

**IMPACT OF PRIVACY EXEMPTIONS IN OPEN RECORDS LAWS ON COMPUTER-ASSISTED
REPORTING**

by FEDOR ZARKHIN

Project Committee:
David Herzog, Chair
Scott Swafford
Mark Horvit

DECEMBER, 2013
Missouri School of Journalism

ACKNOWLEDGEMENTS

First and foremost I would like to thank my project committee members for all their help and support throughout this process. Their wise words and pithy criticisms are what allowed me to come this far, so thank you, David Herzog, Scott Swafford, and Mark Horvit. I'd also like to thank all of the Columbia Missourian editors who have been such a source of inspiration these past two years. Katherine Reed, John Schneller, Liz Brixey, and Jeanne Abbott, I am forever indebted to you for the love of journalism that you inspired in me.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
Chapter	
1. INTRODUCTION.....	1
2. WEEKLY MEMOS.....	2
3. LETTER FROM SUPERVISOR.....	29
4. EVALUATION OF WORK PRODUCT	32
5. EVIDENCE OF WORK.....	35
Articles	
Interactive graphics	
SQL queries	
Mapping analysis	
Scrapers	
Records requests	
6. PROFESSIONAL ANALYSIS	
Literature Review.....	86
APPENDIX	
1. Project proposal.....	101
2. Questionnaire.....	122

Introduction

I came to the University of Missouri's School of Journalism with vague and idealistic goals of changing the world for the better. I couldn't think of a career path more ideal for satisfying my desire to learn, write, and be useful in some way to humanity.

Two years into the program, I have a more practical understanding of journalism, but that doesn't mean I've strayed from my original goals. The initial idealism remains, but it has been complemented by a set of skills I plan on using throughout my career. The skillset I least expected to acquire and value as much as I do is that of computer-assisted reporting. Taking David Herzog's data journalism class, working at the National Institute for Computer-Assisted Reporting, and working with too many datasets to count – all these experiences have solidified my love for numbers, spreadsheets, and SQL queries.

I am confident in my skills as a data reporter and equally confident in my desire to develop them further while participating in investigative projects. I am now ready to take what I have learned and get to work in the "real world."

My career goals are not quite clear to me yet, although I will certainly incorporate CAR in my career. My tentative plan is to gain experience as a computer-assisted reporter and also as a beat reporter in the United States, and to then take the skills I learn and use them at a publication in Russia. The Russian media have a lot of progress to make if they are to adequately serve the watchdog role in Russian society. I want to be part of that progress.

Weekly Memos

Week 1

Hi folks,

I've been at the *Palm Beach Post* for about three months now. In that time I've been working on a lot of long-term projects, I've been helping other reporters with their data projects, and I've done a few stories of my own.

This week, I continued to work on a project to visualize property appraisal data. The goal is to make a map that will let readers search for their home and see how their house was appraised in 2013 versus 2012, and how much more or less they will have to pay in taxes.

New appraisal data for the county will be released in August. We asked for this year's data from the appraiser to build the app, and I've been preparing the hundreds of thousands of records for use.

Neither I nor the interactive projects manager know exactly how we will make this app, but there are a number of options. Recently, for example, I discovered a wonderful template built by Derek Eder. It requires some HTML, Javascript, Google Maps API and CSS knowledge on my part, but Eder did the vast majority of the necessary work.

I've also been working on a story looking into repairs that are being done on a flyover connecting the regional airport to Interstate 95. To put it briefly, the flyover was built by the Department of Transportation about a decade ago to vocal opposition in the county. Soon after the project was completed, inspectors found egregious errors in the

contractor's work. Among other things, my story will address the company's track record and the mistakes they made in their work.

This week I continued my battle with numerous police departments for dispatch data. The problem appears to be that if I know where an officer was dispatched and why, that could violate victims' rights. I will continue trying to find a solution to this problem. So far, I haven't even managed to get the sheriff's office to list the statutes under which I can't have the data. The PIO says that to give me all the exemptions, the department would have to actually fulfill the request. There are millions of records, and the expense of redaction would be prohibitive.

Last week I got a database with fire response times for all the fire departments in the county for the past five years. I averaged the times by zip code and mapped them. As expected, the response times increase in the less populated parts of the county. I have yet to look into the data more closely in search of stories.

Finally, I've been working with Jen Sorentrue, the county reporter, on a story that hinges on data.

I'll try to give you the simplified version. The county inspector general issued a report earlier this year stating that based on an audit she did, a local waste management company owed the county \$300,000 for dumping trash outside the county. Prior to her becoming the IG the county contracted a forensic auditor, who found that the company owed the county \$700,000.

The IG claimed that she excluded the illegal dumps found by the first auditor in her audit. But when Jen and I requested the forensic auditor's data and joined it with the IG's data, we found that she had double-counted illegal dumps.

We emailed the IG our findings, and she said she is checking her data.

I made very little progress on my project this week. Mostly I've been working on formulating the specific questions I will ask on my survey. However I phrase them, I know I'm going to be asking a lot from responders. I need to find a way to introduce the survey in a way that makes them believe their responses would provide a service to journalist-kind.

Mark, if I could take you up on your offer to let me use the IRE member mailing list, I'd very much appreciate it.

This week I will finish the flyover story and continue plugging away at all the projects I described above. I think Jen Sorentrue will write the IG story this week, but I doubt she'll involve me in the writing very much. I need to have property appraisal data ready next week.

Thank you very much for your attention.

Best,

Fedor

Week 2

Hi!

Hope you are all doing quite well.

I spent most of Monday and Tuesday working on an interactive map of aircraft accident data. Over the weekend I took a crash course in HTML and CSS then figured out how to adapt Derek Eder's searchable map template and make the map I wanted.

I've been researching how to make a searchable app for property appraisal data, and I think I'm going to stick with Derek's template. The other options require learning so much that I'm not sure I'll be able to do it on time.

Because there are limits to how much data Google Fusion will display, I'll have to split the current shapefile – which has about 450,000 polygons – into five or so pieces, then make a separate map for each.

I have to admit, this is a monster of a project. A query to calculate the total owed in taxes per property has been running over a week now. Fortunately, I have Adam Playford's wisdom to guide me along. Adam was the data guy at the Post before I came, and he's very good.

Wednesday through Friday I worked on my flyover story, which I mentioned in the previous email. It's a bit complex, what with all the technical details and the pacer documents to sift through, but it's interesting. I'll have it done this week.

I've taken a pause from trying to get dispatch data. Most likely I'll end up either requesting data for a few specific crimes, or I'll just request the data that some agencies already post online.

Saturday I was on rotation for the weekend GA shift, and Sunday I was asked to do a story on youth reactions to the Zimmerman trial. I asked for Monday off to work on my Master's project.

This week I will finish my flyover story. I've also got a few things to do for other reporters, including helping one of the education reporters navigate a student discipline database, analyzing a court database for the social services and immigration reporter, and extracting some data from the Department of Corrections felon database for Pat Beall, an investigative reporter.

I'm on my way to finishing the questionnaire, though I think it could use some more work. You should receive an invitation to complete the questionnaire in a separate email.

I will email it out as soon as I make changes after getting feedback from David and my colleagues.

Thank you so much!

Week 3

Hello Gentlemen,

I just sent you a revised version of my form. I like this one a lot more, though I think it might have become a little clunkier. I did tighten the introduction, though.

So, as I said, I took last Monday off. Tuesday through Friday I worked either on my flyover story (coming soon, I hope!) or on data stuff for other reporters. For example, one of our reporters needed to get a shapefile of parcels where new developments had been proposed in part of our county. It took me half a day to make ArcGIS do what I wanted it to do (stubborn program, I have to say). This required joining our parcel shapefile with a spreadsheet with about 20 parcel numbers. The rest was

pretty easy (Google Fusion is my friend), and now that I know how to make it work, the whole process won't take longer than 10 minutes.

The education reporter gave me a pretty unwieldy Access database with student discipline data. I have no clue why the data's such a mess, but I'm currently in the process of finding a way to get the reporter – Allison Ross, Mizzou grad – what she needs.

I had a query running for over a week and a half, and after speaking with Adam Playford, I killed it. He explained that I must learn indexing if I want to run gigantic queries. That's something I'm doing right now.

It looks like the Post is starting out on a massive criminal justice project. The exact outlines are vague, but I do know a lot of it hinges on data. The editors told me to have the court data in perfect shape within two weeks.

Last week, Joel suddenly remembered that I'm an intern and that it might not be wise to stake the paper's reputation on my work. So I'm going to redo everything I did with the court database and write down every single step I take, SQL queries and all. This will involve using regular expressions to fix some errors, converting values such as 12298619 to 1986-12-29 and other things like that.

Lawrence Mower, an investigative reporter, has been wanting to learn SQL for a while, and last week we finally covered the basics.

Lawrence has been working for a while on a lottery story based on data he got months ago. We found some very interesting things in the data, namely a few people winning way too often, but now we're having second thoughts. Lawrence ended up also

requesting a database the Sun-Sentinel requested, and we found some major differences between what we got and what the Sentinel got.

As I said, I sent all three of you the most recent version of my questionnaire. I'd like to email it Tuesday. I think it's pretty much ready, but I've been staring at it for so long that I can't be sure.

This week at the Post will involve a lot of work with data. I'll finish reading up on indexes by tomorrow, I hope, and then I'll re-run queries calculating total taxes owed for the massive property appraisal project. At the same time I'll take the court data and start re-processing it.

I just realized I don't have any pending records requests. This is a travesty and will be amended ASAP.

Have a great week, and thank you for your time!

Week 4

Hi folks,

I hope all of you had a marvelous weekend.

So, as you can see, it's been taking me much longer to finish the survey than I anticipated. In its current form, the survey provides responders with different questions depending on whether or not they got the data in response to their request. I'm also asking people to only talk about their last request, as David suggested. Otherwise it'll be harder to make generalizations. I'm going to send the survey to a few more reporters that I know tomorrow, and will email it to the NICAR listserv either Thursday or Friday. This is the final deadline I'm setting myself.

Last week I kept working on court data and my property appraisal map. Both are inching closer to fruition, with the court data being almost finished.

The appraisal map is going to take quite a bit more work. The problem is that I'm dealing with lots of data, so if I don't run a query right, I have to wait an extra ten minutes to rerun it. It's frustrating, but I'm getting closer. The obstacle I'm facing now is with condos... I won't bog you down with the details, but the crux of the problem is that I'm not getting the number of parcels that I should. Fortunately, the property appraiser's office is very helpful. They've been walking me through some of the problems I'm facing.

In the course of a conversation with the appraiser's office I discovered that they keep a record of every single construction permit in the county and the municipalities since 1990. My editor and Paul Blythe, the watchdog editor, were quite excited about the potential of this data. So, once I get it, I'll be working with Jen Sorentrue, the county reporter, on growth and stagnation in the county over the past ten years.

The fact that I'm working with Jen as opposed to the i-team on this is not ideal, it seems. In essence, Paul heard I wanted to request this data before Joel, and he sort of "took" the story. Joel was not happy. So, clearly, I have to think more about who I talk to, about what, and when.

Paul and Holly Baltz – one of the managing editors – want to kill the solid waste authority story. The problem is that we got the data our story hinges on from a source with an obvious stake in the matter. Joel, on the other hand, believes the story should run. There are two options now: Either Jen writes a story about ALL the grievances against the inspector general, or the i-team takes over the story.

Last Friday the Florida Department of Education released school grades for all elementary and middle schools in the state. I made the spreadsheet of local school grades for the print graphic. Believe it or not, this took almost the whole day because it involved lots of table joining and lots of cleaning. Over the weekend I made an interactive map of school grades, which should run tomorrow. It's already online, and you can click here to see it. It's very similar to the plane accident map, except it features a slider. And sliders are very nice.

This week I have to make some serious progress on my property appraisal map. I have about three weeks until deadline, and I'm not nearly as far along as I'd like to be. Parallel to that I'm going to finish the court data.

Pat Beall's four-part series on private prisons is coming out in about three weeks. I've read the first part and it's really, really good. Joel and Pat would like me and the reference library folks – Niels Heimeriks and Michelle Quigley – to make some interactives.

Some general reflections. Working here is awesome, but the way I'm wired, I thrive on product. Deadlines and bylines in the paper are what really keep me going. So I've been craving an awesome story, preferably a data one. But finding data stories isn't a piece of cake. I need to keep learning as much as I can about the county and I must drill down to find the stories.

I apologize for the long memo!

Also, thank you for all your help and advice throughout this process. I really appreciate it!

Best,

Fedor

Week 5

Hello folks,

Hope all of you are doing marvelously.

Please don't laugh, but I extended my survey deadline till tomorrow. The reason is that I think NICARians are more likely to see it Monday morning than any other time during the week. Please tell me if you believe otherwise.

I've had quite a few people check out the survey and I think it's ready to go. I just need to take the plunge and click 'Send.' Be brave, Fedor!

My flyover story finally ran Saturday. It was interesting reporting the story because I really had to learn a lot, and got to dip my toes into subjects as varied as bridge construction and Chapter 11 bankruptcy proceedings. It was also kind of surreal to have one of our lawyers read the story to make sure I didn't write anything that could get us sued.

