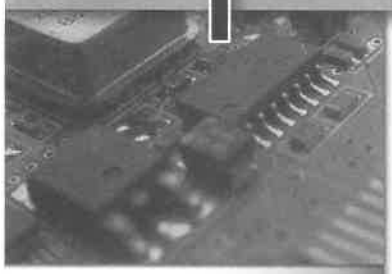
Uplink

March/April 2007
Vol. 21, No. 2

LAST PRINT EDITION! NEW UPLINK INTERACTIVE AND IN FULL COLOR!
Details on p. 18

published bimonthly by
The National Institute for Computer-Assisted Reporting



DOWN ON THE FARM

SUBSIDIES SPREAD CASH FAR FROM STRUGGLING FARMS

By Sarah Cohen, *The Washington Post*

Skeptical of the claims of politicians who say farm subsidies are needed to save the small family farm, Dan Morgan, Gilbert M. Gaul and I spent 2006 answering a simple question about federal subsidy programs: Who reaped the benefits from the taxpayer dollars, and did it help the struggling small family farmers that Congress repeatedly promises that subsidies will save?

Time after time, surprising answers came back. The subsidies included $1.3 billion meant for working farmers but paid to those who grow nothing; $9 billion in disaster aid that often went to the same farms that were paid by the taxpayer-funded $20-billion crop insurance program; $635 million for a severe drought that went to ranches and dairy farms that suffered no serious drought and had no losses; and $817 million in payments to farms that had already reached their legal payment limit.

And all of it was paid without breaking any rules or laws.

Unlocking USDA data

Many of the nine stories that we published as part of the "Harvesting Cash" series, which ran in waves from July through December, depended on the farm subsidy database and other publicly available data from both the USDA and other agencies.

We were lucky that the *Post* had

continued on page 10

In the spotlight:

Tracking cotton subsidies 3

Farmsubsidy.org opens access . . . 5

Reporting on migrant farm labor 7

Data and resources from IRE and NICAR 8

For more about farms investigations see:

- **Mar./April issue of *The IRE Journal***

FIRST VENTURE

FLUFF RULES IN PUERTO RICAN LEGISLATURE

By Omaya Sosa Pascual
El Nuevo Día

Politics in Puerto Rico are complicated and intense.

The current legislature also could be described as frivolous and wasteful, according to a recent computer-assisted investigation by *El Nuevo Día*. The San Juan newspaper used electronic legislative filings to show that Puerto Rican legislators choose quantity over quality when proposing new measures. They devote their energy to saluting special occasions or launching trivial investigations, while neglecting serious economic and social problems.

First, I'll offer a brief explanation for those of you who do not know the political status of our island and how it works. After 400 years of Spanish rule, Puerto Rico became a possession of the United States in 1898 as a result of the Spanish-American War. Through the years, the political relationship between the continent and the island evolved until, in 1952, Puerto Rico became a U.S. commonwealth. In practical terms, we share currency, citizenship and military with the United States, but the island residents manage their own local affairs. We do not vote for the president and do not pay federal taxes. As a result, politics develop in a totally different scenario on the island.

The dominant political parties are local. They're built on the three

continued on page 12

HIDDEN HISTORY

CENSUS SHOWS BLACKS DRIVEN FROM THEIR HOMES

By Elliot Jaspin, Cox News Service

Old census records – I mean really old census records – are a pretty unlikely place to look for a news story. Even current census numbers aren't that exciting and, unlike a fine cognac, sitting around for 100 years doesn't help.

But as I began exploring a database of population figures that stretched from 1890 to 1930, I stumbled across something perplexing. In county after county throughout the United States, I found that the black population would suddenly and mysteriously disappear.

In places like Forsyth County in Georgia and Missouri's Lawrence County, one census would show blacks living in the county, and then a decade later, they were gone. Why had whole communities that had existed since slavery suddenly left? And, once gone, why had blacks never moved back?

At the time, I didn't know that I was looking at the unique signature left by one of the uglier events in American history: racial cleansings.

Like most people, I had never heard of America's racial cleansings. In 1998, I found myself on assignment in Berryville, Ark. With a little time to kill, I wandered into a small history museum and marveled at a display of antique embalming equipment in one room and mounds of bric-a-brac on a table in another. Then I

continued on page 13

INSIDE NICAR

MEYER WINNERS EXAMINE STOCKS, HOSPITALS,SCHOOLS

By David Herzog, NICAR and Missouri School of Journalism

Three newspapers recently won the 2006 Philip Meyer Award for their use of social science methods in reporting on stock options, hospital care and public schools.

The Wall Street Journal took first place for "Perfect Payday," a series of stories that uncovered the practice of secretly backdating stock options to benefit corporate insiders. Reporters Charles Forelle and James Bandler used a statistical model to show how the backdating timing probably didn't happen by mere chance.

Gannett News Service won second place for "Special Report: Rating Hospital Health Care." Database editor Robert Benincasa and reporter Jennifer Brooks analyzed national treatment data to show how patients in poor and rural areas weren't getting recommended care.

The Philadelphia Inquirer received third place for its investigation of Camden, N.J., schools. Melanie Burney, Frank Kummer and Dwight Ott found some test scores at some of schools in the troubled district were higher than would be expected. The revelations led to the resignation of the superintendent and close oversight from the state.

Congratulations to the winners. It's great to see journalists using the tools of social science to report on matters that are important to us all. More information about the winners and the Meyer Award is available online at *www.ire.org/meyeraward/06winners.html.*

Contact David Herzog by e-mail at *dherzog@nicar.org.*

ABOUT OUR CONTRIBUTORS ...

Brigette Alfter is the Brussels correspondent for the Danish newspaper *Information*, and member of Farmsubsidy.org. She teaches and researches European Union FOI issues and is currently developing a FOI Web site for journalists..

Sarah Cohen is a database editor specializing in local and national investigative projects at *The Washington Post.* Cohen is a former IRE Training Director and author of "Numbers in the Newsroom."

Megan Clarke joined *The Atlanta Journal-Constitution* in 2005 as the computer-assisted reporting specialist. She worked as a data analyst at IRE and NICAR and received a master's degree from the Missouri School of Journalism.

Mark Greenblatt is an investigative reporter at KHOU-Houston. His journalism honors include the IRE Award, four Edward R. Murrow Awards, including the national Murrow for the best investigative report in 2006. He is a Mapping Boot Camp alumnus and a graduate of the Missouri School of Journalism.

Elliot Jaspin is a reporter for Cox Newspapers. In 1989, he founded what is now the National Institute for Computer-Assisted Reporting, a joint program of Investigative Reporters and Editors and the Missouri School of Journalism.

Tommy Kaas is partner in Kaas & Mulvad, a company founded in 2007 with Nils Mulvad. Previously, he worked as an editor and co-founder of DICAR, the Danish International Center for Analytical Reporting, and as assistant managing editor at the Danish news agency 4S.

Nils Mulvad is a partner in Kaas & Mulvad, a company founded in 2007 with Tommy Kaas. He is a co-founder of Farmsubsidy.org and the Global Investigative Journalism Network and a former executive director of the Danish International Center for Analytical Reporting.

Omaya Sosa Pascual is an investigative and special projects at *El Nuevo Día* in San Juan, Puerto Rico. Her work has been recognized by Puerto Rico's press associations and the National Association of Hispanic Journalists. She's an alumnae of the May 2006 CAR Boot Camp.

Neil Reisner is a professor of journalism at Florida International University. He worked as a reporter and/or editor at *The Miami Herald, The* (Hackensack, N.J.) *Record* and the *Miami Daily Business Review* and served as training director for IRE and NICAR.

Bits and Bytes

Campus Crime Update

How safe are the campuses of your local colleges or universities? The IRE and NICAR Database Library recently updated the Campus Crime database so you can find out.

On October 1st of each year, schools are required to report crime statistics for the three most recent calendar years. The data cover 2001 through 2005 and contain 8,771 records for the most recent year. On campus and off-campus crime, arrests, discipline and hate crimes are reported in separate tables that can be linked. More limited information is available for 2000. For more information and sample data, visit the Database Library's Web site *www.ire.org/datalibrary/databases/viewdatabase.php?dbaseindex=54.*

RE spring training

variety of IRE's upcoming on-the-road aining events will feature CAR sessions:

he *Watchdog Journalism series*, (*www.e.org/training/watchdogjournalism.html*) esigned for reporters and editors at nall or mid-sized papers or bureaus of irger papers, will feature CAR training at everal events. You must register for the Vatchdog Workshop to attend the CAR essions; space is limited.

Better Watchdog Workshop, April 14-15, in Syracuse, N.Y. (*www.ire.org/training/betterwatchdog/Syracuse07.html*)Beginning Excel.

Better Watchdog Workshop, April 28-29, in Oklahoma City. (*www.ire.org/training/betterwatchdog/Oklahoma07.html*) Beginning Excel.

IRE will conduct a *Money in Politics workshop* (*www.knightnewmediacenter.org/seminars/archives/election_08_covering_politics_in_cyberspace*) April 20-21 in Los Angeles in conjunction with the Knight New Media Center seminar on politics in cyberspace.

CAR Boot Camp, May 13-18, in Columbia, Mo. (*www.ire.org/training/bootcamps/may2007.html*) offers intensive training in spreadsheets, databases and how to use CAR in your news organization.

DOWN ON THE FARM

TRACKING COTTON SUBSIDIES AND LEARNING SQL SERVER ON THE FLY

By Megan Clarke, *The Atlanta Journal-Constitution*

Cotton is still king in Georgia. In 2005, the federal government pumped $317 million in subsidies into the state for cotton alone, making it Georgia's top subsidized crop.

Georgia's take was a piece of $23 billion in subsidies paid nationwide that year. American farm subsidies are criticized around the world for artificially depressing crop prices, and *The Atlanta Journal-Constitution* decided to see how those payments help or harm farmers in the United States.

To tell the story, we chose cotton subsidies because of the amount of money involved and the crop's historical importance in Georgia. My collaborators were reporters Ken Foskett and Dan Chapman, and projects editor Jim Walls.

The U.S. Department of Agriculture's Farm Service Agency is responsible for distributing subsidy money. I requested payment data for the entire country dating back to 1994. FSA initially assured me the data included all subsidy payments. But after some research and checking with a fellow CAR person, I discovered I needed databases for three more programs with nearly $4 billion a year in subsidies: tobacco transition payments, commodity certificate payments and cotton cooperative certificate payments.

By requesting a fee waiver because the information was of "public interest," I avoided paying for any of the data. The payment database took about a month to receive, but the others, which are not requested as often, took a little longer. I would recommend requesting expedited action in your FOIA request. Most federal agencies do not recognize such a request, but FSA did.

Once I received 70 GB of data – 180 million records – it was clear that my usual software, Microsoft Access and FoxPro, were not going to cut it.

With trouble-shooting help from our then-CAR editor David Milliron and the book "Sams Teach Yourself Microsoft SQL Server 2000 in 21 Days" by Richard Waymire and Rick Sawtell, I took a crash course in SQL Server.

The capabilities of SQL Server were overwhelming at first. It can handle 1 million terabytes of data, versus 2 gigabytes in Access. (The drawback to SQL Server's power is that it comes with a hefty price tag.)

I learned the organization and the differences in SQL statements on deadline. SQL Server syntax slightly differs from SQL commands in FoxPro (wildcards, for example and excluding the semi-colon after each line in a query). One handy feature with SQL Server is the color-coded command text. SELECT, DELETE, INSERT all turn blue when used properly and values turn red when your syntax is correct.

If you're transitioning from Access or FoxPro, stick with what you know (e.g. familiar queries) in the beginning. Once you get used to the software, it's easier to transition to advanced features. I continue to use Access and FoxPro for quick-hit analyses or daily CAR, but SQL Server has offered more advanced options for more organized, efficient analysis.

SQL Server also provides more advanced security and a great feature called transaction logs, which documents almost every change made to the database. Like most database management software, SQL Server provides advanced processing tools.

For the subsidy story, I used summarizing queries to produce aggregates by several different categories. My biggest interest was finding oddities and outliers. (When you see a person with a Beverly Hills address collecting cotton subsidies in Georgia, it piques your interest.) Through geographic and time-ordered slices of the data, I was able to spot trends. Besides calculating the typical aggregate totals, my objective was to create tip sheets for the reporters in order to identify individual farmers and programs for them to look into.

continued on page 4

DOWN ON THE FARM

COTTON

continued from page 3

Combining the processing, mapping, analysis and memos, I worked on it for about five months.

The payment database includes a separate name table, which includes farm operator names and addresses. The tables are linked by a customer number, which is a surrogate for the tax ID that FSA normally uses. All of the different subsidy datasets can be linked using the customer number. The payment data also includes state codes, county codes, program names, types of crops, date and amount. It does not include the location of the farms, other than the county and state. I started cross-referencing the payment data with entity data, which identifies the members of each entity that collects subsidies. I found many recipients living quite a distance from the farms they are collecting on. People in Atlanta, New York City and even Saudi Arabia and Japan were collecting subsidies for U.S. farms.

