

Uplink

July 1996

A newsletter for the National Institute for Computer-Assisted Reporting

Wish you were here Uplink update

Ola, bonjour, NPUBET, goddag and hallo. In other words, hello.

In IRE and NICAR's ongoing effort to reach international journalists, we dedicate this issue of Uplink to resources useful to reporters at home and abroad. Marjan Agerbeek of The Netherlands shares her first CAR experience. Mike Gordon of Canada tells us how he won a battle for data. Freelance reporter Gwen Carleton interviews Danish journalists about their battle for data access.

Natalya Shulyakovskaya reviews the Russian Sources for Reporters homepage, part of the Hillman Project, along with other Web sites that have useful international information.

And IRE Executive Director Rosemary Armao brings us up to date on Periodistas de Investigacion, a project made possible by the McCormick Foundation to start an IRE-like organization in Mexico.

Inside

Danes battle for data
Page 3
Canada explores CAR
Page 4
Tech tip
Page 8
On the Internet
Page 10
Bits, Bytes and Barks
Page 12

Mexico institute revs up ¿Qué pasa?

By Rosemary Armao
IRE executive director

Periodistas de Investigacion means "Investigative Journalists" in Spanish. It also means IRE is going international for the first time in a big and systematic way.

It's risky, we know. An IRE/NICAR organization may not work for our Mexican colleagues. Our members could object to organization time spent on international networking when reporters here have so many problems.

Such arguments seem familiar. Three years ago, IRE, with a good idea, a potful of money from the Freedom Forum, and a great deal of trepidation, set out to do something it had never done in its then 17-year history. The result of that venture, as we all know now, is NICAR, Uplink, thousands of journalists trained, heightened attention to the need for computer-assisted reporting at all media outlets and mounting pressure for laws that open electronic records.

Here we go again

So, here we go again, with NICAR as our model. As you read this, Lise Olsen, formerly a database specialist for *The Virginian-Pilot* in Norfolk, Va., a frequent IRE/NICAR speaker and a trilingual reporter, is in Mexico City renting an office and hiring staff for Periodistas de Investigacion. Olsen will be the managing director and her major hire will be a Mexican journalist who will work with her as training director of the new organization. In three years, if this works as we expect, the training director will take

over as the boss, Periodistas will be a wholly Mexican organization, and IRE could be working with colleagues in another country.

Olsen, 31, is a graduate of the University of Nebraska at Lincoln. She's been at *The Virginian-Pilot* since 1988, except for a year she took off in 1993-94 to travel to Argentina, studying Spanish and writing freelance articles. Since

Continued on page seven

Fishing in international waters

Reeling in Web data

By Natalya Shulyakovskaya
St. Louis Post-Dispatch

There are a few problems with getting international data electronically: It's spotty and not too many countries collect it themselves—let alone put it on the Web. So, we have to rely on either international organizations or U.S. sources to collect our beloved numbers.

My attempt to put together information on some international data and sources of information resulted in Russian Sources for Reporters, part of the Hillman Project, at www.reporter.com/hillman/russian/

The pages are the seed of a plan to develop an international tip sheet and

Continued on page two

Uplink

July 1996

Volume 8, Number 7
A newsletter for the
National Institute for
Computer-Assisted
Reporting

Editor

Brant Houston

brant@nicar.org

Managing Editor

Jo Craven

jo@nicar.org

Assistant Managing Editor

Andrew Lehren

andy@nicar.org

Senior Contributing Editors

Richard Mullins

richard@nicar.org

Rosemary Armao

jourca@nuccmail.missouri.edu

Staff

Wendy Charron

Seth Hemmelgam

Neil Reisner

John Sullivan

Wallace Winfrey

Copy Editor

Marilyn Lehren

Uplink is published every month by the National Institute for Computer-Assisted Reporting, 138 Neff Hall Annex Columbia, MO 65211. (573) 882-0684.

Subscription price is \$35 (U.S.) annually, \$50 outside the United States.

Postmaster: Please send address changes to NICAR.

Send e-mail to

nicar@nuccmail.missouri.edu.

NICAR is a joint effort of Investigative Reporters and Editors and the University of Missouri School of Journalism.

NICAR services include hands-on newsroom training on computer-assisted reporting, special academic and advanced training in data analysis. NICAR is supported by a grant from The Freedom Forum and other foundations intended to help the institute deliver its services nationwide to news organizations and associations.

Continued from page one:

Fishing internationally

library of resources about Russia for English-speaking reporters.

The second half of the project is to develop a collection of resources in the Russian language for Russian journalists writing about the United States.

A third goal is to provide information about investigative and computer-assisted reporting in Russian.

Since most international data cover all foreign countries, these pages are not limited to information about the new states of Eastern Europe. I couldn't possibly find and list all available information, but there are listservs and other human resources that you can use to dig further.