If you read the story, though, you might be underwhelmed by the relative insignificance of the subject matter. After all, the story's about \$1.8 million in repairs on a \$156 million bridge that aren't even being paid by the state. Still, it was good practice.

The school grade map was a success. My Google Fusion Tables API got 25,000 requests in the day and a half the map was on our home page. Every time a user asks the map to show different information counts as a request. The Post doesn't have interactive graphics like that often, so the editors were impressed.

I have received the building permit data from the property appraiser and am now in the process of making it usable. We don't know yet where it will take us, but Paul Blythe – the watchdog editor – wants me to work with his reporters on a story on growth in the county over the past ten years.

I've made some great strides on the property appraiser maps. I had a few very technical obstacles to overcome, and it took about half a week to solve them. But for every problem that I solve another ten appear, so this is just one step.

It has become very clear to me that I need to learn Python and Javascript if I want to do well at the Post. For my maps I've confined myself to making simple adjustments to existing code to make them. I don't have the skills to make significant changes, and this is a problem.

In addition, I use a GUI – Navicat – for all of my data processing. I write SQL for the actual queries, of course, but I think Python would give me a lot more versatility. Finally, a few of the editors want somebody who can scrape websites. I can only do the most basic scraping.

I won't have time to learn Python at work, but I'm going to try to devote a solid hour to Python each day after work. Wish me luck!

This week I'm going to work on my property maps. As of now, I've managed to join a 400,000-row table with a parcel shapefile, select a portion of it, convert the result to KML, make a Google Fusion Table, and make a map similar to my school grades map.

Now, I need to 1) group individual condo units into condo buildings, 2) decide how to split up the county for the maps 3) figure out how to make Google Fusion display thousands of polygons even when I'm zoomed out of the map.

Then I have to do all of this again once we get this year's data, which should be any day now. I've told Joel that once we get the data, I could need three solid weeks to do the work.

I'll also start playing around with the permit data this week. I'll just run a few basic queries but I'll give it some more concerted effort after I'm done with the property maps.

I didn't do anything with the court data last week, though what I have to do shouldn't take me longer than half a day to do.

Joel and Holly Baltz – one of the managing editors – want my next story to be data-inspired. I want that too, but we don't have anything for a story yet.

Tomorrow I'm going to send the survey and read through the responses I get the rest of the week.

Thanks so much for your time!

Week 6

Hello Folks,

Hope all is going well.

Apologies for the late memo.

There isn't much new to tell about my work... I've spent all of last week and the first two days of this week working on my property appraiser maps. Getting the data

ready has been an arduous and time-consuming process, but I'm getting quite close now to being done.

I think the process has been very good for me. For one, I've learned new SQL queries and I've really internalized a lot of techniques that previously I knew just a little about. I've also become better-acquainted with mapping software, namely ArcGIS and QGIS. Very useful stuff.

Still, that doesn't make me immune to error. Yesterday I discovered I have to redo a lot of what I've done because of a rather stupid mistake: instead of averaging condo property values I summed them.

Last week David sent my survey to NICAR. I didn't get many responses, so I sent it again a few days later. Still not many responses. I am sort of panicking, and this is inspiring me to keep contacting everyone I know and don't know to fill out the survey.

For the rest of this week I'll be working on the property maps and on getting in touch with data reporters. The maps will run Sunday, but they need to be done Thursday.

Thank you. Have a great week.

Week 7

Hello,

This week was quite busy. I worked almost full-time on my property appraiser maps, which ran this morning. There wasn't a day this week when I left earlier than 11 pm. All worth it, though, because I pretty much met my goals. My Google Fusion API has already gotten 15,000 requests, and the maps should get many more clicks tomorrow.

This project was a fantastic learning experience. Throughout the project I probably ran more SQL queries than I had in the entire year before, and I also learned a lot of GIS software, with some Javascript along the way.

The maps had a name search option, but I had to take that down after a reader pointed out that users could find people whose addresses are confidential, such as the sheriff. I'll remove those records tomorrow and put the name search back into the maps.

Now that that project is done, I think I've done the most I can with Derek Eder's template. I'd very much like to get to the level where I can write my own apps. But that's not what I'm going to focus on now. In terms of programming, the current priority is scraping. To that end, I'm going to ask Joel tomorrow if he'd be all right with me spending an hour a day learning Python.

I've been so busy with the property maps that I haven't had the chance to work on my research. The situation looks very bad. I have a list of people to contact, and I'll start making calls and sending a lot of emails during my lunch breaks. This looks to me like something of an emergency. I can't decide if I should cut questions out of the questionnaire.

This week I'll help Lawrence Mower with lottery data. I've taught him the basics of SQL, though, and he's been trying to do a lot of things on his own. Hopefully he will inspire others to learn, too, and I will be free to pursue my own stories.

I'll start analyzing the building permits data this week, in addition to mining the appraiser data for more stories our county reporter can work on. The appraiser data is

very, very rich. You can find out everything from what a house's floors are made of to when it was last sold to how many bathrooms it has. Maybe there's a bathroom floor story in there.

Thank you for your time.

Week 8

Hello folks,

I hope all of you are doing quite well.

Last week didn't feel very productive, but I think that's because I do not have a concrete project to work on right now. Mostly I'm scouting for ideas and checking out possible leads given to me by editors.

For example, Joel wanted me to check some data to see if there was the potential that a landowner was getting help from Riviera Beach to buy properties cheaply. Turned out that was not the case. The properties in question were all bought in 2006, and Joel said this was supposed to be a relatively recent occurrence.

Right now I'm looking into medical practitioners in Florida, trying to find evidence of any kind of abuse. For example, I joined our court cases table with a table of all licensed medical practitioners in Florida. I think that perhaps if I find that multiple people convicted of abuse worked at one location, there could be a pattern. At this point though I don't know exactly what I'm looking for. I'm waiting for something to jump out at me.

Last week I also worked on my sheriff's office blotter scraper. I've made incredible progress relative to where I was when I started, but now I'm stuck again. It's

quite a complicated site to scrape. At the moment I can, technically, scrape 16 years of names, addresses and booking dates of arrestees, but I can't scrape the charges in such a way that they would be linked to their respective arrestee.

I've emailed about 60 people asking them to fill out my form. I'm emailing everyone individually, addressing them by name and mentioning their news organization in the body. I make sure to ask them to forward the email to their colleagues, as well. I will continue to do this throughout the week.

This upcoming week I'm going to keep digging into medical practitioner and court data. I'm back to that point where I am in desperate need of a data story, and I'm not getting all that much direction from my editors. They have ideas, of course, but it's entirely up to me to turn an idea into a viable question to report.

Monday the property appraiser should have answers to my questions about the building permit data. Once they do, that will take up a lot of my time. Lawrence Mower will be finished soon with a project of his, which means he will go back to working on his lottery story full-time. And that'll mean lots of data work for me. He's been learning SQL, though, so he will do some of the data work himself.

I do have to say I feel immensely lucky to have this freedom. I'm sure many people work years to be able to have the freedom to just look around for stories. I better not waste this privilege. I'm guessing it would be easy to fall into doing a massive visualization project like the appraiser maps, but I would prefer to avoid that.

Thank you for your time and have a magnificent week.

Week 9

Hello Gentlemen,

I hope the start of the semester has gone smoothly and that all is well.

Last week I continued my fishing expeditions for stories. I've done quite a few joins, looking for nursing assistants charged with elderly abuse, teachers charged with violent crimes, pharmacists charged with prescription pill trafficking, and the like. Though I did find examples of all those things, I didn't find any patterns, certainly nothing worth a deeper look.

It's fallow moments like these when I really hate not knowing that much about this county. What I'd like to do is find a way to get at the stories hidden in plain view, such as the issue of undocumented migrants. Who systematically abuses human rights here, and how?

Something will turn up soon enough, I'm sure.

It appears that I'll be working closely with Mower on the lottery story. We found that a few of the people have won in one lottery game an improbably high number of times, and we've been trying to figure out how.

I've sent the questionnaire in individualized emails to over a hundred reporters, and currently have 31 responses. I'm not done with the CAR conference attendees email list that Mark sent me yet, and will finish sending those emails this week. I am encouraged by the thoroughness of a lot of the responses.

This week I will look through a database of phone calls made by county commissioners over about a year. I'm not sure what I'm looking for, but for now I'm just

toying with the data. The biggest challenge is in identifying who exactly the commissioners are calling.

It's possible we're going to request the Florida slice of the OSHA database from NICAR. If we do, I'll be the one analyzing it and looking for stories.

Once Mower is free to work on the lottery story, I'll join him on that full-time.

Thank you for your time.

Week 10

Folks, I apologize for the delay.

I spent most of last week working with Lawrence on court data. What we have is a database of lottery sales broken down by month, game and store, and a database of lottery winners, broken down by name and store. A few months ago we found some incredible outliers in the data, quite a few of them people associated in some way with convenience stores. We also found one man who has won almost \$2 million in less than half a year on a game with a \$5,000 prize. Among other things, Lawrence wants to find as many 'insiders' as possible, as well as find the luckiest stores. That's why we got sales data – we want to compare sales to wins.

The greatest breakthrough for me last week was in scraping. I wrote three scrapers, and am, to be honest, quite proud of myself. The first scraper was for [this site](#) that automatically lists the West Palm Beach Police Department's dispatches. It only shows 200 records, so I have my scraper running all the time, scraping the page every hour for new records. There might be something interesting there in a few months. The problem, however, is that anything besides traffic related calls is labeled 'intelligence-

led policing,' which is hardly useful. Still, it might be interesting to see where the police are making the most traffic stops – there's geo-data behind what you see on the site.

I also wrote two scrapers for the civil court site. Because there's such an enormous amount of data the scrapes must be targeted, such as a scrape of all cases involving the Palm Beach County Sheriff's Office. The scrapers get information such as the case number, filing date, case type and case status and plug it into a PostgreSQL table. From there I can query the data to my heart's desire.

I continued to email people last week and I continue to collect responses. I've had to start harassing people at work to fill out the survey. They're also sending it to people they know.

This week I have been sick, and I left early Monday and took today off. I should be good to go tomorrow, though, and I'll continue to work on lottery data. As soon as I'm done with that I'll start cleaning building permit data. There are some problems with it, but the appraiser has basically said "Tough." There's also more court data work to do, as usual.

With regards to my research, I'm going to start calling newsrooms this week. A lot of the bigger places didn't respond to my emails, so I'm going to have to start being a bit of a pest.

In other news, the Atlanta Journal-Constitution is going to fly me out next week for an interview. I had a pre-interview interview this morning with Shannon McCaffee, Victoria Hicks and John Perry, and apparently it went well.

Thank you for your time.

Week 11

Hi Folks,

The Post offered me a job Friday and I'm going to accept it. I thank all three of you for your help, support and instruction over the last two years. I'll do you proud, I promise.

They want me to start this Friday, but because that would be before the end of my project I'd need your approval to start then. Another option would be to ask the Post to postpone my start date a few weeks.

I'm not going to the AJC interview because I'm set on the Post. The job here will offer me much more freedom, including the freedom to write, scrape, work with the i-team and do web projects. The AJC job would be limited to helping reporters with data and records, and I wouldn't be able to write unless I have an exceptional story to work on.

Wednesday through Friday I worked with Lawrence on lottery data. We've been running some heavy queries trying to find anomalies in wins by store. A query that took me a while to get right was the ration of scratch game payouts to scratch game sales by store. We're also looking into winners who are connected to stores that sell lottery tickets, seeing what we can dig up about their backgrounds. This week I will most likely go with Lawrence to some stores to talk to the 'insider' lottery winners.

The i-team, a few of the reporters and the reference library folks have started meeting weekly to discuss progress on the court data project. Last Wednesday we just threw around ideas. One of the obstacles we have is with Hispanics: usually they are

counted as Whites, so analyzing anything for race/ethnicity becomes quite difficult. But, of course, there are a lot of other things to look at.

I wrote another scraper last week, quite a bit more complicated than the previous ones. It is for this site. There are a few minor things I need to fix up before I actually scrape all the site's data.

I only got ten more responses, though there are about five people who've promised to fill out and haven't. I didn't get to calling anyone last week, though I definitely will this week. I haven't gotten a single response from ProPublica, even though I sent at least seven people emails. So this week I will really start harassing people.

Besides for continuing to work with Lawrence, I'm going to start giving the permit data a good solid look. Paul Blythe, the watchdog editor, is most interested in where the growth has been going on. Fortunately each building permit has a parcel number, so I'll just join that with parcel shapefile. I already see a great interactive in the making!

Thank you for your time.

Week 12

Hi folks,

I apologize for the delay.

Last week I continued working on the projects that have been assigned to me, as well as pick up a potentially good story.

In addition to running lottery queries with Lawrence, I've started working more with the court data. A group of reporters has started meeting weekly to talk about ideas for the data, which has been enormously helpful. Our criminal court reporter, Daphne Duret, comes to these meetings, and her input has been invaluable.

I used the data last week to help our breaking news reporter, Alexandra Seltzer, produce an in-depth piece of people driving with suspended licenses. One such person recently ran over and killed a child, so the follow-up story, which very much relied on the court data, was timely and appropriate.