It became apparent we needed to rework our definition of "farmer." We discovered that college students, churches, prisons, universities, corporate pension plans and even a sports bra company collect farm subsidies. Familiar names also popped up in the data, too including Chevron Corp., CNN founder Ted Turner and Georgia Gov. Sonny Perdue.

Late last year after we published our project, FSA released a new data set of subsidy recipients, called 1614, that news organizations have not had until now. The dataset, named for Section 1614 of the 2002 Farm Bill, tracks all subsidy payments to a human being when possible. The new data is only 6 GB and contains 66 million records for payments from Oct. 2002 to June 2006.

Previously, some payment records stopped at the corporate level and did not disclose the owners cashing the subsidy checks. The 1614 data connects the dots for you. With the old data, I was able to uncover most of the owners behind corporations with a few extra queries.

For those organizations that don't have access to database management software like SQL Server, The Environmental Working Group, an advocacy group, posts subsidy data online. On its Web site, *www.ewg.org*, EWG illustrates aggregates and provides a searchable database to look up specific recipients and their entities.

After analyzing the payment records, we found:

- Five percent of eligible farmers collect half of all U.S. farm subsidies.
- Farmers can easily circumvent USDA limits on subsidy payments. Nearly 200 U.S. farm operations collected more than $1 million each in 2005. That's nearly a quarter of a billion dollars to a group that could fit inside a grade school cafeteria.
- Subsidies drive up rural land prices and prevent small American growers from farming more land.
- Wealthy Americans and institutions with no apparent need for government aid, including some who live far from their farmland, collect hundreds of millions of dollars in subsidies. Enforcement of income caps for subsidy recipients is spotty.

We discovered that college students, churches, prisons, universities, corporate pension plans and even a sports bra company collect farm subsidies.

The data also turned up great leads on individual farmers. A top USDA official was the nation's single biggest beneficiary of a disaster relief program for orchard owners. Some farm operators accused of scheming to collect millions they were not entitled to later received millions more under new corporate names. Farmers who hit payment limits organized new partnerships to double or triple their subsidies.

The findings highlighted a fundamental flaw with current U.S. farm programs: They benefit landowners, sometimes to the detriment of farmers. That's because cropland owners can collect a subsidy even if they grow nothing, or they contract with a farmer to farm the land and split the subsidy.

We took it a step further and compared subsidy payments with those of another USDA program: food stamps. We found that a third of Georgia's counties collected more in farm subsidies than food stamps.

Agricultural Census data provided an astonishing glimpse at the decline in farms and farmers in Georgia. Compiled by the National Agricultural Statistics Service, the ag census provides state and county-level numbers on everything regarding agriculture and is conducted every five years. NASS will survey farmers in December of this year; the data should be available in February 2009.

AJC Reporter Ken Foskett contributed to this article.

Contact Megan Clarke at *mclarke@ajc.com*.

The findings highlighted a fundamental flaw with current U.S. farm programs: They benefit landowners, sometimes to the detriment of farmers. That's because cropland owners can collect a subsidy even if they grow nothing, or they contract with a farmer to farm the land and split the subsidy.

DOWN ON THE FARM

FARMSUBSIDY.ORG OPENS ACCESS TO EU FARM PAYMENTS

By Brigitte Alfter, *Information*, Tommy Kaas and Nils Mulvad of Kaas & Mulvad

Each year, the European Union pays more than $100 million to subsidize Arla, the seventh largest dairy corporation in the world. In the accounting year 2004, the sum of the subsidies was larger than the profit declared by the Danish-Swedish corporation.

During World Trade Organiation negotiations in December 2005, the EU promised to abolish the export subsidies that provide the largest part of Arla's bounty. The consequence, according to the company's deputy director when confronted with the subsidy figures, would be a restructuring, including moving jobs from Denmark and Sweden to a number of lower-cost countries.

This straightforward business story targeted a Danish audience, but more European journalists now are starting to report in their own language on EU subsidy programs after decades of silence. Almost half of the EU's annual budget is allocated to farm subsidies. In 2007, the amount is €54.9 billion, or $71.2 billion. Yet, until recently, who gets the money has been entirely secret.

Farmsubsidy.org is the driving force in bringing Europe's farm subsidies into the light. This network organization uses Freedom of Information laws and computer-assisted reporting to gather and distribute subsidy data from EU nations. The work to get complete data on the subsidies is vast and far from done, but the potential in the data already gathered is huge. As a result of Farmsubsidy.org efforts, European leaders have promised to publish the data, starting in 2009 (more on that later).

The work really began at a party. Nils Mulvad – who at the time was director of the Danish Institute for Computer-Assisted Reporting, or DICAR, NICAR's sibling in Denmark – attended a family party and met colleague Kjeld Hansen, renowned for his in-depth coverage of environmental issues. They began to talk about the secret subsidies. In June 2004, after a year and a half and a huge FOI fight, the two got the data on who gets what in Denmark.

The data proved what statisticians at government agencies, think tanks and in academia had known for years: The largest landowners and conglomerates get the largest subsidies. Among them are *politicians* (*www.farmsubsidy.org/denmark/Which_Danish_politicians_get_farm_subsidies/221006*) gentry, renowned business people, as well as *state prisons* (*www.farmsubsidy.org/denmark/Danish_prisons_continue_to_get_farm_subsidies/031006*), and a number of large corporations, including *Arla with Denmark's largest share* (*www.farmsubsidy.org/denmark/Arla:_When_subsidies_are_larger_than_the_profit/141106*).

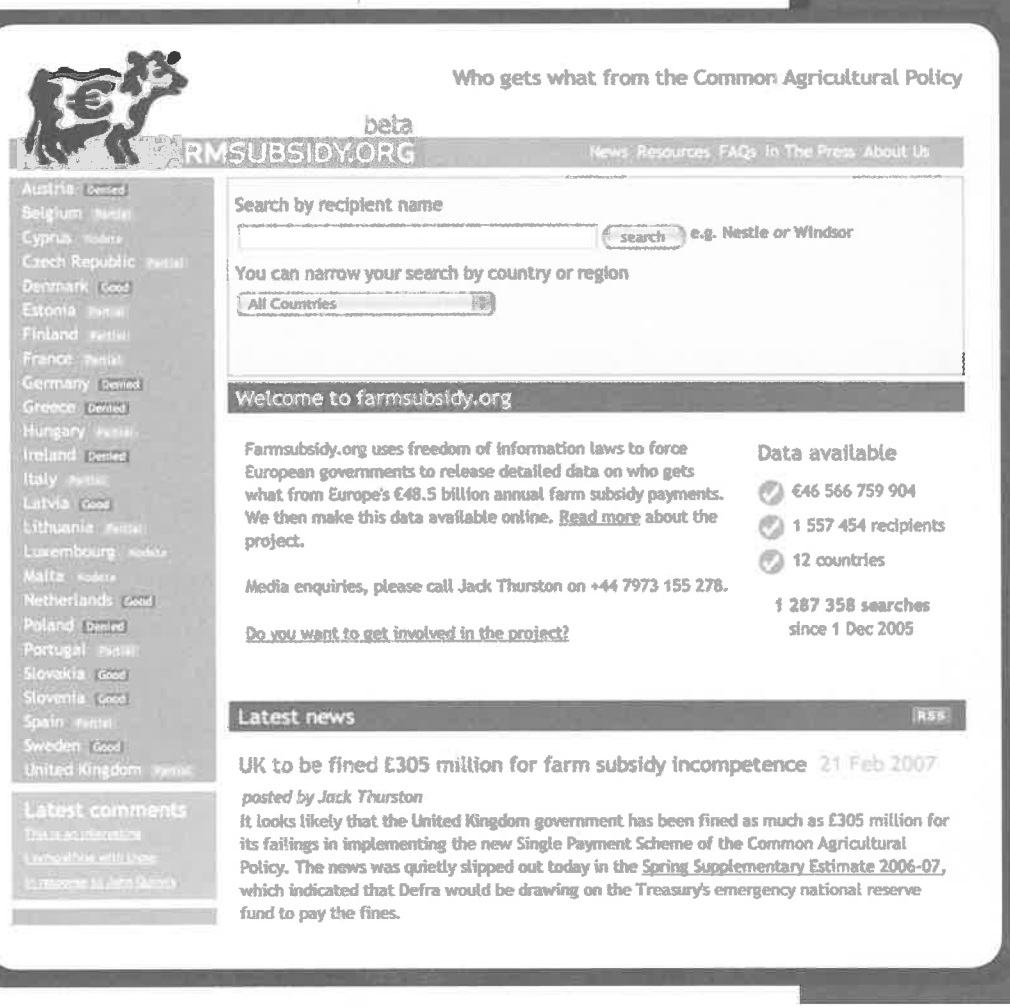
The data provided a powerful reporting tool, made big Danish headlines and created openings for numerous stories to follow. One of the simplest and most appealing stories was comparing a list of the 50 richest recipients with the subsidy data after the EU renamed the direct subsidies to farmers as "income support." A few years after an initial rush of stories in Danish media, it is now obvious how Danish journalists use the data in everyday reporting, for example, mentioning a politician's subsidies as background information in the context of political debates.

In November, *European Voice* named Mulvad the 2006 European Journalist of the Year in recognition of his work with Farmsubsidy.org.

FOI across Europe

The Danish breakthrough inspired others, among them Brigitte Alfter, who was among Mulvad's students at the Danish School of Journalism a decade ago and is now an EU correspondent in Brussels. Her FOI application with the European Commission, which holds all data, became a

continued on page 6



DOWN ON THE FARM

FARMSUBSIDY.ORG

continued from page 5

test case of the *EU's young FOI act* (*http://eur-lex.europa.eu/LexUriServ/site/en/oj/2001/l_145/l_14520010531en00430048.pdf*) that only had come into effect in December 2001.

Alfter filed the application in June 2004, and the commission finally rejected it the following September. The core reason for the rejection was the definition of a database. According to the commission, only information retrieved through "routine operations" could be released. Another argument was a confidentiality clause for EU member states preventing the commission from revealing the data.

The case is still pending at the level of the European Ombudsman (*www.euro-ombudsman.eu.int/home/en/default.htm*). Another option for appeals –although at least as time-consuming as an Ombudsman complaint and far more costly – would have been to go to the European Court of Justice.

The next step obviously had to be addressing other member states. In spring 2005, Alfter filed applications in Germany, where two cases currently are pending in administrative courts. The cases challenge German legislation and legal practices containing some of Europe's strongest privacy and business secrecy protections.

British and Dutch journalists and organisations caught interest, too. British media, including London newspapers *The Guardian* and the *Telegraph*, published stories exposing Queen Elizabeth II, her son Prince Charles and other nobility among the largest beneficiaries of what is politically presented as support for the struggling farmers.

In the Netherlands, the Evert Vermeer Foundation (*www.eucoherence.org/renderer.do/clearState/true/menuId/227658/returnPage/227304*), which opposes EU policies that harm developing countries, achieved access to national data in late 2005 after a number of court cases.

Creating Farmsubsidy.org

By May 2005, it was time for a closer network among those interested in transparency, and a handful of people met in Brussels and drafted the idea for the Farmsubsidy.org network, established by Mulvad and British researcher Jack Thurston. It was time to actively include colleagues in other countries, too. Soon data were released or requested in Sweden, Finland, Poland, Slovenia and a number of other EU member states. When *www.farmsubsidy.org* went online in late 2005, data of varying quality from a number of EU member states were available, and it became possible to trace subsidies for corporations with subsidiaries in multiple member states.

The heaviest arguments used for rejections all over the EU were privacy, protection of business secrets and the fundamental question of access to information held in databases. However, parts of the European Commission started their own struggle for transparency of the data with the commissioner for public administration, Estonian Siim Kallas, taking the lead. Late last year, after a long struggle with some governments, the EU passed a decision declaring that all subsidy data are to be published, beginning in 2009.

Time to lean back and relax at the Farmsubsidy.org network? Not exactly. Particularly in the countries that have resisted transparency most strongly, Network members already can trace movements warning subsidy recipients of the transparency. This easily could lead to a disguise of the true beneficiaries before 2009. Thus, the work continues.

CAR challenges

Work continues on additional information and additional possibilities for the Web site.

Governments provide data in many different formats. Some are uploaded to the Web as PDF files – some, such as Spain's, cover more than 5,000 pages, which are filled with spelling errors. Some countries, such as the Netherlands, only give access to a limited search form, and they try to hinder Web scraping, which makes it very difficult to get a copy of all data for analyzing.

On the other hand, some of the countries are very cooperative and give out data in a standard electronic format (Microsoft Excel or delimited text), which makes it very easy to work with.