In the meantime, here are some useful sites:

- If you are on a fishing expedition, the good old Census Bureau is the first place to check. Its International Program Center (www.census.gov/ipc/www/) lists several databases covering foreign countries.

My favorite is the International Database (IDB), which covers a full bouquet of demographic and socio-economic data. It goes into such details as prevalence of contraceptives used in different countries and number of babies born to women of different age groups (look out for those teen-agers).

A word of warning: Be careful when talking to bureaucrats about international data: One well-meaning woman at the Department of Commerce tried to sell me IDB for \$40 while the data was sitting nicely on the department's site.

- The Census Foreign Trade Division (www.census.gov/foreign-trade/www/) is a good place to pick up statistics on international trade and U.S. trade surpluses and deficits with its top trading partners. Census also has data on immigration and numbers of foreign-born U.S. population.

- The State department dishes out some really good reports (skip the travel advisories) (http://www.state.gov/www/about_state/dos_welcome.html, bottom of the page). One is on terrorism patterns in 1995, another one is

on country background and international organizations reports, and another on U.S. Assistance to the New Independent States for FY 1996.

- The CIA's World Fact Book is another good source. The CIA compiles this every year. It describes each country's political system, major players, land area, environmental laws, life expectancy and other crucial bits of information in flat text. The CIA has it in a database, but does not release it as such. Find it and other CIA publications at its homepage (<http://www.odci.gov/cia>).

- Another great source for international data is the World Health Organization (www.who.ch or gopher://gopher.who.ch:70/11/). A list of the WHO's databases and links to them are at <http://www.who.ch/whosis/whosis.htm#databases>. Most of the data come in fixed-width text format. World Health Organization — Epidemic Incidents is a great source of data on outbreaks of tuberculosis, measles, etc., in different countries. Data are in text format, tabular, fixed length, with some text mixed in.

- UN Criminal Justice Country Profiles (gopher://uacsc2.albany.edu:70/11/newman/crjdoc) is another valuable source, especially with the crime and prison explosions that nations around the world are experiencing. It describes nations and their criminal justice and prison systems. See the State Department's reports on human rights practices in different countries, too.

- Agricultural/Trade Databases from USDA (gopher://mann77.mannlib.cornell.edu:70/11/data-sets/international) is a good source for agro-economic information. Data comes from the USDA's Economic Research Service. You can download it, but the documentation also tells how to order data on a disk. Tables are readable in Excel. Russian commodity prices are there, as well as a bunch of European and African data.

Most of the pages containing data description (some systems refer to it as metadata) also list names, phone numbers and e-mail addresses of researchers who collect the information. I didn't have much luck e-mailing them, though.

Natalya Shulyakovskaya can be reached at (800) 365-0820 ext. 8183, or send e-mail to natalya@pd.stlnet.com

Danes battle for access

By Gwen Carleton
Freelancer

Does it ever seem all of modern bureaucracy is standing between you and the database you desire?

Take heart. At least you don't live in Denmark.

The country known for Hans Christian Andersen, big government and interesting furniture has another claim to fame: It has some of the most restrictive data access laws in northern Europe.

In Denmark, obtaining government data in print form works like this: Journalists interested in a particular issue approach the relevant government agency, describe the issue and, hopefully, receive information, explained Nils Mulvad and Flemming Svith, two Danish journalists who attended NICAR's March bootcamp.

No fishing allowed

"In Denmark, you can only get access to information if you know quite a bit about the issue already," Mulvad said. In other words, no fishing expeditions allowed.

Once the journalist and bureaucrat agree upon the issue under discussion, the journalist must be concise. In the register of deeds' office, for example, "you may only ask three questions, then you must come back the next day," Mulvad said.

It is not surprising then that electronic information is strictly off limits. But Mulvad and Svith, reporters for *Jyllands-Posten*, Denmark's largest national daily, are working to open up the country's trove of public data.

The task has not been easy. Mulvad and Svith are part of a group of journalists, editors and journalism educators that has been negotiating with government officials for months. But in February, they reached a breakthrough: The government agreed to consider opening electronic indices of files, mail and other official documents to journalists.

Such access — which includes only lists of documents, not the documents themselves — is likely to begin this fall, Mulvad said. "It was really exciting after that meeting. Before then, it was like the conversation was taking place in two different rooms."

The prospect of improved openness is exciting for Danish journalists because of the sheer volume of information gathered and archived by the government. The educational system, health care and the national transportation system are just a few of the powerful industries operated by the Danish government.

The country's generous social welfare system requires further government record keeping. Some citizens appear in government databases because of what they receive, such as unemployment insurance or student stipends. Others appear because of what they owe, such as taxes or time in the national penal system. But nearly all Danish citizens appear in at least one database, and most appear in several.