Now we are looking into violations of injunction orders. There are numerous cases of women being killed by the men who had previously been violent towards them, and Daphne would like to take an in-depth look into this issue.

Recently a tip came in to Joel that people regarding a non-profit called Wounded Warriors of South Florida that I am looking into.

I have an enormous amount of things to do in the near future. The priority this week is the non-profit story, which Joel doesn't want me spending more than two weeks on. There is no unequivocal proof that the allegations are accurate. Still, it turns out that it's possible to find out whose names have been searched in the D.A.V.I.D system with a simple request to the Florida Department of Highway Safety and Motor Vehicles. Hopefully that will give me something more to work with.

This week I start working on a map based on our lottery data.

I have almost fifty responses now, which is acceptable but not nearly as many as I had hoped for. Some publications didn't submit a single survey, despite me having sent

the survey to up to seven reporters. I am particularly disappointed with ProPublica. I tried reaching Jennifer LaFleur by telephone and I emailed her twice, but to no avail.

At this point it's high time to start working on the article itself. More responses will come as I nag friends and they nag their friends, but that won't get in the way of the work I'm doing now. So this week I will analyze the basics of my data then read through each response. I have yet some papers to email, though, so I'll keep doing that.

In other news, my new landlord, who is Cuban, doesn't speak a word of English.

Welcome to South Florida!

Have a good week.

Week 13

Folks, I really do apologize for the delay. I'm sorry!

It turns out I actually start working next Monday because that's right after a pay cycle ends.

Last week I kept working on my veterans nonprofit story. I miscalculated and had an interview with some of the current board members before I was completely ready. They gave me the impression that the folks badmouthing them to me are simply upset that they lost control of the nonprofit.

I did find a pretty good lead, however. So, I requested all searches in the DAVID system made by the Palm Beach County Sheriff's Office detective on the board in the last year. The clerk at the Florida Department of Highway Safety and Motor Vehicles sent me the first two pages of my request with almost everything redacted, to show me

that I wasn't going to be able to get the names of people searched. The agency's lawyers have yet to get back to me on whether or not the clerk's decision was correct.

Last week I also worked on domestic violence cases in the court data. Daphne Duret, our criminal court reporter, wanted to know how many people with injunction violations had prior criminal convictions, among other things. I supplied her with this information.

I also started working on the lottery sales map. In the process I discovered that some stores have multiple identification numbers, reflecting changes in ownership. This poses a slight problem, and must obviously be fixed. I'll also have to rerun some of the previous queries, though I doubt the results will be significantly different.

The rest of this week I will work on the lottery map as well as continue working on the veteran's nonprofit story. I have a number of people left to call for the latter, and I've got to keep working with the lawyers to get the records I want.

For my project I've looked at the data and can offer some basic stats. On average, respondents have worked as journalists for 13 years, and there's quite the spectrum in terms of types of reporters: There are data reporters, investigative reporters, an online editor, a multimedia reporter, and more. 38 of the respondents work for newspapers.

Most of the respondents needed the personally identifying information in their requests, which is quite fortunate for my project. Forty of them either got the data they requested or got a version of it.

I have not yet done the work of digging through each response and identifying trends. From what I've seen so far, however, the requests are quite varied, which is good. Most of the people who responded put some effort into formulating complete responses.

This Sunday I will send you a timely memo with a much closer analysis of the responses. Hopefully I'll also have more of them to analyze.

Once again, my apologies for the delay.

Best,

Fedor

Week 14

Hello folks,

Last week went pretty much as predicted.

With my veterans nonprofit story, sources are slinging a lot of mud at each other. The only thing I have that could provide concrete proof of anything is a list of the names searched in the DAVID system by the detective on the nonprofit's board. This week I requested that list, despite the fact that all the names will be redacted. I've also asked for all searches done on a few people the detective might have searched.

I've been steadily reworking the lottery data because of a problem I found recently. The table updates have been taking up to two days, so that's been a bit slow.

I've also started on a new "request everything in sight" phase, which is always fun. The Florida Fish and Wildlife Commission Inspector General found an employee had used her P-card for \$6,000 worth of personal purchases, so I requested all the FWCs P-

card purchases since 2011. That's something I can and most likely will request from more agencies.

I also found a form for use of force incidents that Department of Corrections employees must fill out after an incident, and I requested the database that I assume that information is pushed to. I also asked the DOC for a breakdown of reported rapes and rape prosecutions by prison.

Tomorrow is my first day as a full-time employee, which I'm celebrating by going to Miami for a First Amendment Foundation seminar. Hopefully I'll meet Barbara Petersen.

I'm almost done with the lottery data, so I'll just have to rerun some of the old queries. There weren't that many stores with bad IDs, so I doubt the results will be very different from what we already have.

The general idea I'm getting from the responses is that there are kinds of data that are within definite categories of accessibility. Anything to do with DPPA or HIPAA, for example, is hard to fight. Government employee names and salaries, meanwhile, are almost routine. I have at least four responses from people who requested employee salaries.

Between those two categories is a grey area where people can succeed if they fight or negotiate. Agencies will deny records for seemingly obscure and irrational reasons, and the best way to fight that is to keep trying, of course. I was most impressed by a former AJC reporter's method. His request was denied, so he asked the state's

Secretary of State for an opinion. The secretary ruled in the reporter's favor, so the records were handed over. Short, sweet, to the point.

There are a lot of good insights in the responses. Many situations are unique, of course, but most techniques are transferrable.

This week I'm going to write a first draft of my paper for the project. I need to get into full-time project mode, with deadlines fast approaching. Thank you.

Letter From Supervisor

To: David Herzog, Committee Chair

When The Palm Beach Post hired Fedor Zarkhin as a one-year database intern in April 2013, we anticipated Fedor would bring enthusiasm to the role. We expected he would dig, both as reporter and data guy. We knew he would follow leads and crunch numbers with persistence and good humor.

What we didn't anticipate was that Fedor would teach himself programming skills to scrape websites for vital information. We didn't think he would set and maintain a goal of seeking one database a day from government record-keepers and push one agency until it calculated the cost of responding to his request to be \$2.7 million. We didn't think that even as the newspaper contemplated voluntary buyouts to reduce its work force, we would be persuaded less than six months after his start to offer Fedor a full-time job as The Post's lone data reporter.

But all those things came to pass. Fedor has flourished as a dogged reporter, best seen in his comprehensive accounting of an interstate highway glitch repair job, which resulted in a 1A story, to his willingness to help other reporters make their stories better and teach them data techniques.

Fedor's biggest task has been to handle an avalanche of digital records – the entire database of the Palm Beach County Court system, going back to the early 1900s. The docket data alone tallied more than 20 million rows. He received the data in nearly a

dozen text files and converted the data to SQL . Using Navicat to analyze the data, Fedor has found flaws that even the court administrators didn't know the data contained. Even as we build toward comprehensive stories on Palm Beach County's justice system, Fedor has made the files useful for daily reporting, helping to localize a national story about racial bias in drug arrests.

Most impressive is Fedor's ability to teach himself how to do things that he had never done before. His ability to scrape the West Palm Beach police logs, which they themselves post online but refuse to deliver directly to us, gives us another critical way around bureaucratic obstacles to open records laws.

Working with an I-team reporter on state lottery data, Fedor not only helped him organize the data but taught him SQL so now he can analyze it himself. He did this without calling attention to it at all but just because it made sense.

Fedor brings a reporter's way of thinking to his analysis of records and a data expert's thinking to vexing story problems. For instance, faced with accusations that a police detective on the board of a charitable organization used his police resources for his private-sector work, Fedor found a little-known database kept by state police that proved the accusations and could help us in countless ways in the future.

He has asked agencies for lists of all the data they keep, opening a digital world that for years no one at The Post attempted to tap.

Fedor has been a pleasure to work with, eager to pursue stories and never too busy to be bothered to help with even the most mundane concern. His work here has been excellent and we are excited to have him on the payroll as a full-time employee.

Joel Engelhardt

Investigative Team Editor

The Palm Beach Post

voice: [561.820.4732](tel:561.820.4732) | cell: [561.346-9489](tel:561.346-9489) | fax: [561.820.4407](tel:561.820.4407)
jengelhardt@pbpost.com | www.palmbeachpost.com

Evaluation of Work Product

My goals coming to the Palm Beach Post were two-fold: to gain experience preparing and analyzing data and to write pieces of my own. I believe I succeeded with these goals, in addition to gaining skills I didn't even expect.

Throughout my time here I have worked with many databases, including the multi-million record tables in the court and lottery data. I have performed what for me were rather complex joins and analysis with relatively obscure data from the local government. I have requested a few dozen records, with each one gaining experience in how to fight for them and as well as knowledge of public records law.

Most importantly, I think I developed my approach to computer-assisted reporting. Although my experience at the University of Missouri had already taught me to be careful with data, I have become more thorough and less trusting of data in my time here. Not to mention, of course, the actual data skills I learned while here, including regular expressions and numerous SQL techniques I was not previously aware of.

I also think I have become better at using data to look for stories. Initially my approach was to request everything in sight, then not look too deeply because I wouldn't know exactly what I was looking for. Now, I know it is best to have a specific idea of what you're looking for before requesting data. The more systematic and thought-out the approach, the more likely money and time will not be wasted.

I did not have many opportunities to write my own stories, but the one I did write – on a big flyover construction project – allowed me to do all my own reporting. It

was particularly satisfying because the subject was of local importance, and the paper ran numerous letters to the editor in response to my story. It was not a data-driven story, but it did hinge almost exclusively on records available either on PACER or requested from the Florida Department of Transportation.

The project also helped keep my deadline reporting skills fresh with a five-week rotation on the weekend general assignment shift.

I was pleasantly surprised in this project by learning skills I did not anticipate I would learn, namely data visualization and web scraping. I am proud to say that I even made something of an impact at the paper when it comes to visualization, being the first, to my knowledge, to introduce searchable and very interactive maps. The property value map I made was my most substantial project at the paper, allowing users to search for properties by name, address, and property value range. It took over a month to do, and I learned a lot from it, from mapping software to data cleaning.

I also learned the fundamentals of scraping, which was a very pleasant surprise. I dabbled in Python prior to coming to the Palm Beach Post, but it always seemed inaccessibly complicated for me. I finally broke through after a week of frenzied coding, and now feel comfortable tackling scraping projects. We have not yet used any of the data I've scraped in a story, but I am sure the opportunity will arise.

Despite the generally positive nature of this project, there were some things I was not pleased with. For example I did not get sufficient opportunities to write. It appears that computer-assisted reporting skills are in such high demand that I have to leave most writing to the beat and investigative reporters.

In addition I think my work was stymied because of an absence of anybody to learn from. It would have been extremely beneficial to have an experience data reporter to learn from, but there wasn't one at the Palm Beach Post, so I relied on friends from the University of Missouri and a former Post employee for advice.

Evidence of Work

Stories

Local Tax Revenue to Grow – Aug. 27, 2013

BY JENNIFER SORENTTRUE AND FEDOR ZARKHIN

Palm Beach County property owners would pay a maximum of \$98 million more in property taxes and special assessments next year under proposed tax notices mailed last week, a Palm Beach Post analysis shows.

Property owners can see how their property's market value and proposed tax bill compare to their neighbors' by visiting The Post's exclusive interactive map atpalmbeachpost.com/go/propertyvalues. The interactive, based on the Palm Beach County Property Appraiser's 2013 tax notice data, also allows visitors to create a variety of maps showing properties in certain market value ranges or whether their values rose, fell or stayed the same since last year.

Under the tax rates proposed for 2013-2014 in each of the county's taxing areas, local governments and special assessment districts would generate about \$2.9 billion in revenue next year, up 3.5 percent from this year's level of \$2.8 billion, The Post's analysis found.

The proposed tax rates are the maximum that local governments have set for the budget year that begins Oct. 1 and are used to calculate the preliminary tax notices that the property appraiser's office sent out last week. Taxing authorities can lower their rates before finalizing their budgets in September, but cannot increase them unless they notify all taxpayers in the taxing district by first-class mail of a new proposal.

Some governments have already vowed to lower their rates, meaning that the taxes paid countywide probably will be less than the \$2.9 billion projected.

West Palm Beach commissioners, for example, said last week that they planned to keep the city's tax rate and fire fee unchanged from the previous year, after previously setting the city's maximum tax rate at \$9.55 for every \$1,000 of taxable value, up from this year's rate of \$8.55. The commission also had earlier proposed tripling the county's fire assessment fee.

As a result, West Palm Beach spokesman Elliot Cohen said the tax amounts included on preliminary notices are "not indicative" of what city residents will pay.

Cohen said the city was facing an \$8.5 million shortfall when it set the maximum rate. Since setting the preliminary rate in July, city officials have reduced the deficit to \$2.5 million, he said.

"Setting the maximum has no effect on our residents, but it gives us the flexibility to do responsible budgeting," Cohen said. "We are the largest city in the county. We could not afford to have locked ourselves into any kind of low rate early."

Property Appraiser Gary Nikolits' office mailed the preliminary tax notices on Thursday to all 642,400 property owners. Final bills are mailed by Nov. 1.

West Palm Beach was one of 13 cities and towns that tentatively approved rate hikes. Eighteen have proposed keeping their tax rates flat. Five others will lower rates.