Not only the format but also the content varies widely from country to country. Some countries use the calendar year; others count from the day they open applications for a given program, or they follow the so-called FEOGA year used in the EU budget (an administrative concept similar to a fiscal year in the U.S.).

Unfortunately, the subsidy data is really dirty and changes year after year. For Farmsubsidy.org members, it means trying to create standards and get an overview of the data to understand what to request and what pieces of data are necessary. Mulvad uses his journalism training and CAR experience for scraping, importing and cleaning data. He collaborates with Bob Larkin, a data programmer and project leader who works with MySQL, Drupal (*http://drupal.org/about*) and PHP.

Preparing for mapping is another major project. The group has geocoded the addresses of subsidy recipients, a process that allows them to be plotted with geographic information system software, and has requested shapefiles of agricultural land that are kept by authorities in each country. This will allow Farmsubsidy.org to geographically display and analyze data about subsidies and other EU statistics.

We also have rated the score of transparency of government authorities according to how much data they give out and the quality of the data. Now we're close to finishing a process on integrating further indices to prepare it for demands for more data. That will facilitate the process of following the money.

The main new thing on the Web site for the coming period will be combining documentation with the transparency index. It will involve uploading new data, including subsidy programs and possibly also the main layout in a multilingual system, making it easier for people who speak languages other than English.

The whole idea is to open opportunities for new types of stories, call it the next wave, in which we look at the aim of the subsidy program and investigate how it actually works in practice. We look into the effects of the schemes that were not taken into consideration, when they were introduced and whether they now might be outdated.

The entire EU farms subsidy system is going to be submitted to a so-called health check this year and next – even more reason to request the data now – instead of waiting until 2009.

See data at *www.farmsubsidy.org*, for the U.S. data, visit the Environmental Working Group's Web site, *www.ewg.org*.

Contact Nils Mulvad by e-mail at *nils.mulvad@kaasogmulvad.dk* and Brigitte Alfter at *brig@mail.dk*.

DOWN ON THE FARM

REPORTING ON MIGRANT FARM LABOR

By David Walle, IRE and NICAR

When a team of reporters for the *Palm Beach Post* began preparing packages on migrant farm laborers (*www.palmbeachpost.com/moderndayslavery/content/moderndayslavery*) and the dangers of pesticide (*www.palmbeachpost.com/localnews/content/news/special_reports/carlitos/index.html*) in 2003 and 2005, they pursued multiple angles.

They found instances of modern-day slavery, explored the politics that touch migrant lives and calculated the unseen costs of migrant labor. Covering the story required traveling to its sources: in the crop fields of Florida, to the streets of Guadalupe, Mexico, and the trails taken by smugglers to bring migrants into America. Inquiry into the abuses also meant digging into reports provided by state and federal agencies.

Whenever the federal government contracts a state agency to enforce its laws, the agency must collect data and file reports to the federal government, says Christine Stapleton, a *Palm Beach Post* reporter who worked on both migrant packages. Stapleton suggests that reporters first read the contract, which explains the type of data and the format in which it's kept. Reporters can then request specific databases or record fields from the state.

For example, the *Palm Beach Post* tapped the state health department to check for housing violations caused by migrants living in overcrowded homes. Reporters compiled demographic profiles of labor contractors by getting their phone numbers and birthplaces from the state department of professional registration. While working on the pesticides package, they searched worker compensation data for chemical injury claims filed against agricultural employers. Stapleton shares more advice on investigating agriculture in Tipsheet 2086 (*www.ire.org/resourcecenter/viewtipsheets.php?number=2086*).

The following resources also can aid investigations into farm labor conditions:

National Agricultural Workers Survey
(*www.dol.gov/asp/programs/agworker/naws.htm*) – A report of demographic and employment characteristics of the U.S. crop labor force. Released in 2005, the data is taken from face-to-face interviews of farm workers.

Migrant and Seasonal Farmworker Enumeration Profiles Study
(*http://bphc.hrsa.gov/migrant/Enumeration/EnumerationStudy.htm*) – Profiles of workers in ten states (Alaska, California, Florida, Louisiana, Maryland, North Carolina, Mississippi, Oklahoma, Texas and Washington). It includes county estimates of migrant and seasonal workers and each state's labor demand for crops.

Urban Institute
(*www.urban.org*) – Includes various academic articles about immigrants and plenty of statistics.

The Inter-American Development Bank
(*www.iadb.org*) – Has posted two useful papers, "Data on Immigrant Families" and "Remittances 2005," which discusses how migrants send money back to families in their native countries.

Sarasota/Manatee Farmworker Supporters
(*www.smfws.com*) – A collection of national articles about immigration that are updated daily and archived through 2003.

Migrant Farmworker Justice Project
(*www.floridalegal.org/migrant.htm*) – Includes a facts section on farm workers in Florida.

IRE Data, stories and tipsheets
The Wage and Hour Enforcement Database has information on violations, penalties and employers for more than 700,000 closed cases between the 1980s and 2004 (*www.ire.org/datalibrary/databases/viewdatabase.php?dbaseindex=46*). The database covers areas including child labor, overtime, minimum wage and medical or family leave.

WINK-Fort Myers, Fla., reported that between 20,000 to 50,000 people are smuggled into the U.S. each year as indentured servants. Some are forced into confined living conditions or prostitution to pay off their smuggling debts. Story no. 21928 (*www.ire.org/resourcecenter/view.php?number=21928*).

Tom Knudson, Hector Amezcua and Seth Van Booven of the *Sacramento Bee* reported extensively on the mistreatment and abuse of Latino workers who plant and thin federal and private forests. This entry is the online version of a newspaper series. Story no. 22280 (*www.ire.org/resourcecenter/view.php?number=22535*).

Contact the Database Library at 573-884-7711 to order copies of the databases. Information about these and other databases is available online at *www.ire.org/datalibrary/databases*. IRE members can download tipsheets for free. Contact the Resource Center at *rescntr@ire.org* or 573-882-3364 to order copies of these stories or tipsheets.

DOWN ON THE FARM

RESOURCES FOR REPORTING ON AGRICULTURE FROM IRE AND NICAR

These examples can be found at the IRE Resource Center:

Samuel P. Nitze and Beth Reinhard of *The Miami Herald* used local property data to show that "under a 1959 state law intended to preserve agriculture, developers reap huge property tax breaks by herding cows or raising crops in the most unlikely settings. Some pay less than the average homeowner in annual property taxes on parcels slated for multimillion-dollar projects." One developer saved a quarter of a million dollars last year by placing cows on land containing industrial warehouses. Florida has lost about 8 million acres of farmland since the law intended to preserve such property went into effect. This selection includes a methodological description. (Story No. *22277*, 2005)

Allen G. Breed and Martha Mendoza of The Associated Press reviewed property tax arrangements around the country to find that "millions of dollars in property tax breaks intended to preserve farmland are going instead to companies that bulldoze farms to build housing subdivisions, malls and industrial parks." (Story No. *21066*, 2004)

Paula Lavigne of *The Dallas Morning News* found that "the United States Department of Agriculture does not know how often organic rules are broken and has not consistently taken action when potential violations were pointed out." Audits and inspection reports point to weak oversight of the certifying organizations that bestow official organic status on behalf of the USDA to more than 20,000 producers worldwide. (Story No. *22710*, 2006)

Read more about farms in *The IRE Journal* (Mar./April 2007)

The Resource Center also offers tipsheets designed to aid agricultural investigations:

Brigitte Alfter provides tips about how to investigate the farm and agriculture industry in the U.S. and Europe. She also gives information about FOI in the European Union. Furthermore, she provides many helpful Web links which can aid reporters in their reporting. (Tipsheet No. *2681*)

Ken McCall's PowerPoint presentation from the 2003 IRE Conference in Washington, D.C., focuses on a study done by the Dayton Daily News that analyzed animal waste regulations in Ohio. He uses his own example to help guide others through the process of building a database when data is not available from regulatory or state agencies. (Tipsheet No. *1933*)

Because farm subsidies account for nearly half of all European Union grants, investigating fraud cases can lead to some great stories. This tipsheet discusses the various aspects of these subsidies and where one is most likely to find fraud. (Tipsheet No. *2568*)

This tipsheet provides an "essential" list of Web sites for journalists to use when investigating agriculture-related stories. It also gives the addresses for Listservs and discussion group sites. (Tipsheet No. *1098*)

To order books, tipsheets or story reprints, contact the IRE Resource Center (*www.ire.org/resourcecenter*) at 573-882-3364.

Data

The **Consolidated Federal Funds Report**, or CFFR, *www.ire.org/datalibrary/databases/viewdatabase.php?dbaseindex=18* includes records of federal expenditures or obligations to states, counties and local agencies, including grants, procurement contracts, salaries and wages, direct loans and insurance. The data can be broken up by agency and location of recipient down to county level. This database includes records on the expenditures of more than 20 entities working in the agricultural sector.

The **Federal Assistance Award Data System** *www.ire.org/datalibrary/databases/viewdatabase.php?dbaseindex=38* contains records of federal assistance awarded to state and local governments as well as all major programs giving transfer payments to individuals, loans or insurance. This data contains more than 120,000 records directly relating to agriculture.

The **Single Audit Database** *www.ire.org/datalibrary/databases/viewdatabase.php?dbaseindex=42* includes audit material on every non-federal entity that expends $300,000 or more of federal funds per year. As opposed to CFFR and FAADS, the Single Audit Database looks at spending, not receiving federal assistance.

MORE RESOURCES FOR REPORTING ON AGRICULTURE

U.S. Department of Agriculture's National Agricultural Statistics Service, *www.nass.usda.gov*. NASS uses state-level partners to track loads of agricultural data, including planting, harvest and yield data for field crops, breeding and slaughter numbers for animal agriculture and information on all kinds of specialty crops, from Christmas trees to tobacco.

- The USDA also tracks farm numbers, size, income and demographics on farm operators. Monthly and annual reports are supplemented with a more detailed Census of Agriculture report released every five years. The preliminary phase of the 2007 Census of Agriculture is currently underway, but initial reports from respondents are not expected until around February 2009. *www.nass.usda.gov/Census_of_Agriculture/index.asp*

- Other examples of NASS data include government purchases of agricultural products, vegetative condition images and cropland data layer images. *www.nass.usda.gov/Charts_and_Maps/index.asp*

- The 2005 Aquaculture Census includes U.S. and state-level data for everything from baitfish and sport fish to ornamental fish and mollusks. *www.nass.usda.gov/Census_of_Agriculture/2002/Aquaculture/index.asp*

The **Meat, Poultry and Egg Product Inspection Directory**, which lists establishments that produce meat, poultry and/or egg products regulated by USDA's Food Safety and Inspection Service.
www.fsis.usda.gov/regulations_&_policies/Meat_Poultry_Egg_Inspection_Directory/index.asp

Residue Violators Alert List, monthly list of individuals or firms responsible for repeat drug, pesticide or other chemical residue violations in animals presented for slaughter. *www.fsis.usda.gov/PDF/Residue_Violators_List.pdf*

Quarterly Enforcement Reports provide a summary of the enforcement actions FSIS has taken to ensure that products that reach consumers are safe, wholesome and properly labeled *www.fsis.usda.gov/regulations_&_policies/Quarterly_Enforcement_Reports/index.asp*

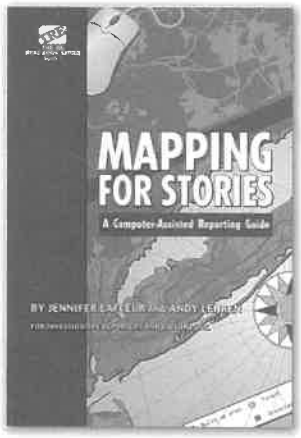
FSIS Audit reports of U.S. Trading partners details slaughterhouse inspections. *www.fsis.usda.gov/regulations_&_policies/index_of_certified_countries/index.asp*

Mapping for Stories

Learn how to map data for daily news stories and larger projects. This book gives professional journalists, as well as journalism educators and students, the foundations for using mapping software to dig deeper into stories. It includes story examples in every section to help show how these skills translate into daily journalism. Downloadable maps and data allow reporters to work along with each chapter and follow lessons at their own pace. It is structured to help journalists complete better stories.

IRE members: $15 each
Nonmembers: $25 each

Plus media-rate postage: $4 for the first book; $2 for each additional book.
E-mail questions to rescntr@ire.org



ORDER NOW!

By PHONE
Call 573-882-3364 with your VISA, MasterCard or American Express

By MAIL
Send your check to IRE Resource Center: 138 Neff Annex Missouri School of Journalism Columbia, MO 65211

By WEB
Visit www.ire.org/store for online ordering or to download an order form

DOWN ON THE FARM

USDA

continued from page 1

successfully sued the USDA over the subsidy data in 1995. We'd never done anything with the data once it was released in 2000, but the lawsuit helped shake loose other records during the "Harvesting Cash" project because the department believed we would take them back to court.