Public opposition

Danes are acutely aware of how much information the government holds about them. As a result, public opinion is squarely against increasing access to electronic databases.

Danish journalists have a few helpful online resources, including Profile, a service similar to Dialog, and Pol Info, which offers newspaper articles, a database of female sources and other sources. Mulvad and Svith said the Internet is a promising source; however, with few Danish sites and almost no government sites, its uses are still limited.

Still, there have been successes. Since *Jyllands-Posten* installed PCs last year, Mulvad and Svith have purchased Excel, Access, Paradox and SPSS. They created a database of sources for the newsroom after a big story left the staff scrambling for names and telephone numbers. Spreadsheets helped out with a story about debt and price-gouging in the national telephone industry. When Danish physicians began encouraging young women to take large doses of folic acid, Mulvad and Svith used e-mail and listservs to investigate other countries' policies.

As for their biggest challenge, gaining access to government data, Mulvad said he is cautiously optimistic.

"We're testing the borders to see which files we can get out," Mulvad said. "At least we'll get something."

Gwen Carleton can be reached at (608) 245-0607, or send e-mail to gwen@nicar.org

For a taste of Danish, visit the *Jyllands-Posten* at www.jp.dk/. The site, which is updated daily, presents a thorough cross-section of the day's news, from domestic and international events to business, sports and features. Visitors can read story summaries for free, or pay a fee and gain access to the full articles.

Dueling for Canadian data

By Mike Gordon
TV Ontario

In August 1994, I was an intern reporter with the *Montreal Gazette* when I requested a database of construction permits from the City of Montreal. I was strung along for a few months before filing a formal access to information request in the first week of December 1995. The final word from the city's access commissioner came in February 1995: "No way, Jose."

So, off I went to the Commission d'accès à l'information (du Québec). Backlogs in the commission's system meant that I was not offered a hearing until mid-October.

In the meantime, I had moved from Montreal to Toronto to start my current job. The commission didn't find me at my old address, and I did not hear about my date until the day before. I couldn't go, so my date was rescheduled to Dec. 8, 1995.

The decision

I waited and waited for the decision. Finally, on Feb. 7, I received a certified mail envelope with the decision. Nervously, I flipped to the end of the four-page decision. Here, I read the wonderful words (for those of you who read French):

"POUR CES MOTIFS: La Commission ACCUEILLE la demande de révision et ORDONNE l'organisme de communiquer au demandeur la base des données qui contient les demandes pour les permis de construction sous forme sortie informatique."

That means I won, and the city was ordered to give me the database.

The issues

For those of you who are interested in the legal issues:

The city argued that while they accepted the information I wanted was public information, the structure and organization of the database was such that protected information that I did not seek was mixed up in the same file matrix. They argued that to extract the information I sought would require them to create a new document, a software program. Article 15 of the Access to Information Act says no organization has to create a new document. I argued that in no way was the programming required

to manipulate a database included in the sense of "a new document" as mentioned in the Act. The judge ruled in my favor.

She wrote:

"Avec gard, je ne crois pas qu'il s'agisse d'un cas de l'application de l'article 15 de la loi. Les renseignements que desirait le demandeur existent. Aucun calcul ni comparaison n'est nécessaire au sens de l'article 15 de la loi. Il ne s'agit pas de la création d'un nouveau document mais d'une simple manipulation d'un programme informatique."

Basically, what she said is that the information exists; no calculation or comparison of data is required to create the information I wanted. This was not a case of creating a new document, but the simple manipulation of a computer program.

Continued on page five

Database of Databases

Canada explores CAR

By Robin Rowland

Ryerson School of Journalism, Canada

The Database of Databases project (<http://www.ryerson.ca/~ricarc/pchome.htm>) is aimed at jump-starting computer-assisted reporting in Canada. Visitors to the site will find lists that may be searched by agency, database or city. Searches reveal, among other things, an agency's address, telephone number, whether the agency maintains a database, and information on access.

This ongoing project aims to:

- Provide lists of government, association, foundation and private sector databases not only to journalists but to the academic community as well.
- Complete the "Canadianization" of CAR exercises for Canadian students.
- Begin the liaison between the journalistic community and the people who run the

Continued on page five

The Quebec Launch
Pad for Journalists at

<http://>

www.CAM.ORG/

~andymon features
links to government
departments,
sources, and up-to-
the-minute news
from Quebec, as well
as dozens of other
useful links for
journalists who cover
or have an interest in
the Canadian
province. Most links
are in English and
French, but some are
in French only.

Continued from page four:

Canada explores CAR

databases, so they will understand the needs of journalists and other researchers.

It's estimated that 80 percent to 90 percent of government information will soon be on computer.