Without rate reductions, residents in the Jupiter Inlet Colony taxing district would see the largest increase in their total tax bill, The Post found.

The median tax hike for the district's 244 property owners would be \$1,035 under the proposed rates, The Post's analysis shows. That amount includes taxes and fees paid, not just to the town, but to all taxing authorities and special assessment districts listed on the town residents' tax notices. These include entities such as the county government, School Board and Florida Inland Navigation District.

The biggest factor in the increase, however, is that the town of Jupiter Inlet Colony set its tentative tax rate at \$5.95 per \$1,000 of taxable value, up 26.6 percent from last year. Town officials have said the rate hike is needed to cover the rising cost of fire-rescue service.

Meanwhile, countywide property values have continued to rise.

Countywide market values climbed by 6.3 percent in 2012, growing to \$162.3 billion on the first day of 2013 from \$152.7 billion on the first day of 2012, according to data received from the property appraiser's office and analyzed by The Post.

As a result, taxable values have also climbed.

Nikolits said the market values assigned by his office are typically 85 percent of what the property would actually sell for.

The Post found market values in Loxahatchee Groves slid last year, while those just to the north in The Acreage climbed.

Market values along the State Road 7 corridor in the county's Agricultural Reserve also rose last year, The Post found. Nikolits said new housing developments made the area more desirable to buyers.

Former Loxahatchee Groves Vice Mayor Dennis Lipp, a government budget watchdog and member of the Taxpayer Action Board, said governments need to do more to cut spending.

Even if local governments keep their property tax rates flat, they still collect more money because of rising property values, he said.

“More governments really need to live within their means as opposed to spending every nickel they can get their hands on,” Lipp said. “It is the same old story. Government grows the size that it can possibly grow to.”

Nikolits said Friday that only four of the county’s taxing authorities — Loxahatchee Groves, Briny Breezes, the Children’s Service Council and the Health Care District — have reduced their tax rates to a level that will generate less property tax revenue than they collected this year.

The rest of the taxing authorities would generate as much or more revenue under the proposals, Nikolits said.

The hikes come after years of falling property values removed billions from the county’s tax rolls. That plus state-mandated tax caps forced local governments to slash spending.

“The disappointing part is now the fiscal restraints seem to be off,” Nikolits said.

Countywide market value

2012: \$152.7 billion

2013: \$162.3 billion

Increase: \$9.6 billion, or 6.3 percent

Countywide proposed tax bills and special assessments

2012: \$2.8 billion

2013: \$2.9 billion

Increase: \$98 million, or 3.5 percent

Parts of PBIA flyover failed within two years

WEST PALM BEACH – As it rushed to finish building the \$156 million airport flyover in the early 2000s, contractor Modern Continental South substituted Styrofoam for grout, didn't bolt some bridge components together and failed to fully install others.

The state paid the contractor a \$1.4 million incentive to meet deadlines before it handed over a defective interchange that is costing the contractor's insurer \$1.8 million to repair. The bridges never posed a risk of collapse, state officials say.

Still, they say, repairs could be much more costly later if the joints are not replaced now.

As a result, three of the ramps that connect Interstate 95 with Palm Beach International Airport are closed for repairs, with another ramp to close in September.

Modern Continental's parent company, which figured prominently in one of the nation's biggest road construction projects — Boston's disastrous Big Dig tunnel — filed for Chapter 11 bankruptcy in 2008. That was four years after completing the flyover.

State officials blame poor workmanship for causing the bridges' 24 expansion joints, critical components between concrete segments, to fail. The joints are expected to last at least 25 years, said an official with the company that makes the joints, Watson & Bowman.

The closings come nine years after the interchange opened. It took the state two years to discover the problems, four to negotiate with the contractor's insurer for the repair money and another three to start the work.

While the defects pose no danger, the joint repairs are needed now, state officials say. When joints fail "the ride might be a little rough across the joint, but it's not like we're gonna have a catastrophic failure," said Alberto Sardinias, a manager with the Florida Department of Transportation.

The flyover, which carries about 16,000 vehicles a day, drew heated opposition before its construction. The project pitted neighbors who feared it would encourage airport expansion against business and tourism leaders who said it would make the airport more accessible.

The contractor's parent company, Modern Continental Construction, filed for bankruptcy after its work on one of the nation's largest highway construction projects turned tragic. A passenger in a car died in Boston's Big Dig when 26 tons of concrete panels gave way in a highway tunnel and crushed her vehicle. In a separate Big Dig incident, a tunnel wall blew out and flooded a section of the highway.

Federal prosecutors filed criminal charges against the company over those incidents and others in 2008. The company filed for bankruptcy the next day. The Massachusetts attorney general reached a \$21 million settlement later that year for the ceiling collapse and later the company was found guilty on 39 counts of making false statements. But the company's path to insolvency was well underway before it finished the PBIA flyover.

In 2001, about a year after construction began, federal court records show that Modern Continental failed to pay some debts. By 2003, it lost the ability to get new contracts because it couldn't get an insurer to back its bonds. The flyover was slated to finish late that year, but dragged on until December 2004.

Modern Continental installed most of the joints in the final 15 months, project officials say.

Both the \$15 billion Big Dig, which moved a stretch of interstate highway underground, and the PBIA flyover were hampered by delays, rising costs and construction mistakes.

The flyover was expected to cost the state \$54 million in 1990, but the final cost was almost three times as much. Numerous problems delayed the project, including high turnover in local company management and insufficient money to pay subcontractors.

Expansion joints fill gaps between bridge segments to allow them to expand and contract as temperatures change. The joints, made of metal and synthetic rubber, prevent water or debris from damaging critical — and expensive — steel cables underneath.

In 2006, inspectors found crucial components inside the joints had deteriorated because they were improperly installed, Sardinas said. Among other mistakes, the contractor had failed to bolt together the metal rails that motorists drive over — which put them out of alignment — and didn't fully install the components that anchor the joints to the bridge segments.

Inspectors also found that at least one of those components was supported by Styrofoam — which is used during an intermediary stage of joint installation — as opposed to the much harder, concrete-like grout that was required.

Components on the surface had cracked and rusted as a result and water seeped into the bridge. But the problems were detected early, Sardinas said, and there was no evidence that the cables had corroded.

Four years after the discovery, Modern Continental's bonding company — the equivalent of an insurer — agreed to pay for the repairs, said Calvin Johnson, FDOT's lead attorney in the case. It took another few years before the department hired a contractor and repairs could begin.

Kiewit Infrastructure South, which is replacing the old joints, closed the airport ramps in mid-June and is expected to finish by October. Unlike when the first joints were installed, a representative of the manufacturer is overseeing the installation, Sardinas said.

The flyover was built to meet steep airport growth projections that have only partly materialized. Annual numbers of passengers at PBI reached the projected 5.9 million in 2009, but decreased to 5.6 million since. Airport Director Bruce Pelly blamed the decline on the shrinking airline industry.

Pelly said this summer's initial closure has not affected passenger volume. "It's obvious it takes people coming from I-95 longer to get to the airport," Pelly said. But the bridges are worth it, he said. "Yes, it makes us a better airport and it makes us more convenient."

Staff researcher Niels Heimeriks contributed to this story.

Ramps under repair

Shoddy construction forced transportation officials to close ramps connecting Interstate 95 and Palm Beach International Airport this summer.

1. Closed July 1; to reopen by Sept. 1
2. Closed Sept. 1; to reopen by Oct. 1.
3. Closed June 17; to reopen by Sept. 1
4. Closed June 17; to reopen by Sept. 1
5. Closed April 25; reopened June 14
6. Closed April 23; reopened June 14

Closings could last longer depending on weather and unforeseen circumstances.

What is an expansion joint?

Expansion joints fill a gap between bridge segments to allow expansion and contraction.

They are made of metal and synthetic rubber. They prevent water or debris from damaging critical steel cables below.

'Geek culture' reigns as PalmCon draws 2,000 to convention center – Sept. 21, 2013

<http://www.mypalmbeachpost.com/news/news/local/geek-culture-reigns-as-palmcon-draws-2000-to-conve/nZ4fY/http://www.mypalmbeachpost.com/news/news/local/geek-culture-reigns-as-palmcon-draws-2000-to-conve/nZ4fY/>

WEST PALM BEACH —

Josh Dykstra wanted to dress as a superhero for Halloween, but he couldn't find one with a beard, like himself. So, naturally, he made his own superman character — Facebeast — who fights crime using the shape-shifting powers of his beard.

Now, Facebeast is a full-fledged character in a series of comic books, self-published by Dykstra, 27, of Tampa, and his girlfriend, Kate Carlton. They came to PalmCon Saturday to sell their comic books, and were two of about 2,000 people at the Palm Beach County Convention Center for the annual gathering of “geek culture” aficionados.

In addition to self-publishing comic book writers, this year's PalmCon was host to a broad assortment of people who came to show off costumes they'd worked months to create, sell comic books they'd collected over decades, or to share their passion for their chosen hobby.

Michale Geressy's path to PalmCon started with what appeared to be two misfortunes: getting a divorce and getting laid off from his job. But having no one to share his generous severance package with, he ended up remodeling a 1959 Cadillac to look exactly like the car used by the Ghostbusters. On Saturday, that car was inside the convention center, with Geressy and other members of the Carolina Ghostbusters standing by in full Ghostbuster gear.

With the exception of Geressy, most of the attendees were from the area between Orlando and Fort Lauderdale, Martin Pierra, who founded PalmCon three years ago, said.

Among the locals was Robert Wilson, who has been collecting comic books since 1962, when his mother bought him two comic books at a laundromat. Now, he estimates he has about 25,000 in his collection, most of which were for sale at PalmCon, filling box after box on table after table. He said he's been collecting comic books for two reasons: because the world needs heroes and because comic books are a way of preserving history.

"When you look at an old comic cover that you read when you were 8 years old, you harken back to that time," Wilson said.

Attendance rates have doubled each time since the first PalmCon in 2011, from 500 to about 2,000 this year. Pierra's goal was to have at least 1,000. He moved it to the Palm Beach County Convention Center because it had outgrown the previous venue, the American Polish Club on Lake Worth Road in Greenacres.

"We couldn't handle the people we had last year," Pierra said.

Although Pierra is waiting for the official attendance tally before he makes the decision, he's almost certain PalmCon will be held at the convention center next year as well.

As he tried to cheer up his 5-year-old son — Micah was dressed as Batman — after he was denied a \$50 Transformer toy, Matt Murnane, of West Palm Beach, explained the importance of comic book superheroes in his children's education:

They "help make doing the right thing cool," he said.

Delray man charged with theft from travel agency that employed him – Aug. 19, 2013

<http://www.mypalmbeachpost.com/news/news/local/delray-man-charged-with-theft-from-travel-agency-t/nZS2k/http://www.mypalmbeachpost.com/news/news/local/delray-man-charged-with-theft-from-travel-agency-t/nZS2k/>

BOCA RATON — A Delray Beach man was arrested Thursday after the owner of a corporate travel company alleged he had diverted more than \$130,000 from the company's accounts.

John Pires, 51, faces two counts of grand theft and one of organized fraud, city police said. He was released from the Palm Beach County Jail on Friday on \$3,000 bond.

Pires, who was a salaried employee at Corporate Travel Sales Bureau of Boca Raton, allegedly diverted travel bookings from the company to other travel agencies, according to a probable-cause affidavit.

Neither the clients nor the company gave Pires permission to do so, the affidavit said.

He then split the commission with the other agencies.

The owner of the company, Joan Harvey, discovered the scheme after Pires resigned in 2008. Through an internal investigation she found that he had been diverting commissions since 2005.

Harvey sued Pires in 2009 and received a judgment of \$100,000.

The three charges filed against Pires are for fraud and grand theft committed in his last year at the company.

Farmworker advocates blanket Lake Worth, Palm Springs in support of immigration reform – Aug. 18, 2013

BY ERIN SHAW, FEDOR ZARKHIN

[http://www.mypalmbeachpost.com/news/news/local/farmworker-
lake-worth-palm-
sprin/nZR9C/http://www.mypalmbeachpost.com/news/news/local/farmworker-
advocates-blanket-lake-worth-palm-sprin/nZR9C/](http://www.mypalmbeachpost.com/news/news/local/farmworker-advocates-blanket-lake-worth-palm-sprin/nZR9C/http://www.mypalmbeachpost.com/news/news/local/farmworker-advocates-blanket-lake-worth-palm-sprin/nZR9C/)

LAKE WORTH —

Isabel Vinent’s voice reverberated through the streets as her raspy loudspeaker shouted “What do we want?” at the people streaming down the streets. “Immigration reform,” scores of voices answered. “When do we want it?” she asked.

“Now!”

About a hundred people marched from the Guatemalan Maya Center to downtown Lake Worth and to the Lake Worth City Hall tonight as part of a statewide campaign to increase awareness of immigration issues and to pressure members of Congress to pass a reform bill when they resume the session in September.

At city hall, the marchers sang and listened to organizers speaking for reform.

“I would like a million friends so that together we could sing louder,” a young man sang in Spanish while playing the guitar as Frank O’Loughlin, executive director of the Guatemalan Maya Center, held the loudspeaker for him.