The basic subsidy data is relatively easy to get from the USDA's Kansas City office. Its FOIA office routinely releases the subsidy payments, also known as the "check file," along with a lookup table for names and addresses. Each year's data contains about 24 million transactions covering about 1.5 million recipients. We used the records from 1990, which we'd saved from the original FOIA, and updated with new ones throughout the project. It amounted to about 217 million records in all, but the database was one of the simplest examples I've ever worked with. The only difficulty was its size. (see Fig. 1)

The payments cover most, but not all, of the subsidies given by the USDA. The government said it contains all of the subsidies that are "taxable events". Also, some conservation programs are excluded because the USDA Farm Service Agency says it doesn't have the records.

We stored the basic payment data in an SQL Server database, and it was accessible to the whole team through an Intranet lookup for both recipients and county comparisons. It meant the other team members could look up anything they needed and freed me to do more analysis and exploration.

Money for nothing

Our approach was classic for computer-assisted reporting. In most cases, a source told one of us about a possible problem in a subsidy program. We'd try to figure out how it might look in the database, or in some combination of databases, and then we ran queries to check. Once we found some examples, we would either interview the farmers by phone or visit an area. Then we could refine the queries to estimate the total amount of payments involved.

Two of the analyses posed the most difficulty. In the first story, called "Reaping money for nothing", we wanted to estimate how much in crop subsidies went to those who didn't grow anything. Only the USDA has those records, and it was not willing to release the production figures for each farm.

I took two independent stabs at estimating the amounts with a troubling result: One estimate was almost five times the other. In one case, we compared the amount of government payments that farmers reported in the Census of Agriculture (*www.nass.usda.gov/Census_of_Agriculture/index.asp*), with the amounts actually paid by the government for the same programs over the same time period.

This could only be done for the Census years of 1997 and 2002 and was troubling not just because the numbers were so high. It was also because we had no way to check them; the census method relied on aggregate statistics, and we couldn't drill down to actual farmers or landowners. Census officials agreed a key explanation of the difference was that the census only surveys farms, while farm payments are also sent to passive landowners. (All of our analysis excluded conservation payments and focused only on payments to working farms for active acreage.)

We were lucky that the Post had successfully sued the USDA over the subsidy data in 1995. We'd never done anything with the data once it was released in 2000, but the lawsuit helped shake loose other records during the "Harvesting Cash" project because the department believed we would take them back to court.

The other method looked at the farms directly. Over time, virtually all working farms collect either a price support or a disaster payment that depends on growing, or trying to grow, a crop. So I looked for farms that collected the crop subsidy but none of the other subsidies over five years. This came out to the much lower figure of $1.3 billion over five years (as compared with $1.5 billion for just 2002 using the census approach).

Although agricultural economists we consulted with felt that the larger figure was probably pretty accurate, they agreed that it would be much harder for the second method to cast a net over people we didn't expect, such as landowners who share-crop. So we ignored the census approach and stuck with the smaller number. It had also proven to be very accurate based on the dozens of recipients we had called or visited.

The second difficult story involved a subsidy called the Livestock

Figure 1

	A	B	C	D	E	F	G	H	I
1	CustID	StateFIP	CountyFIP	FarmNum	TransDt	TransAmt	ProgCode	CalCode	ProgYear
2	000132141	24	031	0001107	12/8/05 0:00	211	6740	D8	2006
3	000132141	24	031	0001302	12/8/05 0:00	89	6740	D8	2006
4	000952984	24	031	0001149	2/24/06 0:00	6	6740	D8	2006
5	000952984	24	031	0001515	2/24/06 0:00	18	6740	D8	2006
6	000952984	24	031	0001516	2/24/06 0:00	3	6740	D8	2006
7	000952984	24	031	0001971	2/24/06 0:00	177	6740	D8	2006
8	000952984	24	031	0001978	2/24/06 0:00	29	6740	D8	2006
9	000952984	24	031	0001979	2/24/06 0:00	448	6740	D8	2006
10	000953380	24	031	0001362	2/8/06 0:00	798	6740	D8	2006
11	000953380	24	031	0001362	2/8/06 0:00	82	6740	D8	2006
12	000953380	24	031	0001623	2/8/06 0:00	283	6740	D8	2006
13	000953380	24	031	0001920	2/8/06 0:00	80	6740	D8	2006
14	000982181	24	031	0001677	2/22/06 0:00	34	6740	D8	2006
15	000999489	24	031	0001362	2/8/06 0:00	1059	6740	D8	2006
16	000999489	24	031	0001362	2/8/06 0:00	185	6740	D8	2006
17	001228582	24	031	0000231	2/8/06 0:00	3	6740	D8	2006
18	008402549	24	031	0001928	2/16/06 0:00	534	6740	D8	2006
19	008402549	24	031	0000253	2/16/06 0:00	48	6740	D8	2006
20	008402549	24	031	0000262	2/16/06 0:00	30	6740	D8	2006
21	008402549	24	031	0000285	2/22/06 0:00	14	6740	D8	2006
22	008402549	24	031	0000754	2/16/06 0:00	42	6740	D8	2006
23	009484441	24	031	0001751	2/8/06 0:00	44	6740	D8	2006
24	009580050	24	031	0001315	2/22/06 0:00	162	6740	D8	2006

Payment table | Name table | Sheet3

Ready

DOWN ON THE FARM

Compensation Program. Created in 2002, the fund was supposed to help ranchers and dairy farmers buy feed in areas that were suffering an historic drought. But the government never required anyone to show any losses to qualify.

Lists of agricultural disaster areas were scraped from the USDA website and obtained under FOIA. Another list of presidentially declared disaster areas is maintained by FEMA and was readily available through a FOIA. Once they were cleaned up and coded, comparing the two lists revealed counties that qualified only because of an event that had no effect on agriculture.

The harder part was estimating how much money went to counties that never had a serious drought. The U.S. Drought Monitor, considered the most comprehensive gauge, publishes a map every week of the extent of drought in the nation. Data prior to 2004 are still not in a GIS system. Older maps were colored by hand and only existed as JPEG images on the monitor's Web site (*http://drought.unl.edu/dm/monitor.html*), the director and other members of the group said.

To convert a JPEG into a map, ArcView will import a picture, allow a user to approximately fit it to an existing map, and then use a process called "georeferencing" to match the map to an existing layer. (The picture and the underlying map should be in the same projection if at all possible.) Most of this is done through the georeferencing toolbar. (See Fig. 2) After setting control points – lining up the tip of Florida, Maine, Washington and the Mississippi River, for example – ArcView will create a raster layer. But it still can't be merged with a county map. That requires another step: converting the raster layer to a vector layer, which can be done through the Spatial Analyst extension or in the ArcMap toolbox.

This produces a polygon shapefile that has attributes based on the colors of the original picture. At that point, I could do spatial joins against county maps to find the worst drought condition in any part of the county over the two years. Although the entire process was tedious, it only took about a day or two to get through the entire exercise.

When compared with the program's payments, it was clear more than half of the money from a program created to help people suffering through an historic drought was spent in areas that never suffered anything worse than a short, moderate drought.

More data sources

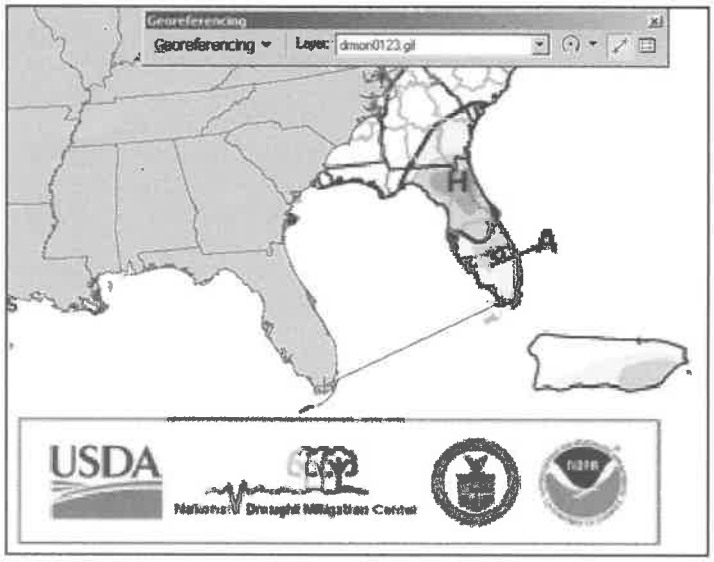
Other stories also depended on comparing various databases with the payment records.

Some of those sources included:

- Historical acreage, production and prices by crop for counties and states from USDA's National Agricultural Statistics Service Web site (*www.nass.usda.gov*).
- Historical daily government estimates of prices for each county in the nation, scraped from the Farm Service Agency's Web site (*www.fsa.usda.gov/FSA/webapp?area=home&subject=prsu&topic=landing*).
- Crop insurance coverage, costs, losses and subsidies by county and crop from the Risk Management Agency (*www.rma.usda.gov*). The databases on the RMA's Web site for download had numerous errors in them. The agency sent us a new copy, but I believe it also updated the data on the Web site at the same time.
- A list of crop insurance agents scraped from the RMA's Web site. We were interested in people who would be expected to cover their costs with insurance.
- Ownership records, called the Entity Relationship File, of individual farms. It took months of complaints and a threat from our lawyer to obtain these, even though we won a FOIA appeal within days of submitting it. They show full ownership records for each subsidy recipient, even when they are complex arrangements designed to legally sidestep payment limits.
- The 1614 data, named after a section of the 2002 Farm Bill requiring USDA to study the effects of limits on payments. This database was released two days before our final story of 2006. In a story on the politics of farm policy, we used it to estimate the amount that would have been saved if a key amendment to the 2002 Farm Bill had passed.

The second difficult story involved a subsidy called the Livestock Compensation Program. Created in 2002, the fund was supposed to help ranchers and dairy farmers buy feed in areas that were suffering an historic drought. But the government never required anyone to show any losses to qualify.

Figure 2



Georeferencing was used to convert a JPEG map image to a raster layer in ArcView.

Contact Sarah Cohen at
cohensh@washpost.com

FIRST VENTURE

LEGISLATURE

continued from page 1

dominant ideological views about the future of the island: full statehood, independence and the status quo. Each race for governor takes on a presidential allure and becomes a matter of consciousness, a decision about who each voter thinks he or she is and who they want to be. As a popular saying goes, politics is Puerto Ricans' favorite sport.

Government gridlock

In 2004, Puerto Rico had a very close election quite similar to the 2000 Bush-Gore presidential race. A controversy emerged over the validity of certain ballots and the manner the in which they were to be adjudicated. The current political environment is very charged and little has gotten done in the last two years. Meanwhile, citizens get faulty services, crime soars and the island is stuck in an economic crisis.

Editor in chief Luis A. Ferré Rangel proposed creating our own accountability system to measure legislators' productivity and assigned the chore to special projects editor Agnes Montano López and me. Montano, who lives in San Diego, Calif., had some CAR experience, but I had no CAR background.

We decided to focus on the legislation submitted in the first session of the year. The bureaucratic process consumes a great deal of economic resources, and, if legislation is approved, it affects the quality of life of all citizens. So, I went to the administrative office of the Senate and House of Representatives and asked for a mailbox, like most lobbying groups do, but we hoped to find a more user-friendly format because there were more that 4,300 legislative pieces alone from January to June 2006. We started the project with about two months to complete the database, analyze it, request more information, get interviews and write stories.

A trip to Boot Camp

In the midst of this, my editor sent me to IRE's CAR Boot Camp at the Missouri School of Journalism to learn how to use Microsoft Access. I learned how to use Microsoft Excel in more useful ways (for example, sorting data and making charts to track trends), some basic Access tricks and how to transfer back and forth between both programs.

The most important thing I took back was the idea that government offices keep a lot of public information in different electronic formats that remain virtually unknown because nobody ever asks for it.

I came home determined to do just that. The legislative aides and clerks had the information. The legislative aides and clerks had the information, but each body kept it in totally different formats each body kept the data in totally different formats, and they had a three-month backlog in data entry. Just as I was about to enter the missing data myself, one of the aides mentioned that each legislator must submit an electronic copy of every piece of legislation to the Office of Legislative Services, which posts the information online.

The site, *www.oslpr.org*, was great, and the staff was really knowledgeable and starving for some attention. Incredibly, the legislation was generally available online within two hours of filing, along with investigative reports and statistics related to the legislature.

Getting results

Projects editor Montano helped with the data-gathering from San Diego. After some trial and error, we could download it directly into Excel tables. We used eight fields: last and first name of the legislator, party affiliation, district represented, date of filling of the measure, the measure number, type of measure (bill or resolution), a brief summary of the content, and subject matter. Half of the fields came from the Web site, and we filled in the rest.