In the United States, much of this material is available to the public and the news media through open records laws. Other material is often pried from government through Freedom of Information requests.

In Canada, we are facing a number of problems.

- A tradition of greater government secrecy.
- Crown Copyright — information belongs to "the crown" rather than "the people."
- Higher costs. The government sees data, collected at taxpayers' expense, as a profit center in an age of cost-cutting.
- There have been only a few requests, so far, for data.

CAR reporters in the United States say you must start with a good story idea, then find the database. The Database of Databases will make this easier.

If a reporter has a story idea, he or she can then look up the Database of Databases and see if the data is available.

In addition, the simple knowledge that those databases exist could suggest story ideas to reporters as well as new, publishable research projects to academics.

How it started

For the past two years, students at Ryerson School of Journalism have completed an assignment on finding what's "out there."

Almost always, the bureaucrats have been astonished that anyone would call and ask for this information. Often they have refused to give more than bare-bones information to the students. And Canadian bureaucrats have been equally surprised that the United States generally makes this information easily available.

So far, the Database of Databases project has identified 60 databases used by federal, provincial and municipal governments, police forces, fire departments, industry and professional associations, research foundations and the private sector.

This is just scratching the surface. There is a lot more out there.

Objective

The objective will be to create a foundation for an ongoing Database of Databases project created by the Ryerson School of Journalism as a service to journalism and academia. Among other things, the project would like to:

- Identify as many databases across Canada as possible.
- Identify U.S. databases that contain Canadian data.
- Create a Database of Databases in the industry standard .dbf format, as well as comma separated values and AskSam formats.
- Identify and obtain databases that can be used for student training and research.
- Create a foundation for the project to continue using students as researchers. The database could then be updated on a regular basis.

Robin Rowland can be reached by e-mail at eridani@inforamp.net

Continued from page four:

Journalist wins data duel

So, I won a case. Maybe it'll be a useful precedent for other reporters in Canada.

In preparing for my case, I was helped by the CAR Caucus of the Canadian Association of Journalists, especially Robert Washburn of the Cobourg *Daily Star*. They helped me get some pro bono legal advice from the LaFleur, Brown law firm in Montreal. Charles Bury of *The*

Sherbrooke Record attended my hearing on behalf of the CAJ.

Hope this bucks up the spirits of those of you engaged in similar legal battles. It takes a long time, but it is possible to win.

Mike Gordon can be reached at (416) 484-2600, ext. 2534, or send e-mail to mike_gordon@two.org

Robin Rowland of the Ryerson Polytechnic University School of Journalism in Canada,

recommends a database called the "Public Information Dataset Manifest and Generator Registration

Information," which tracks all the hazardous waste generators, carriers and receivers in Canada. "If a U.S. company on the border trucks its waste across the border to Canada or vice versa, it is listed," Rowland says. "U.S. data is dirty. It lists 'USA' under province in some conditions."

To get the free database, send three high-density PC disks to: Ontario Ministry of Environment and Energy, Environmental Monitoring and Reporting Branch, Environment Information and Systems Section, 40 St. Clair Ave. West Floor 12, Toronto M4V 1M2

Excel goes Dutch

Foreign news editors
of the Japanese news

agency Kyodo

maintain a homepage

at [http://](http://ourworld.compuserve.com/homepages/tyoko/)

[ourworld.compuserve.com/](http://ourworld.compuserve.com/homepages/tyoko/)

[homepages/tyoko/](http://ourworld.compuserve.com/homepages/tyoko/)

E-mail comments on

the page to Tsukasa

Yokoyama at

tyoko@po.ijnet.or.jp

By Marjan Agerbeek
Trouw, The Netherlands

Marjan Agerbeek of Trouw, a daily in Holland, spent five weeks with The Philadelphia Inquirer CAR team last fall and five weeks at the Minneapolis Star-Tribune. She also attended the 1995 NICAR convention in Cleveland. Once back in Holland, Agerbeek completed her first CAR project, which was published in two parts, beginning on the newspaper's front page, and exposed poor financial planning on the part of the country's 13 cash-strapped universities.

Last January, when I came back from the United States, I was eager to start my first CAR project.

As an education reporter, I decided that the project should be about education. Another quickly made decision was that the project had to be about universities, because there are only 13 in the whole country. If necessary, 13 records could be imported by hand. And calculations with 13 columns could be checked with an old-fashioned calculator. I felt safe, knowing I could defeat bad luck, such as data that didn't want to be imported into my spreadsheet, or other dirty tricks software plays on clumsy beginners.

Finding the data

Each year, the Dutch education department publishes a book with data about universities. I looked it through and found three uncomplicated tables with information about the staff at every university, including how many employees have temporary contracts, how many have steady ones, the number of employees in the distinctive job levels — that kind of stuff.