The songs and the speakers were at times drowned out by the honking of the passing cars on South Dixie Highway, responding to a “Honk for immigration reform” sign held by one of the marchers.

Victor Herrera, of Greenacres, stood on the sidewalk with his friend holding a large cloth with a monarch butterfly painted on it.

Herrera, who is a sophomore at Palm Beach State College and is a member of Students for Equal Rights, explained that the monarch represents immigration reform because yearly the butterflies migrate from Mexico to California, but without problems at the border.

“Between Mexico and America it’s just a line,” Herrera said.

By nightfall candlelight was sprouting across the crowd in front of city hall.

Earlier Sunday, Vincent nudged past the Little Debbie snack cakes at a tiny convenience store on Lake Worth Road and thrust a flier for immigration reform into the owner’s hands.

“Our immigrant community consumes in your store, and we would expect that you support us,” she said, in a voice whose pleasant tone didn’t hide its firmness.

In Spanish, she asked him whether her group — the Remember November Caravan — could put up a poster in the window. He agreed.

Then she asked him for a donation. He declined.

So went the early part of the caravan’s effort Sunday to rally support for the reform bill now stalled in Congress.

The group represents six nonprofits pushing for reform. Led by the Farmworkers Association of Florida, the caravan is visiting all 27 congressional districts in the state in 36 days, traveling in a big purple van with “Remember November” painted on the windows.

The name invokes the significant role Hispanics played in the 2012 presidential elections. Their wish? That residents and business owners would call their representatives in Congress and urge them not to scrap the bill the Senate approved this summer.

“I’m really amazed at their commitment and sacrifice,” Vinent, the deputy director of the Florida Immigration Coalition, said of the caravan volunteers, who number 30. Seven traveled with her Sunday to Lake Worth and Palm Springs.

“Some will sleep on the floor tonight. They will eat what people can donate. They are putting their lives, work and schooling on hold to support immigration reform.”

In June, the Senate passed a bill calling for increased focus on workplace enforcement and access to citizenship for immigrants who came to this country illegally. Members of the Republican-dominated House of Representatives have said they do not support a path to citizenship for those immigrants.

David Benson, who joined the group in Apopka 10 days ago, is the only one in the caravan who didn’t speak fluent Spanish. He doesn’t mind.

“I can’t see myself being anywhere else,” he said. Benson has lived in a Hispanic neighborhood in Dade City for the past three years and has three Hispanic tenants living in his home.

“I’ve seen what they go through, and Lord almighty, it isn’t pretty,” Benson says.

Benson said he had a friend come to him who was badly beaten and bloodied, but begged him not to call police for fear of being deported. Benson also knew a woman who was beaten and raped and still did not want the authorities called.

“If they can’t speak, I’ll do my best to be their voice,” he said.

After visiting another convenience store, a barber shop and a clothing outlet, the group crossed Lake Worth Road to visit a local grocery, C&D Produce.

“See this?” Jose Amateco asked as he pointed to the fresh produce. “Farmworkers picked this.”

Amateco, another caravan member and a senior biology student at St. Leo University outside Tampa, said he has long supported farmworkers’ rights to decent wages and acceptable working conditions.

He’s volunteered for the nonprofit Farmworker’s Self-Help since high school. He said he wants to help others in another way some day, as a doctor. Amateco is a Dreamer — a beneficiary of the Development, Relief and Education for Alien Minors Act — which qualifies undocumented youth eligible for a six-year, conditional path to citizenship that requires completion of a college degree.

“I had the opportunity to go to school and get an education. I had a lot of help, but so many others don’t,” Amateco said.

Among the states, Florida has the third-most people who could benefit from a renewable reprieve from deportation, according to the Immigration Policy Center. In

Palm Beach County, about 30,000 people are migrant and seasonal farmworkers, the largest number of any county in Florida, according to a national study.

After an afternoon in the hot August sun talking with business owners and handing out fliers to passers-by, Vinent gathered the group before they were off to speak at Sacred Heart Catholic Church downtown.

A volunteer delayed her request to hand out fliers to one more person.

“Muy bien,” Vinent said after a pause. “De le. De le.” Tell them.

Some young people, disappointed by Zimmerman verdict, say prejudice is part of their daily lives – July 14, 2013

<http://www.mypalmbeachpost.com/news/news/crime-law/some-young-people->

[disappointed-by-zimmerman-](http://www.mypalmbeachpost.com/news/news/crime-law/some-young-people-disappointed-by-zimmerman-)

[verdic/nYqZ5/http://www.mypalmbeachpost.com/news/news/crime-law/some-young-](http://www.mypalmbeachpost.com/news/news/crime-law/some-young-people-disappointed-by-zimmerman-verdic/nYqZ5/)

[people-disappointed-by-zimmerman-verdic/nYqZ5/](http://www.mypalmbeachpost.com/news/news/crime-law/some-young-people-disappointed-by-zimmerman-verdic/nYqZ5/)

LAKE WORTH — Alphonso Carter saw nothing out of the ordinary in the George Zimmerman trial verdict.

“I never go in a white neighborhood,” the 15-year-old from Delray Beach said. “They always think we’re gonna steal.”

Other young people out and about Sunday in Lake Worth and West Palm Beach — some of them the same age as Trayvon Martin, the 17-year-old black teenager Zimmerman was acquitted of killing late Saturday — saw the verdict the same way.

For Carter and other Palm Beach County youths the Zimmerman trial was all about race, and the prejudice they see in Zimmerman's decisions the night Martin died is a part of their daily lives, they say.

Andre Marshall of Lake Worth simply didn't see the logic behind the jury's verdict, saying that if he was in a position similar to Martin's, he would have fought back.

"I've got to protect myself," Marshall, 25, said. "If you're trailing me and you're not a police officer, then I don't know who you are."

Ellison Robins, also of Lake Worth, distilled the case to what he saw as its bare elements.

"I just see a white man shooting a black boy and getting off," Robins, 20, said standing outside of the predominately black Lake Worth Church of God after Sunday's service.

And both the verdict and the culture he lives in teach him to be careful. He said he has to second-guess how he carries himself in public and make an effort for people to trust that he's not a "bad kid."

"It just makes it hard," Robins said. "After a while you can put people into a position where they want to do wrong."

But some say it's time for the media to move on.

"I swear to God, I thought this was over last year," Russell Read, 19, of Coconut Creek, said.

Standing outside the Cruzan Amphitheatre, awaiting a performance by New Orleans rap artist Lil' Wayne, Read said the media are distracting the public from more important events around the world. And besides, not everything was known about the night Martin died, he believes.

“It’s hard for me to believe that Trayvon Martin wasn’t doing something wrong,” he said. “I’m sure he was looking around, doing something sketch.”

Stories I contributed to:

Licenses can be revoked but offenders keep driving, sometimes with deadly results

BY ALEXANDRA SELTZER

<http://www.mypalmbeachpost.com/news/news/licenses-can-be-revoked-but-offenders-keep-driving/nZ9nJ/http://www.mypalmbeachpost.com/news/news/licenses-can-be-revoked-but-offenders-keep-driving/nZ9nJ/>

How was it that Steven W. Barnes, a person with four DUI convictions and a permanently revoked license, was driving?

Eulalia Gaspar kept asking herself that question last week as she arranged to bury her 3-year-old daughter — and she wasn’t alone.

Priscila Perez Gaspar was killed Monday when Barnes, 56, lost control of the truck he was driving and hit her as she walked on the sidewalk along Kirk Road in suburban Lake Worth. That same day, six more Palm Beach County residents driving with either suspended or revoked licenses were caught, according to Palm Beach County Jail records.

But while a judge may ban someone like Barnes from driving, no one can physically stop them.

“The laws are good. The problem you have is that you can revoke somebody’s license but that doesn’t necessarily stop someone from driving,” Deerfield Beach-based attorney Marcelo Lescano said.

“The laws can be very strict and impose these penalties and say, ‘Look, you’ll never legally drive again.’ They don’t have the right to drive but yet they have the power to.” State attorney spokesman Mike Edmondson says prosecutors are powerless “until the legislature decides that the penalties are going to be higher for those types of offenses.” Since January, the Palm Beach County State Attorney has charged 7,718 for driving under either a suspended or revoked license. In 2012, there were 16,439 county drivers who received traffic citations for driving with a suspended or revoked license and 195,467 in the state, according to the state’s Department of Highway Safety and Motor Vehicles.

Almost half of the cases the state attorney’s office handles involve a defendant driving with a suspended license.

“It’s common practice because they don’t care if they have a license or not,” said Ellen Roberts, who helped create the office’s traffic homicide unit before retiring last year.

“You tell them in court that their license has been permanently revoked, which means you can never operate ever, ever again, and then you turn around and you find them in the car.”

Sometimes, it’s as quick as minutes from the gavel dropping after they leave the courthouse.

A driver’s license can be suspended for reasons including failure to appear on a traffic summons, lack of auto insurance, not completing traffic school, suspension for points and a traffic violation resulting in injury or death.

But more seriously, a license can also be revoked depending on a drunken-driving conviction.

An ignition interlock — one that won't let the car start if the driver's blood-alcohol level is too high — is required after a second DUI. In some states, including Connecticut, Arizona, New York and Louisiana, judges order the interlock after one conviction. Other states including Montana, New Jersey and Rhode Island don't suspend a license on a first conviction, according to data from the Governors Highway Safety Association.

Previously in Florida a person convicted of four DUIs could reapply for a license after 10 years of revocation. Today that no longer is allowed.

Barnes, who now lives in suburban Lake Worth, was convicted of four DUIs between 1979 and 1996 — the year his license was permanently revoked. Despite not having a license, he was caught behind the wheel again in 1999, 2002 and 2008. He spent days in jail on two of those convictions and a year for one of them.

Barnes, who was ordered held without bail Saturday at the Palm Beach County Jail on charges of vehicular homicide and driving without a license in a way that causes injury or death, could not be reached for this story.

'Very hard to control someone'

Barnes isn't alone. Just about a year ago and less than two miles south of where Priscila was killed, William Melendez, who had a suspended license, lost control of the Chrysler Sebring he was driving and collided with a 1997 Mercedes driven by David Odom.

Odom was killed and Melendez was arrested on charges of DUI manslaughter, failing to render aid, vehicular homicide, DUI property damage and driving with a suspended license. He was found guilty of DUI manslaughter and failing to render aid.

The state attorney, however, did not act on the suspended license and dropped the other charges.

On Aug. 4, Palm Springs police arrested Julian Lopez, a habitual traffic offender. Lopez was convicted three times of driving with a license that was either suspended, revoked, canceled or disqualified, all within a five-year period. In March his driving privileges were revoked until March 2018.

“It’s very hard to control someone,” Palm Beach County Assistant State Attorney Andrea Robinson said. “What’s not to say they don’t hop in their wife’s car?”

State Rep. Irv Slosberg, who became an advocate for traffic safety after his daughter was killed in a 1996 crash, called Barnes a “menace to society” and said the only way to keep him from threatening public safety again is for him to go to jail.

“I can imagine what she’s going through because I lost my daughter,” Slosberg said. “I’m walking in her shoes. The first couple days, couple of weeks, couple of months, couple of years, they’re unbelievable.”

Slosberg, D-Boca Raton, met with Gaspar on Friday.

“I want justice, not for me but for Priscila. This is not fair,” Gaspar said. “This man should not have been on the street that day. He was. Now my daughter is gone.”

Staff researchers Niels Heimeriks and Fedor Zarkhin contributed to this story

In final face-off with rogue Goodman juror, judge likely to end contempt case with jail time – Sep. 28, 2013

<http://www.mypalmbeachpost.com/news/news/crime-law/in-final-face-off-with-rogue-goodman-juror-judge-l/nZ9Nt/http://www.mypalmbeachpost.com/news/news/crime-law/in-final-face-off-with-rogue-goodman-juror-judge-l/nZ9Nt/>

WEST PALM BEACH — On Monday, former John Goodman juror Dennis DeMartin will come face to face, perhaps for the last time, with the man who holds DeMartin’s freedom in his hands.

The final hearing in DeMartin’s contempt of court case will pit Palm Beach County’s most infamous juror with a man that is now the area’s most powerful judge. And based on Chief Circuit Judge Jeffrey Colbath’s track record as well as recent fireworks in the case, DeMartin appears unlikely to escape the showdown without looking at time behind bars.

“My crystal ball says the court will convict and incarcerate Mr. DeMartin,” Robert Gershman, who represented the 70-year-old pro bono along with Joseph Walsh, said Friday after the last preliminary hearing in the case.

Colbath’s decision in May to charge DeMartin, a Delray Beach retiree, with contempt for derailing Goodman’s first DUI manslaughter trial added one more layer of controversy — and frustration — to the ongoing courtroom drama surrounding the death of 23-year-old Scott Wilson.

DeMartin revealed in the third of three books he’s published since Goodman’s high-profile March 2012 trial that his ex-wife had once been arrested for DUI — a fact he’d

failed to disclose to the court when he was a prospective juror. Colbath used that as the basis for overturning Goodman's conviction and 16-year prison sentence, and also for DeMartin's contempt charges.

Aside from Goodman's upcoming retrial, DeMartin's case is the only high-profile matter the judge has kept since he became the 15th Circuit's chief judge, an administrative role he took over from Circuit Judge Peter Blanc this summer.