We created tables with the same fields for each legislator, for the Senate, for the House and one combined, so we could evaluate performance of individuals, each body and the legislature as a whole. Then we transferred the data to Access for queries. The final database had 6,319 records.

We discovered many interesting and important things. First, in the middle of one of Puerto Rico's worst economic crises, more that half of the legislative efforts were dedicated to congratulating people, often for no extraordinary reason. Newcomer Nelson Del Valle Colón presented the most proposals overall, but almost 92 percent, or 266, were congratulatory measures. When we interviewed him, he was honored. He had no idea of the great administrative cost incurred in filing a piece of legislation.

The second most popular category was resolutions to investigate one thing or another. These accounted for 16 percent of the total. Most of the 690 investigations focused on inconsequential issues that could be solved with a letter or phone call. The best part is that only 6 percent of those investigations were actually completed.

Joint resolutions to allocate funds accounted for 9 percent. We found 340 resolutions to allocate a total of $120 million from the deficit-ridden general fund for all sorts of agencies, organizations and even individuals, including things like repairing constituents' homes and buying home appliances. All of this occurred within a government that had to shut down several agencies, including public schools, in June 2006 because of insufficient funds.

Most disappointing was our discovery that legislation on core issues for citizens got little attention. Public health measures only accounted for 2 percent of the pieces filed, and education measures were even fewer. Agriculture, technology, telecommunications, and tourism (one of our most important industries) were relegated to less than 0.3 percent of the total.

The series of seven stories drew praise from experts who study legislative issues in Puerto Rico. It also generated interest from a local university that wants to maintain the database through its public policy institute. As I write, Montano, reporter Israel Rodríguez Sánchez and I are preparing to do the exercise all over again to evaluate the latest legislative session.

Contact Omaya Sosa Pascual at
ososa@elnuevodia.com

RACE

continued from page 1

noticed a picture on one wall.

In the top of the frame was a photograph of a farmer and his wife taken some time before the Civil War. Below the picture was the farmer's will. On separate lines, he carefully recounted each of his earthly possessions, parceling them out to family and friends. It was what you would expect until I reached the fourth or fifth line. Wedged between livestock and land were five slaves to be given away.

Blacks had lived here in the past, but during the several days I spent in the Berryville area, I had not seen a single African American. As an afterthought in the last interview I did before returning to Washington, D.C., I asked a woman where the blacks were.

"Oh, the Klan keeps them out," she confided.

Census yields clues

When I returned home, I downloaded some 1990 Census data from the U.S. Census Web site and performed a query in Microsoft Access database manager to find all the counties in Arkansas with few, if any, blacks. To my astonishment, about a third of the counties popped up. I expanded my search to the South and later to the Midwest and parts of the North. The list of white or virtually all-white counties grew accordingly.

When I mentioned this to someone visiting from Georgia, she told me the story of blacks being driven out of Forsyth County years ago. Had the same thing happened in other states? Is this why I was seeing so many lily-white counties?

I found a site at the University of Virginia (*fisher.lib.virginia.edu/collections/stats/histcensus*) that contained county-level census data going back to the nation's very first census. It was a simple matter to download data for about 30 states from the Web site and import them into a database. (The Virginia site offers online queries and customized maps, but it also links to the University of Michigan's Inter-University Consortium for Political and Social Research, which offers raw census data from 1790 to 1950.)

Using Access, I wrote a program (see box) that read through the Web site's database, found those counties where the black population had collapsed, and then saved that county data to a new file. The program ordered the Web site's database ("Cnty-cen" in the program) by state, county and census year and then reading one county at a time into an array. Within that array, the program hunted year by year through the census of blacks and, if it found a drop, emptied the array – the county data – into a "suspect" database.

Option Compare Database Script, designed for use in Microsoft Access is available at *www.nicar.org/techtips.html*. To use, cut and paste it into a new module, then edit database and field names to reflect your data.

At this stage, I made a number of assumptions that proved to be fairly useful. I defined a black population collapse as a drop of 50 percent or more between census years. I also decided to look at the years 1890 to 1930. Although there were racial cleansings before 1890, the bulk occurred around the turn of the century. I also found that the 50 percent benchmark proved sensitive enough to find racial cleansings while ignoring the normal fluctuations in population

I had not expected so many counties to show sudden population drops. I had assumed there would a few dozen counties or 50 at most. Instead, I found more than 200.

There were also a couple of inherent problem with using census data. One was that only county-level data had been computerized. If blacks were driven out of one town in a county, there's no way of seeing it. But the greater problem was that census data could only narrow my search to a particular county and a 10-year period.

History lessons

To document a racial cleansing, I needed to know exactly when it occurred. For better- known incidents, like Georgia's Forsyth County, this was not a problem. All it took was a phone call to a local historian or a few hours searching through a county history book. In the vast majority of cases, I had nothing except the county name and the gyrations in the census numbers.

My first approach was to interview a variety of people, from members of the local historical or genealogical society to the history faculty at colleges and universities. When that failed – and all too often it did – I would go to the Library of Congress and read through the microfilm of a state's major newspaper for a ten year period. Day after day, I would scan old newspapers searching for some clue to help explain a population collapse.

As tedious as this was, I came to realize that these old newspapers had become my patient tutors. I had known, for example, that post-Reconstruction blacks were denied the right to vote, but I always had assumed it was a furtive practice. Not so. It was not unusual to come across news stories like one in a Virginia newspaper openly discussing the best way to disenfranchise blacks without affecting white voters. Other stories warned that the bite of the "blue gum nigger" was poisonous or how a local grocer's convention was entertained by blacks parading around in a "cakewalking" contest. They were so outrageous that I began keeping a file of them.

Some stories were so common – blacks being burned at the stake was one category – that I soon found myself skipping over them. In a matter-of-fact, quotidian voice, these old newspapers documented the world in which these racial cleansings took place. It is one thing to be told that several hundred lynchings took place in a year. It is quite another to read about them day after day in news stories that frame them as white men of grim resolve meting out justice to "burly, black brutes."

Tracing racial cleansings

I kept stumbling across short articles, often used as filler, that described all the blacks being driven out of a town or chased from some village. Measured against the violence of lynchings and race riots, the people who wrote about these expulsions apparently considered them interesting but

continued on page 14

RACE

continued from page 13

not particularly remarkable. I got the sense that, while not everyday occurrences, they were relatively common.

The very ordinariness of these racial cleansings presented several problems. As time passed, the memory of these events simply disappeared from the public consciousness. When I talked to historians about racial cleansings, they would try to correct me. What I was looking at, they assured me, was probably one of the effects of the Great Migration during World War I, a period in which southern blacks moved North. Nobody had been driven out, they explained. People just left. Another problem was there were so many possible cleansings that I could never adequately document all of them.

My solution was to limit my story to those racial cleansings where there was a clear ultimatum delivered to leave or die, where it was documented through contemporary accounts, and where the expulsion was "successful," for lack of a better term. I defined "successful" as a county that remains virtually all-white today. Using this approach, not only would I have a manageable number of cases to report on, but there would be no doubt that these were racial cleansings and not some statistical quirk or a footnote to trends like the Great Migration.

Even though I eventually narrowed my story to a dozen cleansings, documenting what took place proved a mammoth task that took eight years. During that time, I often would stop to work on other projects, but I would guess that, if I totaled up all the hours, I spent at least three years researching and writing the history of racial cleansings.

I come away from this project with a new appreciation for the truism that reporters write the first draft of history. When newspapers of the early 1900s did a good job covering the expulsion of blacks, it was relatively easy for me to piece together what had happened. Excellent examples can be found in the *St. Louis Post-Dispatch* and *The Carthage (Mo.) Evening News*. But in far too many cases, history has suffered due to choices made by reporters or editors who lived in that era.

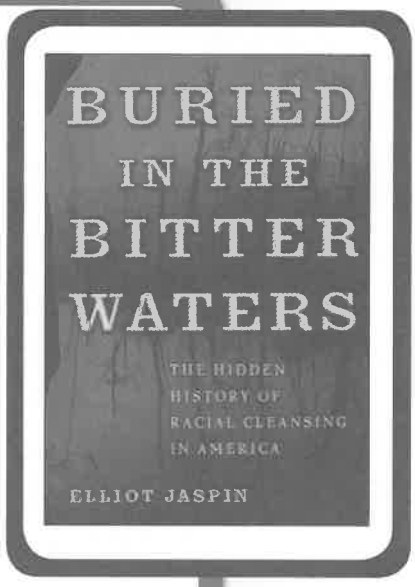
The story today

I am also struck by how critical our past is to how we function as a society. During my research, one editor dismissed my project as a waste of time and said it would have "no impact." But what I found was that our nation has two histories: one white, the other black. While whites had no memory of anyone being driven out, the history of these terrible events is passed by word of mouth in the black community.

But these stories have not been completely lost. The Associated Press series "Torn from the Land" explored 107 cases in which whites – using trickery, legal maneuvers or violence – drove off black property owners and claimed their land.

In 2005, "Sundown Towns" by James W. Loewen (New Press) helped resurrect the history of towns that refused to allow black residents. The book's Web site (*www.uvm.edu/~jloewen/sundowntowns.php*) includes related articles and a database of suspected "Sundown Towns" nationwide.

As with so many things, computers and the Internet are making it impossible to hide this past, too. Sites like *Ancestry.com* and *NewspaperArchive.com* have been converting all those old newspaper microfilms to digital libraries that can be searched electronically. When that process is finished, a reporter will be able to type in keywords such as "Negro" and "exodus" and search thousands of small local newspapers going back over a century.



I also imagine that historical census data at the city and town level will also be computerized in the future. When that happens, it will be possible to resurrect a complete picture of America's racial cleansings.

But all the gee-whiz technology in the world will still not substitute for reporters and editors with the courage to come to grips with their pasts. Unless we take it on ourselves to research and write about what happened in our communities, the nation will be dogged by an incomplete understanding of what brought us to this point.

For more about this story:

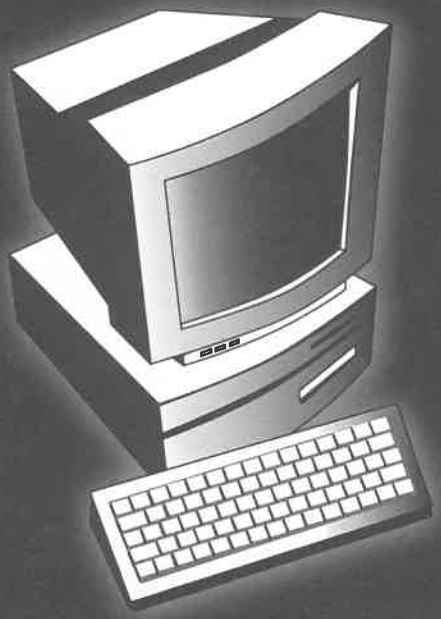
"Buried In The Bitter Waters: The Hidden History of Racial Cleansing," Jaspin's book about racial expulsions in the U.S. was released on March 1. In July, 2006, he published part of his work in a four-part series called "Leave or Die" in the Austin (Texas) American-Statesman. (*www.gjsentinel.com/news/content/shared/news/interactives/lod/index.html*).

The Center for Investigative Reporting and Two Tone Productions have produced a documentary called "Banished," informed by Jaspin's reporting and recently screened at the Sundance Film Festival. Read more about it at *www.muckraker.org/pg_one_investigation-1257-11-0.html*.

Elliot Jaspin can be reached at *elliotj@coxnews.com*.

Even though I eventually narrowed my story to a dozen cleansings, documenting what took place proved a mammoth task that took eight years. During that time, I often would stop to work on other projects, but I would guess that, if I totaled up all the hours, I spent at least three years researching and writing the history of racial cleansings.

Jumpstart Your Career!



COME TO THE MISSOURI SCHOOL OF JOURNALISM AND HONE YOUR SKILLS AT INVESTIGATIVE REPORTERS AND EDITORS AND THE NATIONAL INSTITUTE FOR COMPUTER-ASSISTED REPORTING (NICAR) WHILE GETTING A GRADUATE DEGREE IN JOURNALISM.

THE MORE THAN A DECADE-LONG COLLABORATION BETWEEN THIS TOP JOURNALISM SCHOOL AND IRE HAS RESULTED IN DOZENS OF JOURNALISTS MAKING THE LEAP TO THE SOME OF THE BEST NEWSROOMS IN THE COUNTRY.

IF YOU QUALIFY FOR ADMISSION TO THIS SCHOOL AND A GRADUATE RESEARCH JOB AT IRE, YOUR TUITION WILL BE WAIVED AND YOU WILL RECEIVE A MONTHLY STIPEND WHILE GETTING HANDS-ON EXPERIENCE IN NICAR'S RENOWNED DATABASE LIBRARY AND OBTAINING A MASTER'S DEGREE FROM THE SCHOOL.