These tables were potentially interesting because of political circumstances. At a national level, it was decided that the universities had to become more autonomous in their staff policy. And the university boards had to cut expenses. I wondered whether the universities could shrink without mass layoffs. I also wanted to know what kind of employment policy they had four years ago, when they also had to cut expenses.

My colleagues at the education desk agreed that these questions were worthwhile, so I started the project. I asked the information official at the department of education to send me the book with the university data on floppy disk. The older data I found in another book. These were not available in an electronic version, so I had to type those in.

Hitting paydirt

When the floppy arrived, the data appeared to be in WP 6.1 format. As there is no Windows-PC at the newsroom, I had to work at home, where I have installed Microsoft's Office Professional. I translated the data into Word, which was a piece of cake. And it turned out Excel had no problems with the tables. I was stunned about my luck and the intelligence of Excel.

I analyzed the employees' records of public universities with Excel and found the number of professors, who get the highest salaries, had been increasing, while the total number of teachers had been declining. I also found that this had nothing to do with selective layoffs. The boards had ignorantly hired more professors. So, while the boards were trying to deal with lack of money, they were spending an increasing amount of money on salaries.

I visited three universities to find explanations for the data. I knew which university faced mass layoffs. And I found universities that had cut expenses by using Ph.D. students as teachers, which is only allowed on a small scale.

Three weeks and counting

The project took three weeks. My colleagues found this ridiculously long and were pressing me to publish during the last two weeks. At Dutch newspapers, working three days on one story is considered long.

The stories were hard to read, because they mainly explained the data; I had no time to do some reporting. Nevertheless I'm satisfied. The project was a little CAR example in Dutch, which I needed to show the newsroom what CAR is capable of. The management of *Trouw* appears to be interested. There is a lack of money, but hopefully, by the end of this year, there will be a Windows-PC in the newsroom.

Marjan Agerbeek can be reached by e-mail at agerbeek@knoware.nl

Popular training

Bootcamps fill up fast

By Brant Houston

NICAR managing director

When the institute for computer-assisted reporting began at Missouri in 1989 (then named MICAR), the basic offering was a six-day immersion "bootcamp."

Under the direction of computer guru Elliot Jaspin (and later former executive IRE director Andrew Scott), the four bootcamps a year turned out dozens of CAR reporters who were pioneers in their newsrooms in the early '90s.

In early 1994, when the institute became NICAR, it began an ambitious outreach program of road training that has led to about 135 seminars in two years throughout the United States.

At the same time, the bootcamps have continued to thrive. With the increase in NICAR staff, we have been able to expand the average number of bootcamp participants from 12 to 20. This year, because of the demand, we added a fifth bootcamp and are trying to find a way to add specialized seminars in the coming year..

As result of the expansion, more than 225 print and broadcast journalists have completed the bootcamp since January 1994.

The curriculum at the bootcamps also has changed as the technology has changed. While we still teach database managers, spreadsheets, using 9-Track tapes, and getting and cleaning data, we have incorporated the Internet into the curriculum, showing journalists how to gather and analyze data from the Web sites.

We also have been working to make our curriculum more relevant to international journalists. In the past year, journalists from Denmark, Finland, Australia and England have signed up.

In the coming year, we are faced with the challenge of teaching various skill levels. We intend to keep the basic bootcamp, but we are already developing the materials for an intermediate-level bootcamp. In addition, we will again offer our advanced seminar at the University of North Carolina at Chapel Hill May 4-9, 1997, with CAR pioneer Philip Meyer.

If you are interested in attending bootcamp, the next one scheduled with remaining openings is for Jan. 5-10.

For more information, call NICAR at (573) 882-0684, or send e-mail to nicar@muccmail.missouri.edu

IRE, NICAR and The Richmond Times-Dispatch will offer the Computer-Assisted Reporting Workshop for Minority Journalists in Richmond, Va., Nov. 15-17.

The hands-on training will include navigating the Internet, spreadsheets, database managers, negotiating for data and more. Forty slots are available. For more information, call (573) 882-0684, or send e-mail to Wendy Charron at wendy@nicar.org.

Miss something? Check out the IRE-L and NICAR-L mailing list archives on our websites at <http://www.ire.org> and <http://www.nicar.org>. You can see posts to both lists organized by thread, author and date. The list archives are available in html or in plain text format.

Continued from page one:

Mexico institute revs up

1994, she has concentrated on securing, analyzing and developing stories from databases.

McCormick grant

The Mexico project has been made possible by a three-year \$540,000 grant from the McCormick Tribune Foundation.

With that money, IRE/NICAR hopes to offer low-cost or free training in newsrooms and at seminars across Mexico and to work at setting up some of the same components that have made IRE invaluable to its 3,500 members. That includes a library of investigative projects, a newsletter and other publications, a collection of databases and a homepage on the World Wide Web that reporters can turn to when they need help or advice on a story.

