especially relevant here, as it is represented by the newsroom (or the entire organization) for the purposes of this analysis. How much time and

resources to commit to an innovation – such as computer-assisted reporting – "can be

made by an individual acting independently, by a collective of individuals seeking

consensus or by an authority figure mandating adoption within the system as a whole."

(Singer, 2004) I'm interested in the relationship between the journalists employing the

idea and the support they receive from their employers who must pay for it.

Singer also discusses a longitudinal study of computer usage that explains how

computers were "entrenched as newsgathering resources" by the late 1990s. She

mentions another relevant study that addressed the "diffusion of computer-assisted

reporting in newspaper newsrooms (and) found complexity to be a key factor and

emphasized the importance of peer communication." (Singer, 2004) Communication in

newsrooms will no doubt be an important factor in the success rate of computer-assisted

journalists.

These theories all tie together when trying to answer that one basic question about

how successful computer-assisted journalists maximize quality journalism production

while satisfying their employers' priorities for the bottom line.

Much investigative journalism falls within this paradigm of cost-benefit analysis.

This project will bridge the gap between these two seemingly disparate priorities.


*Methodology*

Interviewing is a tried and true method for determining best practices. For example, in

answering the question "What are the best CAR practices needed for journalists to be

successful in their careers," Le Nghiem Thi Xuan conducted in-depth interviews with

"young journalists who are considered successful in their jobs." This master's student

ended up choosing students who had won IRE awards (Le, 2007). While our analysis

focuses are different, (CAR skills vs. competing priorities) the methodology is similar.

Thomas Lindlof outlines the seven basic objectives of interviewing:

Learning about things that cannot be observed directly by other means;
understanding a social actor's perspective; inferring the communicative
properties and processes of interpersonal relationships; verifying, validating, or
commenting on data obtained from other sources; testing hypotheses the
researcher has developed; eliciting the distinctive language – vocabularies,
idioms, jargon, forms of speech – used by social actors in their natural settings;
and achieving efficiency in collecting data. (Lindlof, 2005)

There are various things to keep in mind during in-depth interviews.

Ethnographic interviews are "the most informal, conversational, and spontaneous

form of interview." (Lindlof, 2005) These are often related to the experience of

the interviewee, in which new things constantly come to light for the interviewer.

Consequently, the interviewer "must be able to identify quickly something of

interest in what is said or done and develop a line of questioning on the spot."

(Lindlof, 2005)

In a respondent interviewer, the researcher asks a series of similar

questions to different subjects. The questions are designed to "elicit open-ended

responses" – the benefit being that "by asking the same questions of all

respondents in roughly the same order, the research minimizes interviewer effects

and achieves greater efficiency of information gathering." (Lindlof, 2005)

Narrative interviews are yet another information-gathering device. These rely on stories, told by the inverviewee. Stories are valuable, "more than a means of amusement, stories encode the information needed y members of society to carry on the most critical activities of social intercourse, economy, politics, art, spirituality, birth, and death." (Lindlof, 2005) While this interview technique "usually depends on a long-term, trusting relationship between researcher and social actor," these types of interviews are conducted after the interviewer has done all he can to "establish the most comfortable conditions for letting a person talk." (Lindlof, 2005)