GRADUATE STUDENTS WORKING IN THE DATABASE LIBRARY HAVE GONE ON TO CAR JOBS AT NEWSROOMS ACROSS THE COUNTRY INCLUDING:

- THE ATLANTA JOURNAL-CONSTITUTION
- THE CENTER FOR PUBLIC INTEGRITY
- THE DALLAS MORNING NEWS
- DATELINE NBC
- DETROIT FREE PRESS
- THE KANSAS CITY STAR
- MIAMI HERALD
- THE NEWS AND OBSERVER IN RALEIGH, N.C.
- THE NEW YORK TIMES
- THE SEATTLE TIMES
- ST. LOUIS POST-DISPATCH
- THE WASHINGTON POST

FOR MORE INFORMATION ON IRE AND NICAR, SEE *WWW.IRE.ORG*.

FOR MORE DETAILS ON GRADUATE STUDIES AT
THE MISSOURI SCHOOL OF JOURNALISM,
VISIT *HTTP://JOURNALISM.MISSOURI.EDU/GRADUATE*

E-MAIL *BRANT@IRE.ORG* OR CALL 573-882-2042.

VIOLENT CRIME TRENDS PREDATE KATRINA EVACUEES

By Mark Greenblatt
KHOU-Houston

MAPPING IT OUT

THE LATEST USES OF MAPPING IN NEWS REPORTING

On one particularly hot Houston evening last summer, a crowd of frustrated citizens flowed over into the halls and aisles of Grace Presbyterian Church. They sensed a growing problem on the streets of Houston and wanted answers from the city's mayor and police chief.

"Failure is not an option," one frustrated citizen told the mayor.

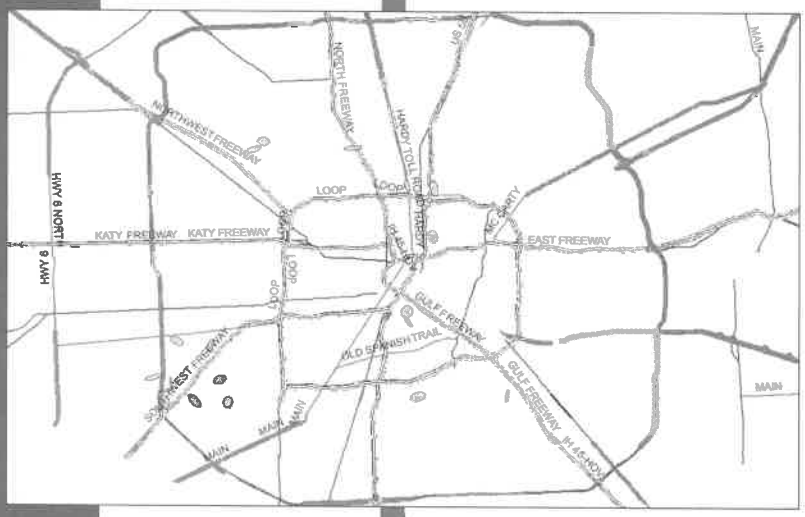
The demands came nearly one year after Hurricane Katrina sent 150,000 evacuees to settle in Houston, and police admitted the homicide rate in this city had dramatically increased.

The official word from the public relations staff at the Houston Police Department: Blame the evacuees from New Orleans for the murder rate increase.

As Houston Police Capt. Dwayne Ready told *The Washington Times* last October, "We recognize that the homicide rate is up as far as raw numbers and as well as percentages relative to the population. ... We also recognize that Katrina evacuees continue to have an impact on the murder rate."

Curiously, the police department remained focused on telling the public about the recent rise in homicides, but when pressed by an increasingly frustrated public about other crimes, the department would sometimes cite statistics showing an *overall* decrease in violent crime. Police sidestepped talking publicly about other violent crime trends or other neighborhoods. They told us they had only studied "the Katrina effect" as it related to homicides.

Executive producer David Raziq, producer Chris Henao, photojournalist Keith Tomshe and I wanted to know why. So, we decided to begin our own independent investigation of crime trends throughout the city. Our analysis would eventually cast doubt on claims that Houston's growing problem with violent crime could be blamed solely on the evacuees. In fact, we saw strong evidence to the contrary.



Data adds detail

We began by requesting the most detailed incident-level data available in Houston for every crime committed in the city throughout the last three years. Each crime record we received contained information on the offense date, an incident number, offense code, police beat, census tract, city, county, time of offense, day of week, premise code and the address of the crime. We received the information on nearly 450,000 crimes in an ASCII plain-text file, which we imported into Microsoft Access.

To search for trends, we honed in on the offense code assigned to each incident. The offense code is a numerical description of a specific kind of recorded crime. We had to use the police department's data dictionary to decipher what each code really referred to because the records track several hundred different kinds of crime.

The offense code would provide details critical to our analysis. The Houston Police Department had told the public the number of serious assaults was down across the city. It was a true statement – but only if you lumped all kinds of those assaults together.

Examining the more detailed offense codes revealed the larger story because it showed the total number of "assaults" that were really shootings or stabbings compared to those that were much less serious. It turned out that the number of less serious assaults were down. But, we discovered a dramatic increase in the

worst of the worst: aggravated assaults with deadly weapons.

We continued our research by transferring our findings to Microsoft Excel spreadsheets to look for changes over time in the number of incidents for each offense code. Looking at the rate of change, our calculations showed that the number of aggravated assaults with deadly weapons in Houston had actually gone up by 22 percent in just two years' time.

For example, during the first seven months of 2004, Houston saw 1,804 incidents of aggravated assault with a deadly weapon – the kind of assault that is so serious one expert called each incident "a failed attempt at murder." By 2005, that number jumped by 9 percent to 1,976 incidents. The surge accelerated to 2,209 in 2006.

Digging deeper, we uncovered more disturbing trends in other crimes. Home burglaries by forcible entry were up more than 25 percent over the last two years. Robberies of gas stations were up 73 percent in one year; and the surge in this crime also began a full year before Katrina.

Revealing hot spots

We also wanted to know how these crime surges were changing our neighborhoods, block by block. In order to look for hot spots, we used ArcView GIS mapping software to show the location of all 6,000 incidents of aggravated assault with a deadly weapon.

The process took us longer than we would have liked because the police department's data needed an excessive amount of editing and fact-checking to make the mapping possible. We used police beat and census tract information to help verify the accuracy of our map. We decided to plot each year's incidents separately so we could look for neighborhood shifts in this crime over multiple years.

Once we had our maps together, we asked Dr. Ned Levine, the author of a program called CrimeStat, to work with us as we continued to narrow in on hot spots.

CrimeStat is a spatial statistics computer program. It was funded by grants from the National Institute of Justice and used by police agencies across the nation to discover crime trends. It allowed us to see statistically significant relationships among clusters of incidents, going beyond what Excel or Access would reveal.

For instance, looking at a spreadsheet alone, we never would be able to see relationships among incidents that might be 15 feet away from each other but in different ZIP codes. CrimeStat has a number of different ways to search for hot spots. We chose a function called "nearest neighbor," which allows you to spot areas that have a certain number of incidents within a set distance of each other.

The research collaboration uncovered a number of new hot spots that had sprouted up all over Houston, and it showed shifts in where clusters of crime were occurring. It was information we knew residents in those neighborhoods – not to mention the police force – would need to know.

However, before rushing to air the story, we wanted to make sure we were right. We asked two criminologists at the University of Houston Downtown and a third expert in crime analysis to check our work. All three researchers confirmed we had run the numbers correctly. What's more, they confirmed the crime surges we discovered were "significant events" that obviously had begun before Hurricane Katrina struck. They concluded that some of the more serious forms of violent crime were indeed on the rise in Houston.

Finally, we took our findings to the Houston Police Department. We presented our analysis to the department's top public information officer, who shared our numbers with the department's crime analysis division.

"They don't have any heartburn over it," Ready said. "So, I would agree it is probably a fair reflection."

The police remained steadfast in telling us that violent crime rates were down if you used overall numbers and considered population. Ready told us he thought his department already knew about all of the hot spots we discovered. He did not provide any documentation when we asked.

With that said, we were finally ready to reveal the city's new hot spots for violent crime. If Houston police knew about them, they certainly had not told the public yet.

In addition, whether or not they attribute the action to our story, the city of Houston recently announced a new task force intended to curb the rising trend in robberies of gas stations throughout Houston.

Online information

We knew we would never have enough on-air time to tell our viewers about crime trends in every neighborhood throughout the fourth largest city in America. We went to our Internet staff and asked them to help us create an interactive map that people could use to zoom in on their own neighborhood.

It was the first time we had ever tried to take detailed maps and make them searchable online to a hyper-local level. Although other news organizations had tackled similar projects using expensive software, we didn't have time to buy it or train our Web staff.

Instead, we came up with a quick and easy way to bring this information to the public that any news organization could use. We exported our maps to some relatively simple PDF files. Our Web staff then took those files and created a simple but extremely valuable way for our viewers to zoom in to see crime incidents along the very streets they live on. They did this by using Macromedia Dreamweaver, software nearly every Web design staff has. Our staff highlighted certain sections of a larger citywide PDF file we provided to them and linked different "hot spot zones" from the larger map to much more detailed PDFs that showed incidents street-by-street in those areas. All our viewers had to do was click on a zone, and they could get to the more detailed maps.

Almost immediately, we found out just how hungry our viewers were for local information like this. Our Web traffic for these searchable maps outpaced the number of hits a standard news story receives by more than 1,000 percent. Viewer after viewer wrote in to thank us for both the on-air report and for helping them to learn more about their own areas online.

We gave our viewers the tools and information they would need to make up their own mind about crime trends in their own neighborhood. What's more, we had expanded into a new era of how KHOU would communicate news and information online.

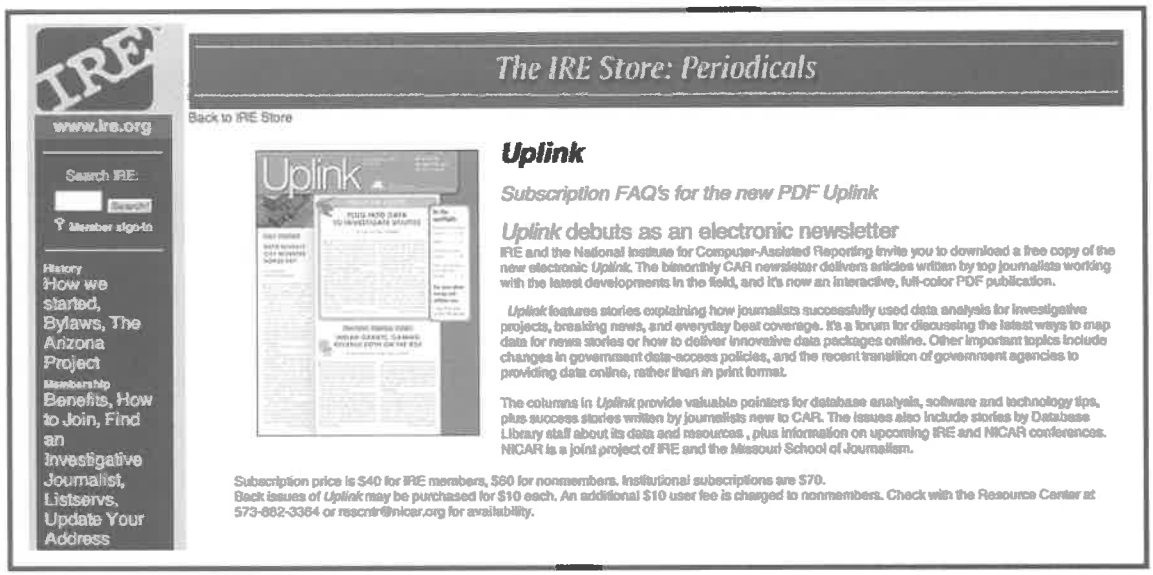
Contact Mark Greenblatt at
MGreenblatt@KHOU.com

Electronic Uplink takes off!

The March-April 2007 issue of *Uplink* will be the last issue printed on paper and mailed to subscribers. The new Uplink will be delivered as a PDF newsletter, bringing you the benefits of searchable text, links to resources from IRE and the Web, color graphics and more.

To celebrate our redesign, we're offering a free preview of the first two issues for a limited time at *www.ire.org/store/periodicals.html#uplink*. Please share the news with colleagues interested in computer-assisted reporting.

Subscribers can look forward to an *Uplink* with extra features in the coming year, such as news updates and tips in between the regular bimonthly publishing schedule, plus new multimedia offerings. You'll also receive access to online *Uplink* archives, beginning with the January-February 2007 issue, and an end-of-year compilation on CD-ROM. Our staff welcomes your feedback on the changes as well as suggestions for new content. You can email comments to *Uplink@nicar.org*.