Computer-assisted reporting, particularly developing databases and on-line research, will be a highlight of the curriculum, which is still

being put together for Periodistas de Investigacion, by Olsen with help from NICAR.

Possibilities unfold

We foresee that Periodistas de Investigacion will widen the international network of journalists that IRE has helped nurture across the border. We expect the organization to foster bilateral investigations on such topics as immigration, drug trading and the environment. Student and intern exchanges also have been discussed. It's an exciting time as the possibilities begin to unfold. We welcome your ideas, your advice, your warnings and your experiences reporting in and on Mexico.

In the meantime, we are learning about whole other systems of taxation, banking, news gathering and ethics. We'll keep you informed.

Rosemary Armao can be reached at (573) 882-1982, or send e-mail to armao@ire.org

Untangling problem layouts ...

Drew Sullivan of the Associated Press recommends a research tool called NameBase, available for free at <http://www.pir.org> Namebase, run by Public Information Research, "is an index of names pulled from books, magazines, newsletters and all sorts of strange and rare sources dealing with the CIA, intelligence, corporate fraud, drug-dealing, the mob and more than a few conspiracy theories," Sullivan says. "If you can't remember what Thomas Clines or Richard Secord was involved in, this is your place." Drew Sullivan can be reached by e-mail at drew@ap.org

By Richard Mullins
NICAR staff

Getting new data is like getting a present. You want to get the data loaded into your PC database program as quickly as possible so you can start playing with it.

But this unwrapping process can be delayed by simple typing mistakes on our part, or, what's worse, there can be small or frustratingly substantial differences between the record layout we have on paper and the data we were given on tape.

Just like alarm clocks going off at 6:15 a.m., this kind of thing happens more often than we'd like. Here are some of the procedures I've come up with to deal with the problems that sometimes hold us up from the fun part: getting to interview data.

For the tape processing, I'm using Nine Track Express and assuming familiarity with its Create File Parameters and Create Field Descriptions steps. Additionally, these tips take advantage of Microsoft Excel's marvelous Text Import Wizard; a graphical word processor such as Microsoft Word; and a file viewer called List, a shareware program I find indispensable.

Testing format

Use .csv as the output format while testing. The Preview feature in Nine Track Express can save you a lot of time. I never process any tape without using it. While the Transfer File procedure is counting off the records moved from the tape to the PC, pressing the "P" key pauses the process and shows you exactly what Nine Track is reading as a single record. If the data at the beginning and at the end looks OK, then you know the record length you are using is correct. I always use the Preview several times to make sure the one record I saw wasn't coincidentally correct and most of the others were wrong.

Mistakes in field lengths are more frustrating. If you are outputting the data to the FoxPro or Paradox format, the preview will not show

you how the data is being split into fields. To find out if it's right, you have to quit Nine Track and look at the results in the database program. If it's wrong, that means quitting FoxPro or Paradox and going back to Nine Track. And if you're using the Windows version of these database managers, then you have the extra step of loading and unloading Windows.

Time saver

Save yourself some time and use .csv as the output format until you know that the data fields are being split in the right places. Then, when the field description is right, change the output format to the database format you want.

The commas and quotes will reveal simple field length problems such as: every ZIP code field beginning with a letter (some column before the ZIP is one byte too long. Of course, this example is too easy. Finding mistakes in data layouts when most of the information is readily recognizable stuff such as names, addresses, amounts and dates is easy compared to data sets that are hundreds of bytes long, with nothing but key fields, and lots of small fields with codes like "EA" and "J1."

When scanning the preview output, go to the end of the line, and work back to the beginning.

Load a small sample of the file (1,000 records or less) in fixed-position format.

Saving samples

I often do this step first with a dataset I haven't processed before. I save the sample file on the floppy I keep for every data project. All finished and tested Nine Track profile (PR2) files go here too. (I know I wouldn't want to have to do all that typing and testing all over again.) I also save small .dbf samples of all the tables in the data set.

I often print a small portion of this file, especially if it's obvious that some record layout debugging is going to be needed. Most data records are longer than 80 characters, so this is where a graphical word processor comes in handy. First, for the printout to be useful, you have to set

Continued on page nine

Continued from page eight: ... Several programs help

the text to a monospace font such as Courier or Courier New. Proportional fonts such as Times Roman will display the text with less width for the i's and l's and more space for M's and W's. The result is text that doesn't look like a fixed-position file. If necessary, I set the point size to eight, use Print Setup to shrink the left and right margins to the minimum and print sideways on the paper (landscape mode), or even choose legal size paper over letter.

Displaying this file with the List program also helps, since one of the features of this program displays a ruler or counter line at the top of the screen to make it easy to see, for example, what characters occur in positions 149 through 156.