And if the tension in the DeMartin case wasn't apparent before, a preliminary hearing began Friday with Colbath calling Gershman and Walsh unprofessional for throwing out a last-minute motion to have DeMartin declared incompetent.

"You filed the motion to have me recused from the case pretty quickly, so why did this one take so long?" Colbath asked, referencing a filing a week earlier in which the attorneys accused Colbath of having a personal bias against DeMartin.

Gershman immediately renewed his motion to have Colbath off the case, a request Colbath made the attorney put in writing before he unceremoniously denied it, just as he had their first request.

Exchanges like these have given the public a glimpse of a much tougher side of the judge who in the past has handled highly publicized cases with a mostly affable exterior.

Colbath is a second generation jurist, the son of former Palm Beach County Chief Circuit Judge Walter Colbath. The younger Colbath was appointed to the county court bench in 1992, where he started a tradition of playing volleyball at the Palm Beach County Jail on Friday afternoons with the prosecutors and public defenders who litigated cases before him.

'Word of the day'

On the circuit court bench, Colbath often peppered morning docket calls with a daily “word of the day” – a tradition he didn’t halt when he presided over trials like Goodman’s and the murder solicitation trial for the infamous Boynton Beach newlywed Dalia Dippolito.

In an interview last year, Colbath said he started it to break the monotony of what was often mundane court hearings, and also to offer a temporary distraction in tense cases.

“I was surprised at how I’d be presiding over murder trials, and the defendant would stop and ask: ‘Hey judge, what’s the word of the day?’ ” Colbath said at the time.

Until the DeMartin case, Colbath’s tough words in cases usually came at the very end. In Dippolito’s case, for example, he called her “pure evil” only in the moments before he sentenced her to 20 years in prison for trying to hire a hitman to kill her husband.

And though he at times appeared to handle DeMartin gently in the three post-trial hearings exploring allegations of misconduct in Goodman’s case, his thoughts on the juror’s behavior sprang up from the pages of the ruling overturning Goodman’s conviction based on DeMartin’s actions.

Colbath also added to DeMartin’s charges a drinking experiment the retired accountant conducted the night before he and five others convicted Goodman — an experiment he detailed in the pages of his self-published juror tell-all, *Believing the Truth*. Though Colbath had expressly forbidden jurors to conduct their own experiments in the case, DeMartin said he drank three vodka and tonics within the span of 90 minutes to see if Goodman would have been impaired based on testimony in the case.

Goodman’s attorneys had made the drinking experiment their first avenue of appeal, but Colbath — who did say he felt it was misconduct — appeared to forgive the misstep by denying Goodman a new trial based on the experiment alone.

Bumbling may not save juror

From the beginning, local attorneys speculating on the case have said that the possibility of Colbath sending the elderly juror to jail is real, despite his often appearing hapless while answering questions in his post-Goodman trial hearings.

Last week, well-known defense attorney Michael Salnick, who represented Dippolito through her trial, said if nothing else, Colbath’s sentence will reflect the seriousness of the charges against DeMartin, which Salnick called a slap in the face to the criminal justice system.

“Based on what he allegedly did, he could have robbed this man (Goodman) of 16 years of his freedom, all based on his misconduct,” Salnick said. “I think the judge is going to look at that, and if he’s found guilty, the judge is going to sentence him based on the facts.”

Judges in Palm Beach County have rarely charged people with criminal contempt of court charges. According to a Palm Beach Postanalysis of local criminal court cases since 2006, only 40 people have been prosecuted with the same contempt charge DeMartin now faces. Most of those cases were handled by other judges, but a glimpse into Colbath’s record with criminal contempt comes in a 2004 case.

That was the year Colbath inherited from another judge the contempt case against Jay Gordon, a man who found himself in trouble with the court for allegedly violating an injunction in a divorce proceeding.

According to court records in the case, Gordon at one point had asked to have the case moved out of Palm Beach County, claiming then-newly elected Circuit Judge Martin Colin, once a main witness in his divorce case, was his ex-wife's "boyfriend."

After a hearing in late 2004, Colbath found Gordon guilty of contempt. But he didn't sentence him right away. Instead, Colbath sent him to jail for a week until he could decide. When Gordon came back before him a week later, Colbath decided to release him but put off his sentencing again for another two months.

In December of that year, he sentenced Gordon to a year of probation, which included a 30-day jail sentence to be served on weekends. An appellate court later overturned Gordon's conviction, saying that Colbath should not have conducted the hearing or sentencing without appointing a public defender to represent him.

Gordon was decades younger than DeMartin, however, whose lawyers say he will temporarily lose certain Social Security benefits if Colbath decides to jail him for more than 30 days.

Further complicating matters, Gershman and Walsh last week filed a notice with Colbath asking the judge to allow them to get doctors to test whether DeMartin was mentally competent to assist them with his defense. Walsh, after Friday's hearing, hinted that he and Gershman would be filing more details about DeMartin's condition ahead of Monday's hearing.

But even before last week, Walsh said in an interview, the whole ordeal was having an adverse effect on DeMartin’s health — manifesting in a flare-up of heart ailments, which DeMartin has battled for years.

Health issues aside, DeMartin – who also complains of memory loss – has not been a sympathetic figure to many in the community, including several members of the public who have written letters to Colbath asking him not to show leniency.

“DeMartin should pay for his exploitation of our system, no matter his age,” Frank Simmons, one of DeMartin’s fellow Delray Beachers, said in a handwritten note to Colbath in June.

Data Interactive Editor Niels Heimeriks and Staff writer Fedor Zarkhin contributed to this story.

West Palm woman charged with DUI, leaving scene after crash in Lake Worth – Aug. 18, 2013

<http://www.mypalmbeachpost.com/news/news/local/west-palm-woman-charged-with-dui-leaving-scene->

[aft/nZR9J/http://www.mypalmbeachpost.com/news/news/local/west-palm-woman-charged-with-dui-leaving-scene-aft/nZR9J/](http://www.mypalmbeachpost.com/news/news/local/west-palm-woman-charged-with-dui-leaving-scene-aft/nZR9J/)

WEST PALM BEACH — Jessica Marksz needed a drink.

That, she told Palm Beach County sheriff’s deputies, was why she had left her car and walked to a nearby Publix Super Market on Thursday after a crash in which she is alleged to have driven the wrong way down a one-way street and hit a car.

Deputies weren't impressed, especially when one caught up with her picking up a bottle of wine.

They arrested her on charges of DUI with property damage and leaving the scene of an accident. She was released later that day from the Palm Beach County Jail after on \$3,000 bond.

Marksz, 35, of West Palm Beach, told Deputy Ashley Hegland she had one pint of Guinness at Igot's Martiki Bar in Lake Worth, though Hegland observed that her speech was slurred and that she had trouble walking.

Marksz is alleged to have caused the accident at the 200 block of North J Street in Lake Worth

Interactive Graphics

Property values maps landing page

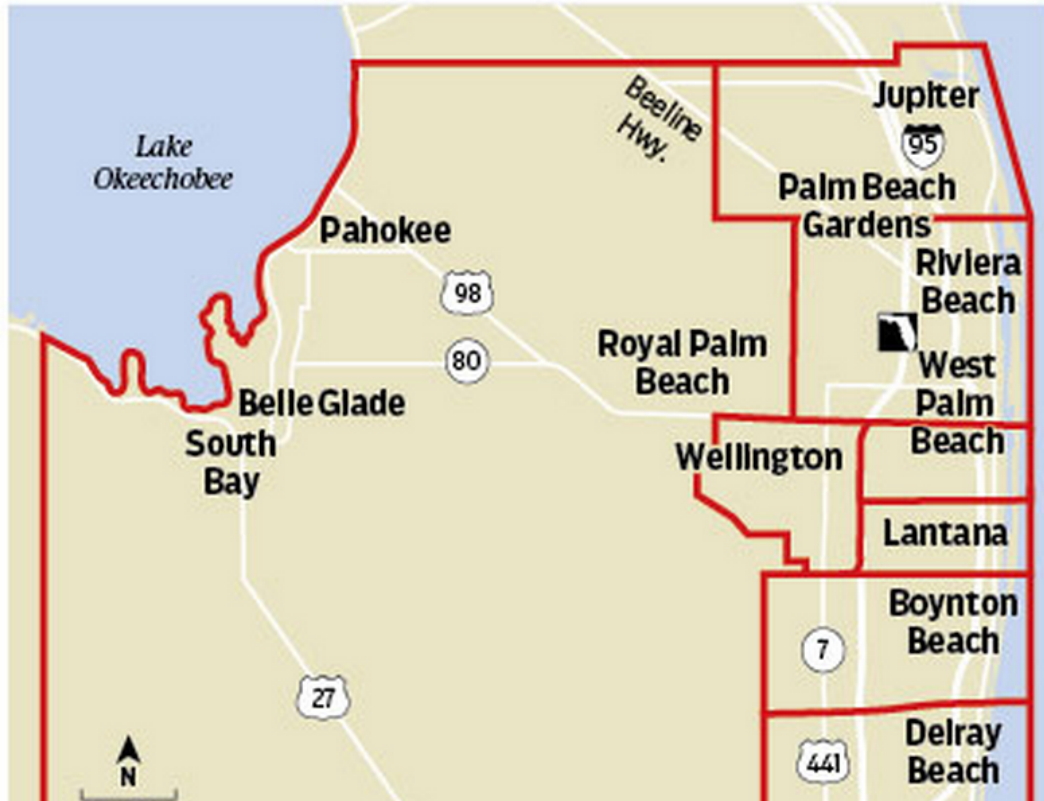
How does your property compare?

By Fedor Zarkhin

Use these interactive maps to check how your property's market value and proposed taxes compare with others' across the county.

These maps were created from Palm Beach County Property Appraiser's 2013 tax notice data.

INSTRUCTIONS: Click on a sub-map for an interactive map of that area.



Property values map example

Address Search

Enter an address or an intersection

within

Name search

Use with address search for best results

Last name then first name, no commas

2013 Market Values

Value Change

● Down

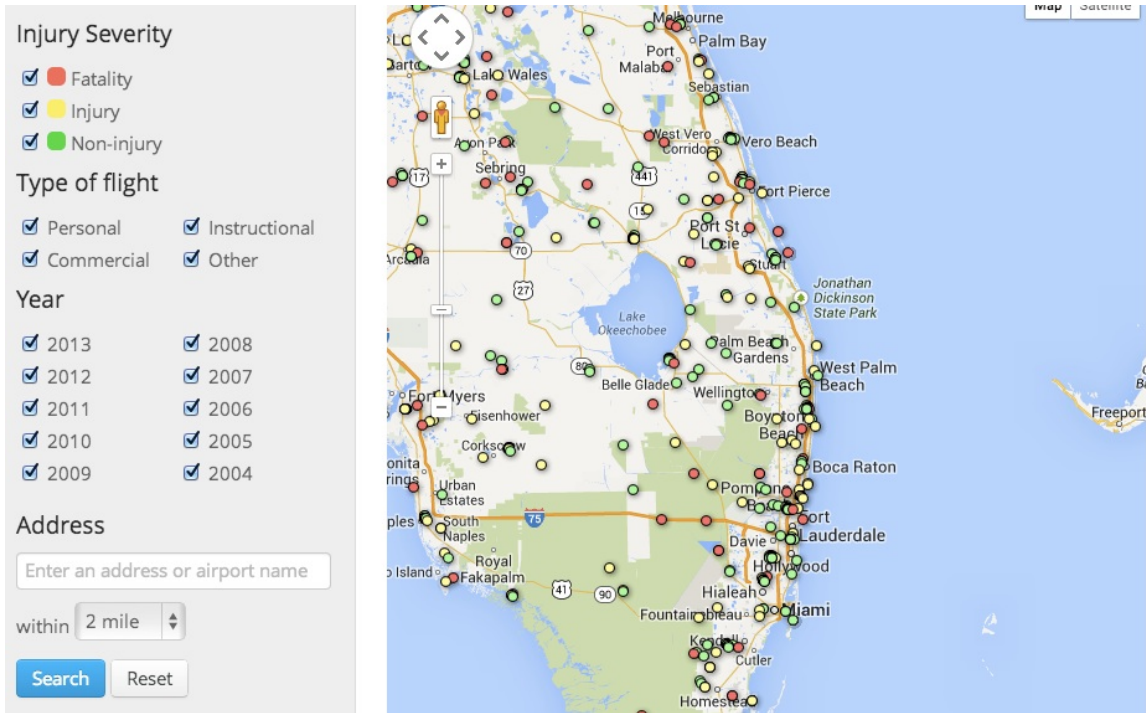
● Up

● No change

63,351 properties found



Aviation Accidents Map



School Grade Map

2013 School Grade

A B C D F I

School type

All

(Click 'Search' to see results)