Here's how future *Uplink* delivery will work:

If you are an IRE member with an *Uplink* subscription, you can access the PDF downloads by logging in on the IRE Web site. After you enter your last name and IRE member identification number, the "Contents" menu on the member login page should display a "Read the Latest Version of *Uplink*" option. (This system is currently activated so subscribers can test their access.)

If you subscribe but are NOT an IRE member, you will receive a letter containing a special login and password, plus instructions for downloading *Uplink*. **YOU MUST PROVIDE AN E-MAIL ADDRESS** to receive Uplink notifications and updates.

Information on the new *Uplink* and subscriptions is also on the Web at *www.ire.org/store/periodicals.html#uplink*. If you have problems accessing the electronic publication, or need to request a password for your subscription, please contact John Green at *jgreen@ire.org*.

FIREFOX EXTENSIONS: A GEEK'S NEW BEST FRIEND

By Neil Reisner
Florida International University

It's only a little more than two years since The Mozilla Foundation released Firefox 1.0. The new open-source browser on the block quickly became the new browser on the block, playing David to Internet Explorer's Goliath.

In January, *Information Week* reported that Internet research firm WebSideStory measured Firefox's share of the browser market at 14 percent. That includes a large following within the computer-assisted reporting community. Users are attracted by the reputation Firefox quickly earned as more stable and secure than IE and loaded with features that IE didn't have.

What even sophisticated users may not know is that the open-source community quickly began to develop more than 2,000 (and counting) Firefox extensions or add-ons – little programs that add functionality, including some that can make dealing with online data easily.

Here are examples of just a few, some enormously useful for computer-assisted reporting and some just useful and/or cool. Find Firefox's download page at *www.firefox.com* and its extension portal *addons.mozilla.org/firefox/extensions*.

But remember: *caveat emptor*. Read the comments on each extension's home page, and don't start playing with them on critical data. Note that some extensions are not yet updated for Firefox 2.0. The extensions are not in any way warranted by Mozilla. Although I have used or played around with most, but not all, of the extensions discussed here, I, too, offer no warranty.

Most extensions do exactly what they claim, but sometimes they don't. They even have been known to crash the browser. In the unlikely event that this happens, see the end of this article for instructions on starting Firefox in safe mode. Safe mode loads Firefox without any extensions. Then it's a simple matter to delete the offending program.

Some of these descriptions are from the text on the Firefox extension site; some of the screen shots are also from the site.

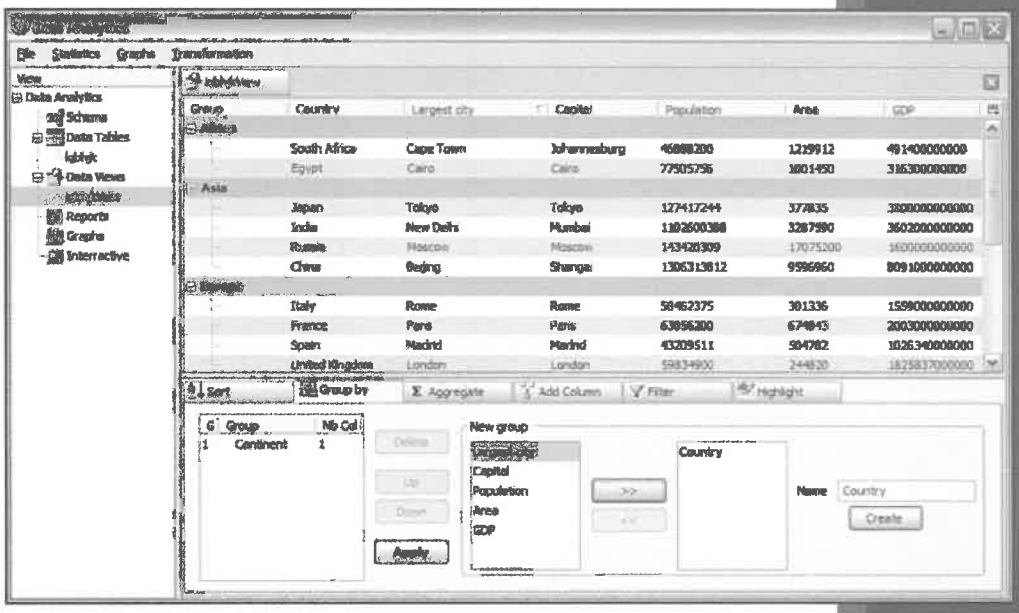
continued on page 20

CAR Picks

Data Analytics

https://addons.mozilla.org/firefox/2010

Allows you to easily import data locked in static HTML tables and then run basic statistics and makes it easy to analyze, graph and generate reports on the results. Runs descriptive statistics, T-tests, Anova and more. Not tested by author.



Data Analytics at work.

Table Tools

https://addons.mozilla.org/firefox/2637

This extension can sort, filter or copy HTML tables. You can search columns using a select filter, which uses a drop-down menu to choose values from a column. The second option is a search filter, which allows you to enter a value, keyword(s) or JavaScript regular expressions. Sorting and select filtering are smart because they auto-detect data types such as date/time stamps, numbers (including scientific notations), major currencies, IP addresses or plain text.

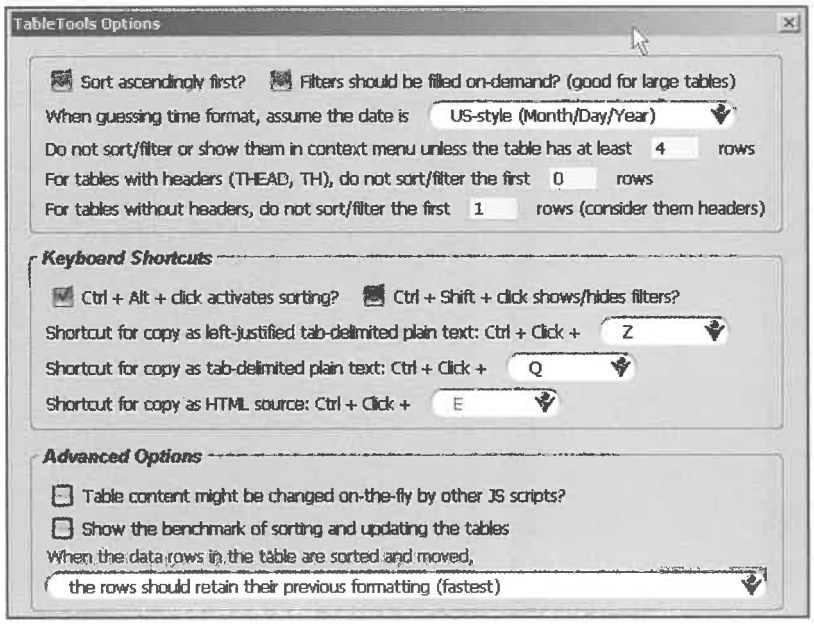


Table Tools option menu.

CAR TOOL

Firefox

continued from page 19

Table2Clipboard

https://addons.mozilla.org/firefox/1852

Copies HTML tables to clipboard as plain text in CSV, tab-delimited and other formats and allows you to paste it into the spreadsheet of your choice.

Copy Plain Text

https://addons.mozilla.org/firefox/134

Copies text on HTML pages without formatting. Set it to trim space in and around the copied text, replace multiple spaces with a single space, remove extra empty lines or convert two or more empty lines into a single line. This feature can be set as Firefox's default copy action.

SQL

https://addons.mozilla.org/firefox/409

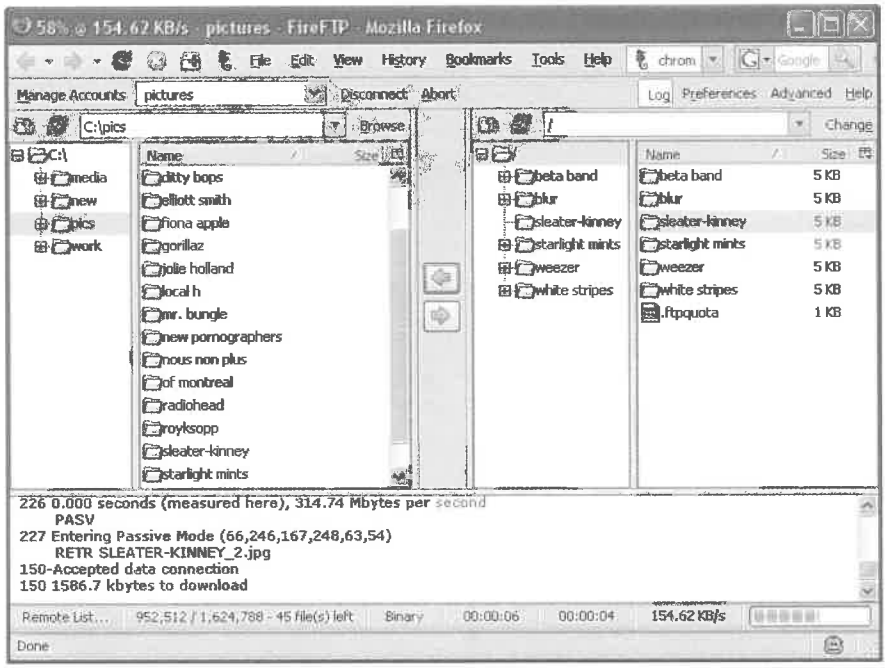
Developed by The Mozilla Foundation to allow applications to directly connect to SQL databases. The online description says: "A web application no longer needs to pass information through a scripting language, such as Perl or Python, in order to receive information it can use. The removal of the layer separating applications and data simplifies the job of the programmer."

Just Plain Cool

FireFTP

https://addons.mozilla.org/firefox/684

A full-featured FTP client that works within Firefox.



FireFTP turns Firefox into a full-featured FTP client.

Gmail Space

https://addons.mozilla.org/firefox/1593

This extension allows you to use your Gmail Space (2.5 GB and growing) for file storage. It acts as an online drive, so you can upload files from your hard drive and access them from every Internet capable system. The interface will make your Gmail account look like a FTP host. After the installation, you'll get an option called GSpace in your Tools menu.

FoxClocks

https://addons.mozilla.org/firefox/1117

My sister lives in Japan, and I can never keep track of the time there. I kept waking her with middle-of-the-night calls until I found FoxClocks. It puts a small clock on Firefox's status bar and keeps track of time around the world. My sister is much happier.

FoxClocks lets friends and relatives living abroad get their sleep by keeping track of what time it is throughout the world.

CAR TOOL

TinyURL Creator
https://addons.mozilla.org/firefox/126
This tool takes long, unwieldy URLs and gives you a short URL to replace it. It works just like *http://tinyurl.com*, but it's built into your browser.

Foxmarks Bookmark Synchronizer
https://addons.mozilla.org/firefox/2410
You can synchronize bookmarks on multiple computers, but the extension must be installed on each computer. It also makes your bookmarks available at *my.foxmarks.com*.

PDF Download
https://addons.mozilla.org/firefox/636
This gives you options for PDF documents: view in Firefox as PDF or HTML, open it in a PDF reader or editor or download the file.

Tab Mix Plus
https://addons.mozilla.org/firefox/1122
It includes such features as duplicating tabs, controlling tab focus, tab clicking options, undo closed tabs and windows, plus much more. It also includes a full-featured session manager with crash recovery that can save and restore combinations of opened tabs and windows.

Down Them All
https://addons.mozilla.org/firefox/201
This extension, mentioned recently on the NICAR-L Listserv, claims to increase download speeds by up to 400 percent. It also lets you to stop and restart a download, select specific features (such as images or links) from a Web page and perform batch download.

If Things Go Wrong – Starting Firefox in Safe Mode
kb.mozillazine.org/Safe_mode

Windows
(This should not be confused with restarting the computer in Windows Safe Mode!)
Find the Mozilla Firefox folder in the Windows Start Menu and choose the "Mozilla Firefox (Safe Mode)" or "Mozilla Thunderbird (safe mode)" shortcut. If you have accidentally deleted this shortcut, you can start Firefox in Safe Mode by going to "Start -> Run" and entering (using Firefox as an example):

```
"C:\Program Files\Mozilla Firefox\firefox.exe" -safe-mode
```

Linux
On Linux, you can go to your Terminal and run (for Firefox):

```
/path/to/firefox/firefox -safe-mode
```

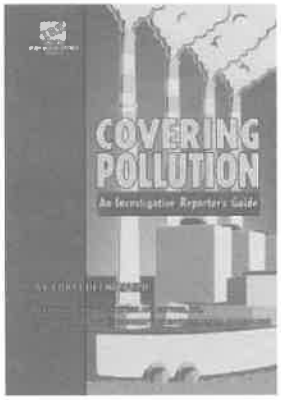
Mac OS X
On Mac OS X, go to Utilities (in the Applications folder) and open Terminal, then run (for Firefox):

```
/Applications/Firefox.app/Contents/MacOS/firefox -safe-mode
```

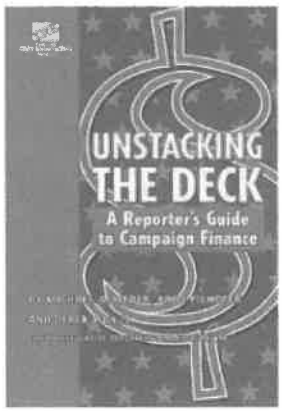
If you have installed the application to another location, modify the path as such. It's the "-safe-mode" command line parameter that's crucial here.