Since fixed-position output only divides the data into records, but doesn't split the fields, you can quickly fill out the Field Description part of the PR2 file by making a single field with a length equal to the data record length. Use character as the data type.

Nine Track Express has a limit on how big a single field can be, so if the record length is greater than 500 bytes, you have to enter descriptions for 500-byte fields until what's left is less than 500 bytes. Then you make the final field length equal to the remainder. For example, if the record length is 1637 bytes, add field descriptions for three fields of 500 bytes, and then one field of 137 bytes.

Other tools

Use Excel on the sample fixed position file. I think everyone who sees and uses the Excel 5 Text Import Wizard immediately likes it because it's so easy. I used it recently on a non-problem data set I had just received because it was the fastest way to get the data imported without a lot of typing of field lengths.

I also like the Text Import Wizard because it's a great teaching tool for the concept of importing data. It presents the perfect picture of what has to occur when fixed-position files are imported into any spreadsheet or database, regardless of the software method.

This picture is also a great help in figuring out problem record layouts. You can use Excel's wizard to scan the records (as with a file viewer) and at the same time, designate where the fields should be split. When the wizard asks for the data type for each column, choose "Text" for every column. By default, the Excel Wizard designates every column as "General," but in the program's intent to be all-purpose and helpful, data such as Social Security numbers or other ID data composed of all numerals and no alphabetic or punctuation characters will be converted to scientific notation like 2.35E7, which is my Social Security number with the confidential parts rounded off.

Spreadsheet limits

Spreadsheets have a limit of about 16,000 rows, but you only need to test a few hundred rows. This is another reason for doing that fixed-position sample output of the data.

After you figure out the correct layout, you have several options. You can import the spreadsheet into FoxPro or Paradox and print out the table structure for reference in correcting the Nine Track field description. It temporarily takes up more disk space, but you could also load the entire tape file as a fixed-position file; empty all the sample rows from the FoxPro or Paradox table and then import the fixed-position file into the empty table.

Because of the difference in how Excel and FoxPro or Paradox store data, there is a limitation to this method. If the data has long character fields that are filled with just spaces, instead of data, those fields will show up in the database file with much smaller widths. The only way around this problem that I can suggest is to replace all spaces in the fixed-position file with some distinct character that is not used in the data. The caret character (^ , or Shift+6) is a good choice.

Richard Mullins can be reached at (573) 882-2127, or send e-mail to richard@nicar.org

**IRE committees
always are on the
lookout for good
volunteers.**

**The Advisory Council
for Periodistas de
Investigacion is led by**

**IRE Executive
Committee Chair
Deb Nelson of the
Seattle Times. To
volunteer, send e-
mail to Nelson at**

**dnel-
new@seattimes.com**

**The International
Committee is chaired
by IRE President John**

**Lindsay of Oregon
Public Radio. To
volunteer, send e-
mail to Lindsay at**

Lindsayopb@aol.com

**The Computer
Committee is chaired
by IRE Director**

**Shawn McIntosh of
The Dallas Morning.**

**To volunteer, send e-
mail to McIntosh at**

**dmnproj@airmail.net
For information on**

**other IRE
committees, send e-
mail to IRE Executive
Director Rosemary**

**Armao at
armao@ire.org**

By David A. Milliron
Gannett News Service

Continued on page eleven

Growing collection of federal databases

From the NICAR library

NICAR offers a number of federal government databases. Here is a list of our growing collection:

NEW • A monthly CD subscription for all 1995-96 Federal Election Commission campaign contributions by individuals and political action committees, plus all presidential matching fund requests.

NEW • The Health Care Financing Administration's 1995 database of all Medicare-funded inpatient work in U.S. hospitals.

- Federal Railroad Administration data for accidents, casualties, and highway crossings. 1991-1995.

- Coast Guard boating accidents, 1969-1994.

- Federal Aviation Administration data, including airplane maintenance work documented in the service difficulty report, pilot licenses and grades, and aircraft registration.

- Home Mortgage Disclosure Act records, for tracking who gets loans and who gets turned down, and finding redlining patterns.

- Federal procurement data, 1992-1994, includes breakdowns by agency.

- Alcohol, Tobacco and Firearms gun dealer records.

- National Bridge Inventory System data, includes inspection grades.

- FBI Uniform Crime Reports, a detailed compilation of crime data that includes statistical breakdowns of individual murders. This includes the new 1994 data.

- Social Security death records, by name and social security number, going back to 1937.

- Occupational Safety and Health Administration violation data includes worker accidents

and exposures to hazardous chemicals by companies.

- U.S. Department of Transportation truck accident and census data. It includes accidents by company and road.

- U.S. Small Business Administration loan guarantees, 1989-1995. This includes the name of the business, address, amount covered by the SBA, and status, including whether the loan went bad.