Free or reduced lunch 0 - 100

0 % 100 %

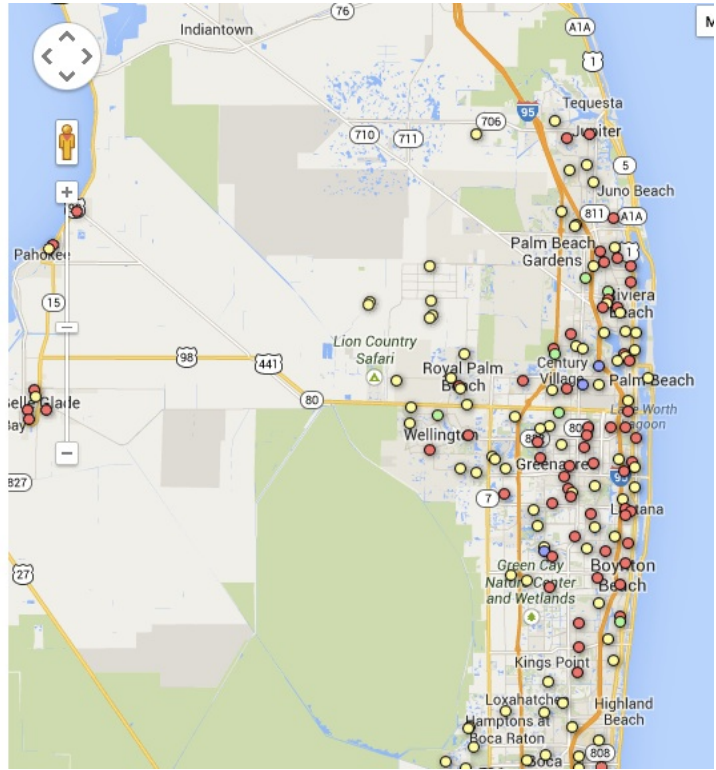
Search for schools near you

Enter an address or an intersection

within 1/2 mile

Search Reset

153 schools found



Examples of cleaning and exploratory Queries:

```
UPDATE nondisp
SET FDATE=
(CASE WHEN f13<='13'
THEN
CONCAT("20",f13,"-",f14,"-",f15)
ELSE CONCAT("19",f13,"-",f14,"-",f15)
END)
---
SELECT count(*), firstname, middleinitial, lastname, PartyType FROM parties
GROUP BY firstname, middleinitial, lastname, PartyType
HAVING count(*)>1
ORDER BY count(*) DESC
SELECT stat, stat_descr FROM mainfile
WHERE stat_descr LIKE "%child%"
GROUP BY stat, stat_descr
SELECT count(*), dyear FROM mainfile
GROUP BY DYEAR
ORDER BY count(*) DESC
SELECT * FROM mainfile
WHERE stat LIKE "825.102%"
ORDER BY dyear DESC
```

```
CREATE TEMPORARY TABLE nofiles3 SELECT count(*) AS count, stat FROM mainfile
```

```
WHERE (disp = "no file" OR disp = "nolle prosse")
```

```
AND type ="CF"
```

```
GROUP BY stat
```

```
ORDER BY count(*) DESC;
```

```
CREATE TEMPORARY TABLE totals3 SELECT count(*) AS count, stat FROM mainfile
```

```
WHERE type ="CF"
```

```
GROUP BY stat
```

```
ORDER BY count(*) DESC;
```

```
SELECT ROUND(nofiles3.count/totals3.count*100,0) AS percent, nofiles3.stat,
```

```
nofiles3.count, totals3.count FROM nofiles3, totals3
```

```
WHERE nofiles3.stat=totals3.stat
```

```
AND totals3.count>1000
```

```
ORDER BY percent DESC
```

Examples of analysis queries:

Joining with medical license data:

```
CREATE TEMPORARY TABLE abusers
```

```
SELECT * FROM mainfile
```

```
WHERE stat LIKE "825%";

SELECT * FROM med, abusers

WHERE med.l_name=abusers.last

AND med.f_name=abusers.first

AND (med.m_name=abusers.middle

OR LEFT(med.m_name,1)=LEFT(abusers.middle,1)

OR (med.m_name IS NULL AND abusers.middle IS NULL)

)

---
```

Lottery Data

The screenshot shows a database query tool interface. On the left is a tree view of database objects, including a 'lottery' schema with various tables. The main window displays a query in the 'Query Editor' tab:

```
1 SELECT * FROM winners
2 LIMIT 1000
```

Below the query editor is a table with columns: Message, Result1, Profile, and Status. The 'Result1' column contains the query results, which are displayed in a table format with columns: ID, PDATE, LOCID, RETNAME, and RETSTREET.

ID	PDATE	LOCID	RETNAME	RETSTREET
1000030085	2000-01-04	048789	HARDEES OF TRENTON	1625 E WADE STREET
1000030121	2000-01-03	005246	CUMBERLAND #9563	3518 E CURRYFORD ROA
1000030222	2000-01-03	050344	WINN DDDE #2220	3565 N LECANTO
1000030234	2000-01-03	007195	THE WHISTLE STOP & THE	82685 OVERSEAS HIGHW
1000030241	2000-01-03	003217	141001877 CIRCLE K	1523 PINE ISLAND ROAD
1000030248	2000-01-03	032098	JR FOOD MART #227	102 W HIGHWAY 90
1000030259	2000-01-03	003044	ALBERTSON'S #4308 M	2770 W BAY DRIVE
1000030289	2000-01-03	045919	KWIK STOP FOOD STORE	3081 NW 64TH AVENUE
1000030337	2000-01-03	031863	PUBLIX #415	33343 US 19 NORTH
1000030371	2000-01-03	050860	SUPER STOP	4144 SW 64TH AVENUE
1000030526	2000-01-03	048430	PINE ISLAND AMOCO	8790 W COMMERCIAL B
1000040101	2000-01-04	011123	111004724 CIRCLE K	4285 NORTH COURTENA
1000040153	2000-01-04	014598	LIL CHAMP STORE #6149	701 S PARK AVENUE
1000040257	2000-01-04	052495	KWIK STOP	5701 N NEBRASKA AVEN
1000040264	2000-01-04	046843	4 POINTS MARKET	6950 LOX ROAD
1000040594	2000-01-04	015257	JOHNSON & JOHNSON #9	1996 HAMILTON AVENUE
1000040602	2000-01-04	003097	MARTIN OIL COMPANY	3260 BEACH BOULEVARI
1000050413	2000-01-05	048761	WEST MARKET	1015 W HIGHWAY 50
1000050414	2000-01-05	048344	SUPER STOP #802	221 N BABCOCK STREET
1000060228	2000-01-06	008450	BIG "B" LIQUORS	1820 SOUTH RIDGEWOO
1000060326	2000-01-06	048735	1507 7-ELEVEN #32274	605 S COURTENAY PARK
1000060557	2000-01-06	040171	CENTRAL GROCERY	948 S CENTRAL AVENUE

Examples of cleaning queries

UPDATE stores_copy AS st, (SELECT count(*),storen, storest, storec, LEFT(storez, 5) AS

zips, min(storeid) AS thething FROM stores_copy

GROUP BY storen, storest, storec, LEFT(storez, 5)

ORDER BY count(*) DESC)a

SET st.storeidfix=a.thething

WHERE st.storen=a.storen

AND st.storest=a.storest

AND st.storec=a.storec

AND LEFT(st.storez,5)=a.zips

UPDATE winners_copy AS win,

(SELECT count(*) AS count,retname, retstreet, RETCITY,

LEFT(retzip, 5) AS zips, min(locid) AS thething FROM winners_copy

GROUP BY retname, retstreet, RETCITY,

LEFT(retzip, 5)

ORDER BY count(*) DESC)a

SET win.locidfix=a.thething

WHERE a.count>1

AND win.retname=a.retname

AND win.retstreet=a.retstreet

AND win.RETCITY=a.RETCITY

AND LEFT(win.retzip,5)=a.zips

UPDATE sales AS st, (SELECT count(*) AS count,storen, storest, storec, LEFT(storez, 5) AS

zips, min(storeid) AS thething FROM stores_copy

GROUP BY storen, storest, storec, LEFT(storez, 5)

ORDER BY count(*) DESC)a

SET st.storeidfix=a.thething

```
WHERE a.count>1

AND st.storen=a.storen

AND st.storest=a.storest

AND st.storec=a.storec

AND LEFT(st.storez,5)=a.zips
```

Examples of analysis queries:

```
CREATE TEMPORARY TABLE a

SELECT DISTINCT(inmate_release_root.DCnumber) AS felon1,

inmate_release_root.lastname AS felon_last,

inmate_release_root.firstname AS felon_first, inmate_release_root.middlename AS

felon_middle, inmate_release_root.birthdate

AS felon_bday, inmate_release_offenses_cps.DCnumber AS felon2,

inmate_release_offenses_cps.county AS felon_county

FROM inmate_release_root, inmate_release_offenses_cps

WHERE

DISTINCT(inmate_release_root.DCnumber)=inmate_release_offenses_cps.DCnumber;

SELECT a.felon1_id, a.felon_last, a.felon_first, a.felon_middle, a.felon_bday,

a.felon_county, non_chain_owners.owner_id AS

lotto_id, non_chain_owners.owner_status AS lotto_status,

non_chain_owners.first_name AS lotto_first,

non_chain_owners.last_name AS lotto_last, non_chain_owners.county AS lotto_county
```

```
FROM a, non_chain_owners

WHERE a.felon_first=lotto_first

AND a.felon_last=lotto_last

AND a.felon_county=lotto_county

CREATE TEMPORARY TABLE all_sales

SELECT sum(sales) AS total_sales, storeidfix, storen, storest, storec, storez FROM sales

GROUP BY storeidfix, storen, storest, storec, storez;

CREATE TEMPORARY TABLE all_wins

SELECT sum(prize) tot_payouts , locidfix FROM winners

GROUP BY locidfix;

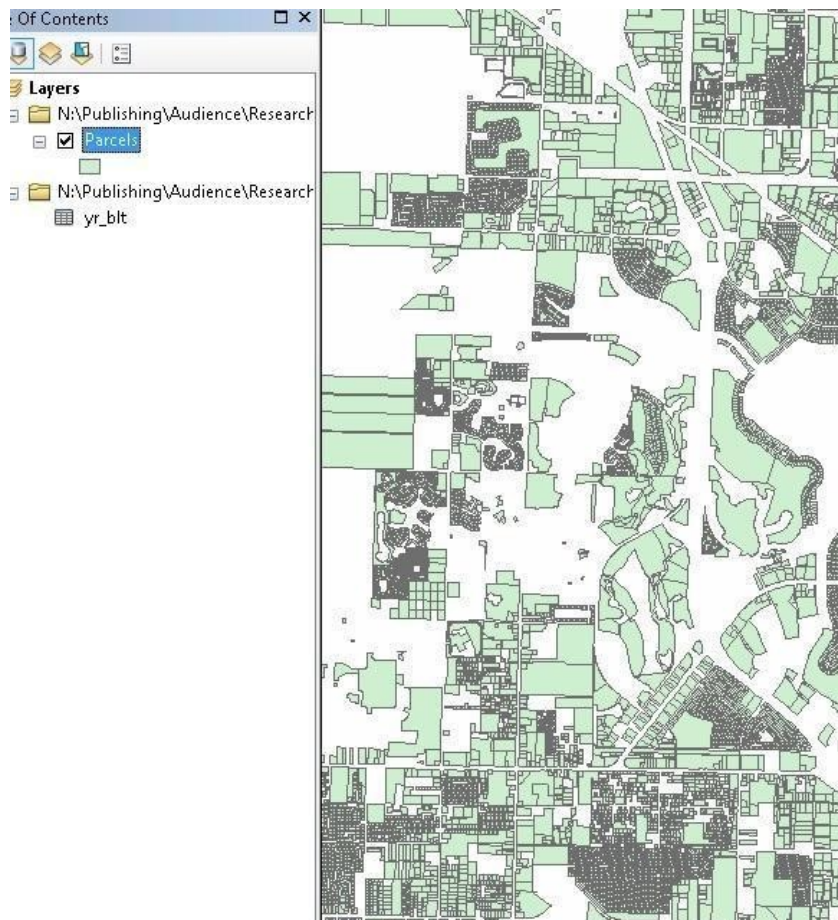
SELECT ROUND(tot_payouts/total_sales,4), storeidfix, storen, storest, storec, storez

FROM all_sales, all_wins

WHERE all_sales.storeidfix=all_wins.storeidfix
```

Mapping Analysis

ArcGIS screenshot



Query to create tables for maps

```
CREATE TEMPORARY TABLE taxes_13
```

```
SELECT
```

```
ROUND(f5+f8+f11+f14+f17+f20+f23+f25+f28+f31+f34+f37+f41+f44+f47+f50+f53+f56+f5
```

```
9+f61,2) AS 2013_taxes, pcn FROM rec3513;
```

```
CREATE TEMPORARY TABLE taxes_13_12
```

```

SELECT pcn,

ROUND(f5+f8+f11+f14+f17+f20+f23+f25+f28+f31+f34+f37+f41+f44+f47+f50+f53+f56+f5
9+f61,2), 2013_taxes AS 2012_taxes

FROM rec3613, taxes_13

WHERE taxes_13.pcn=rec3613.PCN;

CREATE TABLE taxes_and_values

SELECT rec4013.pcn, 2013_taxes, 2012_taxes, marval13, marval12, taxval13, taxval12

FROM taxes_13_12, rec4013