Contact Neil Reisner at nreisner@gmail.com

REQUIRED READING FOR YOUR NEWSROOM



COVERING POLLUTION
An Investigative Reporter's Guide
This primer gives an overview of useful resources reporters can use for local investigations into environmental pollution. Its main focus is to show how to get to the heart of an investigation quickly and without waiting months for FOI requests to be fulfilled. Filled with examples and references to stories, tipsheets and other resources available from IRE and SEJ.



UNSTACKING THE DECK:
A Reporter's Guide To Campaign Finance
Invaluable for pursuing stories about the impact of money on elections, political parties and candidates at the federal, state and local levels.
- Understand the loopholes in soft money restrictions.
- Learn about the use of nonprofits to funnel money to candidates.
- Find out how to track where candidates spend the money they raise.
- Learn how to obtain and use pertinent documents and electronic data.
- Find story examples from the IRE Resource Center.

... and much more

IRE MEMBERS: $15 • NONMEMBERS: $25 EACH • PLUS POSTAGE

POSTAGE RATES: Media Rate - $4 for first book, $2 for each additional book

BY MAIL
Send your check to IRE
138 Neff Annex
Missouri School of Journalism
Columbia, Mo. 65211

BY PHONE
Call
573-882-3364
with your VISA or MasterCard

BY WEB
Visit
www.ire.org/store
to order online or download an order form.

VER 1.0 SCRUTINIZES DATA QUALITY

By Matt Wynn, IRE and NICAR

Dirty data. It's an uncomfortable fact of life.

Maybe it comes in the form of a field with twice as many values as the look-up says it should have or a 15-year-old listed as being born in 1901. Regardless, it's there, and it's something computer-assisted reporting junkies have to deal with.

Last April, the Institute for Analytic Journalism held a workshop to try to address the growing challenge of working with all of these data. Twenty journalists, social scientists, computer scientists, educators, public administrators and GIS specialists met for three days to consider the question: "How can we verify data in public records databases?"

The Institute was created in 2001 by Steven Ross of Columbia University and J.T. Johnson, a professor emeritus at San Francisco State University. In his 35-year journalism career, he regularly moved from the classroom to the newsroom and back. The institute attempts to bring an awareness of methodological tools and skills to journalism and journalism education.

The book "Ver 1.0 Proceedings" (ed. by J.T. Johnson) published by the institute, captures some of the progress that was made during the April workshop. It contains papers, slideshows and presentations by journalists and social scientists discussing issues of database creation and management and what can be done to verify those databases.

"Now, journalists will be able to ask some pointed questions to bring to the discussion of how we validate public records and data," Johnson said in a recent phone interview.

The book includes "A Critical Biography of a Federal Database," by Paul Overberg of *USA Today*. In the paper, he explores the circumstances that led to the creation of the Environmental Protection Agency's Toxic Release Inventory, as well as factors that have affected how the data is maintained today. He talks about the process of creating the list of chemicals TRI would track and the immediate results of making the data public. (This database is available from the IRE and NICAR Database Library, *www.ire.org/datalibrary/databases/viewdatabase.php?dbaseindex=55*).

Overberg summarizes the importance of such an in-depth analysis: "An examination of the social and political forces moving during and even before the creation of the database provide important insight. The database's design, inclusiveness, exclusions, scope are all shaped by details in congressional reports, federal legislation and regulations, lobbying and advocacy."

Marcus Messner and Bruce Garrison from the University of Miami attempted to find discussion of dirty data in existing journalism literature and were left unimpressed. Only a handful of authors, they discovered, had attempted to summarize the errors journalists face in databases. A similarly small number had written about techniques useful in cleaning data.

Johnson said he hopes "Ver 1.0" will help convince others to fill that void.

"We hope that there will be some journalism educators who will look at this and say this is important and try to develop some tools that we can then put into play," he said.

The sheer volume of publicly held databases has made the role of journalists in vetting the data extremely important, Johnson said. Public and private decision-makers draw on the reserves of information to make decisions on everything from disciplining doctors to education to healthcare to voting registration and law enforcement.

That's a lot for journalists to handle. The book's appendix includes two stories that highlight the dangers of incomplete or inaccurate data. The first is a *St. Louis Post-Dispatch* story from 1999 about 700 registered sex offenders who do not live at their listed address, and the second is a *Dallas Morning News* story about errors in the state's conviction database.

An examination of the social and political forces moving during and even before the creation of the database provide important insight. The database's design, inclusiveness, exclusions, scope are all shaped by details in congressional reports, federal legislation and regulations, lobbying and advocacy.

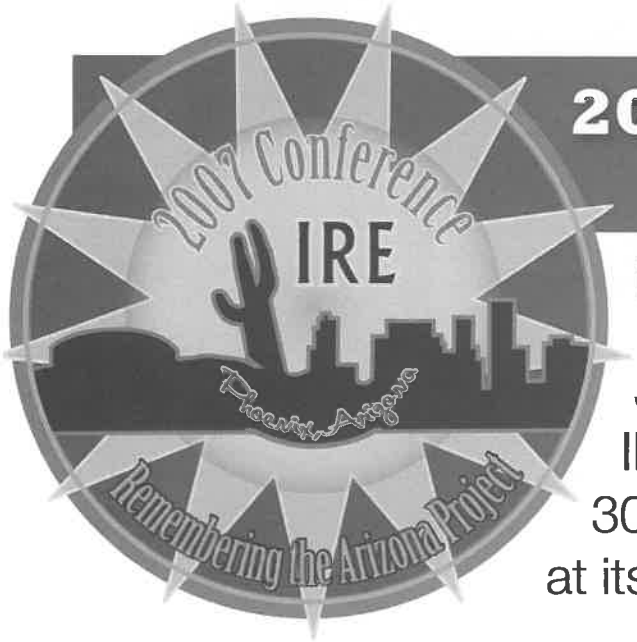
"You talk about watchdog journalism... this is exactly what we should be doing. Saying, 'Look, you're making these decisions, and your data is wrong,' Johnson said.

The most frightening part, perhaps, is that the problem isn't getting any better, said Johnson. Datasets are growing larger and larger, and journalists who may have learned to clean datasets of 1,000 or even several hundred thousand records find themselves unprepared to deal with the realities of so-called "monster" datasets, those that hold up to 100 GB of data.

Journalists probably haven't seen anything like it yet, but it's a problem they may soon have to address. And when they do, it's important they know what they're dealing with, and how best to verify what's there.

The book is available through the Institute's Web site, *www.analyticjournalism.org*.

Contact Matt Wynn at
mwynn@ire.org



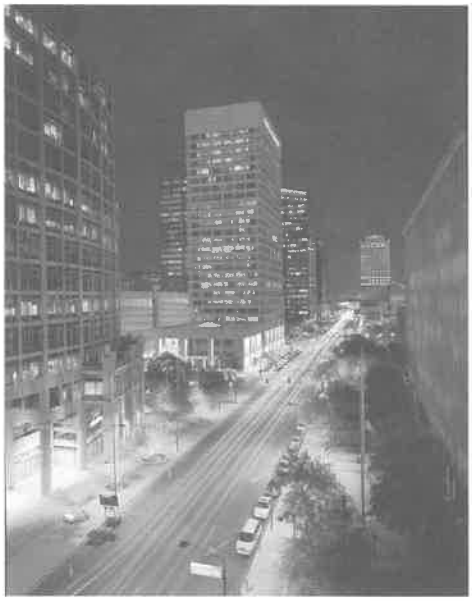
2007 IRE Conference • June 7-10 •

IRE

HOST: The Arizona Republic

Join the best in the business as IRE honors Don Bolles and the 30th anniversary of the Arizona Project at its annual conference.

The Arizona Project was the historic effort by journalists to continue the investigation that led to the killing of Arizona Republic reporter Don Bolles, a 47-year-old husband and father. Bolles and many of those 38 journalists, known as the Desert Rats, were founding members of IRE and were instrumental in shaping its future.



For IRE, the resulting Arizona Project brought national attention and stature. A tiny organization with little money flourished to become what it is today. Thanks to those who have gone before, IRE now has an organization strong enough to take on today's threats to investigative reporting.

Join us in Phoenix for:

- Panels with tips and techniques from top reporters, producers, editors, news directors and writers
- Hands-on training in computer-assisted reporting from the best practitioners
- Networking and mentoring opportunities

Visit www.ire.org/training/phoenix07 for more details, conference registration and hotel reservations.

IRE AND NICAR SERVICES

Investigative Reporters and Editors, Inc. is a grassroots nonprofit organization dedicated to improving the quality of investigative reporting within the field of journalism. IRE was formed in 1975 with the intent of creating a networking tool and a forum in which journalists from across the country could raise questions and exchange ideas. IRE provides educational services to reporters, editors and others interested in investigative reporting and works to maintain high professional standards.

The National Institute for Computer-Assisted Reporting, NICAR, is a joint program of Investigative Reporters and Editors Inc. and the Missouri School of Journalism. NICAR services include supplying journalists with government databases, training programs, tipsheets and data analysis.

Programs and Services

IRE Resource Center: A rich reserve of print and broadcast stories, tipsheets and guides to help you start and complete the best work of your career. This unique library is the starting point of any piece you're working on. You can search through abstracts of more than 20,000 investigative reporting stories through our Web site, www.ire.org/resoucecenter.
Contact: Beth Kopine,
beth@ire.org, 573-882-3364

Database Library: Administered by IRE and NICAR. The library has copies of many government databases, and makes them available to news organizations at or below actual cost. Analysis services are available on these databases, as is help in deciphering records you obtain yourself.
Contact: Jeff Porter,
jeff@ire.org, 573-882-1982

Campaign Finance Information Center: Administered by IRE and NICAR. It's dedicated to helping journalists uncover the campaign money trail. State campaign finance data is collected from across the nation, cleaned and made available to journalists. A search engine allows reporters to track political cash flow across several states in federal and state races.
Contact: Brant Houston,
brant@ire.org, 573-882-2042

On-the-Road Training: As a top promoter of journalism education, IRE offers loads of training opportunities throughout the year. Possibilities range from national conferences and regional workshops to weeklong boot camps and on-site newsroom training. Costs are on a sliding scale and fellowships are available to many of the events.
Contact: David Donald,
ddonald@ire.org, 573-882-2042

Publications

The IRE Journal: Published six times a year. Contains journalist profiles, how-to stories, reviews, investigative ideas and backgrounding tips. *The Journal* also provides members with the latest news on upcoming events and training opportunities from IRE and NICAR.
Contact: Brant Houston,
brant@ire.org, 573-882-2042

Uplink: Bimonthly newsletter by IRE and NICAR on computer-assisted reporting. Often, *Uplink* stories are written after reporters have had particular success using data to investigate stories. The columns include valuable information on advanced database techniques as well as success stories written by newly trained CAR reporters.
Contact: David Herzog,
dherzog@nicar.org, 573-882-2127

Reporter.org: A collection of Web-based resources for journalists, journalism educators and others. Discounted Web hosting and services such as mailing list management and site development are provided to other nonprofit journalism organizations.
Contact: Brant Houston,
brant@ire.org, 573-882-2042

For information on:

Conferences and Boot Camps:
Ev Ruch-Graham,
ev@ire.org, 573-882-8969

Listservs: Matthew Dickinson,
matthew@ire.org, 573-884-7321

Investigative Reporters and Editors, Inc.
The National Institute for Computer-Assisted Reporting
138 Neff Annex
Missouri School of Journalism
Columbia, MO 65211

Editor:
Brant Houston
brant@ire.org

Managing Editor:
David Herzog
dherzog@nicar.org

Asst. Managing Editor:
Jeff Porter
jeff@nicar.org

Senior Editors:
Sarah Cohen
Stephen K. Doig

Art Director:
Wendy Gray

Publications Coordinator & Advertising Contact:
Megan Means
meganm@ire.org
573-884-2360

Contributing Editors:
Alex Richards
Rebecca Townsend
David Walle
Matt Wynn

Editorial:
573-884-7711

Subscriptions:
IRE members $40
nonmembers $60
Contact: John Green
jgreen@ire.org
573-882-2772

Uplink Address
IRE-NICAR, 138 Neff Annex
Missouri School of Journalism
Columbia, MO 65211