- U.S. Small Business Administration disaster loan guarantees, 1989-1994. This includes individuals and businesses, the amount covered by the SBA, and the status, including whether the loan went bad.

- U.S. Small Business Administration's list of minority companies certified for SBA assistance in seeking federal contracts. It includes the name of the company, its address, the owner, type of business and phone number.

- U.S. Department of Transportation hazardous materials accidents database, a collection of roadway, rail, air and waterway accidents from 1971 to 1995.

- U.S. Department of Transportation fatal accident reporting system. It includes all roadway accidents from 1988 to 1994.

- U.S. Coast Guard directory of U.S. merchant vessels. It includes the name of the ship, the managing owner, home port and various descriptive information.

- National Endowment for the Arts, grants, 1989-1993.

For up-to-date prices and more information, call (573) 882-0684, or send e-mail to nicar@muccmail.missouri.edu.

NICAR's week-long bootcamps offer

hands-on training in computer-assisted

reporting skills,

including the use of

spreadsheets and

database managers,

accessing data in

various media, such as

nine-track tapes, and

negotiating for data.

The next open session

is Jan. 5-10 in

Columbia, Mo.

For more

information,

call NICAR

(573) 882-0684, or

send e-mail to nicar@muccmail.missouri.edu

Continued from page ten: On the Internet

ments, Congressional Record and Index, Congressional Reports, Economic Indicators, Economic Report of the President, Federal Register, General Accounting Office, Government Information Locator Service, U.S. Government Manual, History of Bills, House Rules Manual, Interior Reports, Privacy Act Issuances, Public Laws, Unified Agenda and U.S. Code.

- The U.S. General Accounting Office now offers free on-line access to the GAO reports and investigations on everything from HMOs to immigrant students and government operations.

<http://www.gao.gov/>

David Milliron can be reached at

(703) 276-5805, or send e-mail to

dmilliro@gci1.gannett.com

Bits, Bytes and Barks

Moving on

Pontiac, Mich., *Oakland Press* reporter Jeff Green is leaving his job as environment/demographic/science reporter after seven years for a similar position as associate editor at Wards Automotive Reports in Southfield, Mich., a Wards Communication Publication. Green will continue to use CAR skills, adapting them to covering the automotive industry.

Jobs available

- *The Intelligencer Record*, a Philadelphia suburban daily, seeks an aggressive, hard-working reporter. Internet, spreadsheet and database skills a big plus. Send letter, resume and clips to Lanny Morgnanesi, executive editor, *The Intelligencer Record*, Box 858, Doylestown, PA 18901.

- *The San Francisco Bay Guardian*, an alternative weekly, seeks a city editor to manage newsroom and plan coverage of Bay Area news and progressive politics. Minimum three years management, editing, and investigative journalism experience required. Experience with computer-assisted reporting a plus.

Send cover letter, resume, three recent clips, and reference to News Editor Search, *S.F. Bay Guardian*, 520 Hampshire, San Francisco, CA 94110. No phone calls please.

Give us your opinion

In response to popular demand, NICAR is considering scheduling one or more three-day intermediate seminars in the coming year in Missouri and possibly at other sites. We would like your suggestions and thoughts.

Here are the proposed intermediate seminars:

- **Advanced data analysis:** Covering advanced SQL queries, data cleaning, lookup tables and more. Also, problem data from miscoding to messy record layouts. Should we also include problems that government computer tapes offer?

- **Front-ends:** Making data easier to search and use in the newsroom. This would involve using Access and Visual Basic, plus training in putting data up on internal Web sites.

- **Mapping:** While the annual University of North Carolina-Chapel Hill workshop links statistics and mapping, this workshop would focus on the nuts and bolts of mapping. We probably would use ArcView, but also teach principles that could be applied to all mapping.

- Last, we are looking at updating our Internet training again and would appreciate suggestions on that.

Please send messages to Brant Houston, NICAR managing director, at brant@nicar.org

Get them while they last

Uplink 1995 issues are available in a bound edition for \$10 plus shipping. To order, call (573) 882-0684.

New IRE-L and NICAR-L addresses

IRE-L and NICAR-L have changed addresses. They are located on lists.missouri.edu, a dedicated Unix system designed to service discussion groups run out of the University of Missouri. All posts to the lists should now be sent to: ire-l@lists.missouri.edu or to nicar-l@lists.missouri.edu

All commands for subscription changes should be sent to: listproc@lists.missouri.edu

NON-PROFIT ORG.
U.S. POSTAGE
PAID
PERMIT NO. 309
COLUMBIA, MO. 65211

Investigative Reporters and Editors, Inc.
100 Neff Hall
University of Missouri
School of Journalism
Columbia, MO 65211