```

Screenshot:

totval13	totval12	totvalchange	valchange_pct	tottax13	tottax12
302306102	275899696	26406406	9.6	5439464.62	4935953.19
1892905564	1823758044	69147520	3.8	36312972.41	32978532.52
926888960	846226625	80662335	9.5	14671392.37	13575845.85
943826926	850269691	93557235	11.0	15881795.24	14816633.67
903076897	858575616	44501281	5.2	16669844.51	15416895.10
8742363339	8225687113	516676226	6.3	186868521.45	174117275.57
688151378	616594174	71557204	11.6	14355950.99	13428998.72
7652525989	6938947414	713578575	10.3	139128161.31	128274883.47
296511761	297212680	-700919	-0.2	5693005.31	5226993.49
874029980	826852670	47177310	5.7	15430904.06	14910131.62
1037022917	984860768	52162149	5.3	19639124.92	18834149.61
418425437	412289274	6136163	1.5	9183064.55	8879281.65
1943986703	1841220075	102766628	5.6	35202889.94	34262718.90
522024995	496984609	25040386	5.0	13473041.99	12993573.20
90859120	90806769	52351	0.1	1804957.28	1767682.19
259759725	252264587	7495138	3.0	5121935.85	5076025.36
144918062	133472446	11445616	8.6	3741612.48	3522225.87
4999591834	4680827196	318764638	6.8	104749098.08	99335728.82
7714874060	7037576469	677297591	9.6	143919842.17	136283133.95
558062910	522577421	35485489	6.8	11503795.50	10852410.02
9515393785	9031176213	484217572	5.4	187989575.22	184336998.48
1557222846	1396441112	160781734	11.5	34512368.79	33296809.86
3838514718	3594843015	243671703	6.8	78404997.50	74913914.59
18462953	15933706	2529247	15.9	308793.89	294985.53
20698487987	19602955198	1095532789	5.6	350485605.53	342103831.34
868102315	775697339	92404976	11.9	17105802.73	15776825.87
15157024469	13927132964	1229891505	8.8	231421579.70	219654111.64
320600689	294955439	25645250	8.7	5229798.07	5075404.70
5800705	5460193	340512	6.2	105601.92	102218.92
9135874648	8657140701	478733947	5.5	167858854.68	164305136.01
51918038058	49384057649	2533980409	5.1	853483694.84	841824321.51
2447806615	2315306647	132499968	5.7	43034620.01	41568184.90
2061910501	1851625708	210284393	11.4	26409862.70	25570507.02

Scrapers

Data scraped from West Palm Beach Police Department incident site

ase_num haracter varying(20)	datetime timestamp without time zone	offense character varying(255)	location character varying(255)	
0130015001	2013-09-03 00:38:00	STOLEN VEHICLE REC	1665	PALM BEACH LAKE!
0130015002	2013-09-03 00:50:00	LOITERING OR PROWL	300	EVERNIA ST
0130015003	2013-09-03 01:09:00	DOMESTIC TROUBLE/DI	5801	N FLAGLER DR
0130015004	2013-09-03 02:34:00	AGGRAVATED BATTERY	1909	N TAMARIND AVE
0130015037	2013-09-03 04:22:00	GRAND THEFT AUTO	1501	S FLAGLER DR
0130015119	2013-09-03 05:30:00	RAPE	2111	BRANDYWINE RD
0130015006	2013-09-03 06:08:00	DOMESTIC TROUBLE/DI	500	N CONGRESS AVE
0130015036	2013-09-03 06:15:00	BURGLARY - FORCED I	5108	51ST WAY
0130015007	2013-09-03 06:48:00	SIMPLE BATTERY(TOU	2105	PONCE DE LEON A
0130015263	2013-09-03 07:00:00	THEFT-PETIT \$100-\$	1880	N CONGRESS AVE
0130015009	2013-09-03 07:17:00	THEFT	3201	BROADWAY
0130015010	2013-09-03 08:15:00	POSSESSION OF OPEN	701	BELVEDERE RD
0130015379	2013-09-03 09:30:00	THEFT - PETIT	325	EXECUTIVE CENTE
0130015016	2013-09-03 09:44:00	POSSESSION OF OPEN	3999	S DIXIE HWY/SOU
0130015020	2013-09-03 10:08:00	POSSESSION OF COCA	400	PALM BEACH LAKE!
0130015024	2013-09-03 10:49:00	DOMESTIC TROUBLE/DI	645	EXECUTIVE CENTE
0130015026	2013-09-03 11:00:00	THEFT-PETIT \$100-\$	622	AVENIDA ALEGRE
0130015030	2013-09-03 11:32:00	INFORMATTON	1670	SUMMIT BLVD

Data scraped with Palm Beach County Civil Court data scraper

case_num character varying(20)	name character varying(255)	party_type character varying(50)	filing_date character varying(100)	caption character varying(255)
502002CA000285XXO	PALM BEACH COUNTY	DEFENDANT	09-JAN-02	CARRIE NEUNZIG
502002CA006211XXR	PALM BEACH COUNTY	DEFENDANT	28-MAY-02	UNITED STATES OF
502001CC003028XXC	PALM BEACH COUNTY	GARNISHEE	08-FEB-01	FORD MOTOR CREDIT
502002SC001725XXS	PALM BEACH COUNTY	DEFENDANT	23-JAN-02	STATE FARM MUTUAL
502001SC025337XXR	PALM BEACH COUNTY	DEFENDANT	07-NOV-01	JOAN R MARINO
502001SC022244XXR	PALM BEACH COUNTY	DEFENDANT	01-OCT-01	LEE D CORNELIUS JI
501997CA003146XXO	PALM BEACH COUNTY	DEFENDANT	09-APR-97	MARGARET HASSEL
501997CA003337XXR	PALM BEACH COUNTY	DEFENDANT	14-APR-97	G E CAPITAL MORTG
501997CA004715XXR	PALM BEACH COUNTY	DEFENDANT	28-MAY-97	COUNTRYWIDE FUNDI
501997CA006393XXR	PALM BEACH COUNTY	DEFENDANT	18-JUL-97	BANKERS TRUST COMI
502000CA004587XXR	PALM BEACH COUNTY	DEFENDANT	10-MAY-00	MANUFACTURERS & TI
501994CA007793XXA	PALM BEACH COUNTY	DEFENDANT	11-OCT-94	KIMBERLY L GONZALI
502003CA001493XXR	PALM BEACH COUNTY	DEFENDANT	10-FEB-03	US BANK NATIONAL
501997CA008187XXO	PALM BEACH COUNTY	DEFENDANT	15-SEP-97	CARLOS A MATOS
501997CA008238XXR	PALM BEACH COUNTY	DEFENDANT	16-SEP-97	K. HOVNANIAN MORTI
501999CA002494XXR	PALM BEACH COUNTY	DEFENDANT	filing_date	MIDFIRST BANK I
502004SC003081XXX	PALM BEACH COUNTY	DEFENDANT	filing_date	C POLIANDRO V P B
501988CA003100XXO	PALM BEACH COUNTY	DEFENDANT	06-APR-88	RICHARD SEAGER JR

Records Requests

Records request email exchanges:

Fedor:

Per your request, attached is the Traffic section of the Project Development and Environment Study for the PBI Interchange.

Barbara Kelleher

Public Information Director

Florida Department of Transportation District Four

3400 West Commercial Boulevard

Fort Lauderdale FL 33309

Email: Barbara.Kelleher@dot.state.fl.us

954-777-4090 or Toll Free at 1-866-336-8435 ext. 4090

www.dot.state.fl.us

From: Zarkhin, Fedor (CMG-WestPalm) [<mailto:FZarkhin@pbpost.com>]

Sent: Wednesday, June 26, 2013 1:16 PM

To: Kelleher, Barbara

Subject: RE: Follow up question RE: PBIA ramp expansion joint repairs

Hi Barbara,

Yes, Brian sent me the inspection report. I'd really appreciate it if you could help me out with one other thing. Where can I find the traffic projections for the flyovers from before the project started? Also, is there an update, by any chance, on who was overseeing the project in the early 2000s?

Thanks so much for your help.

Fedor

From: Kelleher, Barbara [<mailto:Barbara.Kelleher@dot.state.fl.us>]

Sent: Tuesday, June 25, 2013 4:31 PM

To: Zarkhin, Fedor (CMG-WestPalm)

Cc: Meredith Cruz

Subject: Follow up question RE: PBIA ramp expansion joint repairs

Fedor:

Per our construction project manager there are 24 expansion joints that are being repairs under this contract.

Also, I see that Brian O'Donoghue forwarded to you the independent inspection of the joints. Hope this is helpful.

Call or email if you have any further questions.

Barbara

Records Request for contents of D.A.V.I.D. searches:

To Whom it may concern,

I'm a reporter with the Palm Beach Post, and I'm writing to file a records request for information on people searched in the D.A.V.I.D. system by Florida law enforcement. If you have any questions, please call me on my cell at 814-933-8301.

Thank you!

The request is below:

10/7/2013

PUBLIC RECORDS REQUEST

Pursuant to Article I, section 24 of the Florida Constitution and Chapter 119, Florida

Statutes, I request an opportunity to inspect and copy the following public records:

An electronic database of all searches made by Florida law enforcement in the Driving and Vehicle Information Database (D.A.V.I.D) in the system currently used to track those searches. Although I request all information in that database that is public record, I would expect the database to have the following fields:

- Name of law enforcement agency that conducted the search

- Name or badge number of law enforcement officer making the search
- Location of the law enforcement officer when he/she conducted the search
- Name of person searched
- Date and time of search
- Home address of party searched
- Birth date of person searched
- Demographic information (age, gender, race, ethnicity)

I request these records be provided in an electronic format that can be imported into standard database software. Examples of such formats include an Excel .xls or .xlsx file, an Access .mdb or .accdb file, a text-based delimited file such as .csv or tab-delimited .txt, a .dbf file or an SQL dump readable by standard open-source database software. (A PDF file would not comply with this request because PDF files are not readable by database software.) If this information is stored in a relational database, I request it be provided in its original relational format, not "flattened" or de-normalized.

The Palm Beach Post will not accept this information in printed format and will not pay for printouts.

If you deny this request in whole or in part, I request that you advise me IN WRITING, as required by section 119.07(1)(e) and (f), Fla. Stat., of the particular statutory exemption upon which you are relying and that you state with particularity the reasons for your conclusion that the record is exempt. If an asserted exemption applies to only a portion of a record, please delete the portion that you claim is exempt and release the remainder of the record as required by section 119.07(1)(d).

The Palm Beach Post will pay for the “actual cost of duplication” of the public records as defined by Florida Statutes, section 119.011(1). Please advise me of these costs before taking any billable steps.

Thank you.

Fedor Zarkhin

Database Reporter

The Palm Beach Post

voice: 561.820.4497 | cell:814.933.8301 | fax: 561.820.4407

fzarkhin@pbpost.com | www.palmbeachpost.com

Hi Teri,

Hope all is well!

Could you please go ahead with my previous request, redactions notwithstanding? I’d also like to file another request, this one for specific people searched in the DAVID system. I’ll send a check over as soon as you tell me how much all of this will cost. I’m assuming there’s no way to pay by card, right?

Thank you!

10/16/2013

PUBLIC RECORDS REQUEST

Pursuant to Article I, section 24 of the Florida Constitution and Chapter 119, Florida Statutes, I request an opportunity to inspect and copy the following public records:

An electronic database of all searches made by Florida law enforcement in the Driving and Vehicle Information Database (D.A.V.I.D) system for the individuals listed below, from October 10, 2012 to October 15, 2013:

- Robert McKenna, DOB: 05/16/1966
- David Curtin, MOB: 10/1970
- Robert Keefe, MOB: 10/1978, living in Okeechobee, FL
- Shauna Keefe, MOB: 6/1975
- Maria Monsanto, MOB: 07/1975
- Shaun Winfield Harrelson, MOB: 10/1979

Although I request all information in that database that is public record, I would expect the database to have the following fields:

- Name of law enforcement agency that conducted the search
- Name of law enforcement officer making the search
- Name of person searched
- Date and time of search
- Birth date of person searched
- Demographic information such as age, gender, race and ethnicity, if available

I request these records be provided in an electronic format that can be imported into standard database software. Examples of such formats include an Excel .xls or .xlsx file, an Access .mdb or .accdb file, a text-based delimited file such as .csv or tab-delimited .txt, a .dbf file or an SQL dump readable by standard open-source database software. (A PDF file would not comply with this request because PDF files are not readable by

database software.) If this information is stored in a relational database, I request it be provided in its original relational format, not "flattened" or de-normalized.

The Palm Beach Post will not accept this information in printed format and will not pay for printouts.

If you deny this request in whole or in part, I request that you advise me IN WRITING, as required by section 119.07(1)(e) and (f), Fla. Stat., of the particular statutory exemption upon which you are relying and that you state with particularity the reasons for your conclusion that the record is exempt. If an asserted exemption applies to only a portion of a record, please delete the portion that you claim is exempt and release the remainder of the record as required by section 119.07(1)(d).

The Palm Beach Post will pay for the "actual cost of duplication" of the public records as defined by Florida Statutes, section 119.011(1). Please advise me of these costs before taking any billable steps.

Professional Analysis

Literature Review

Little research has been done on the importance of personal identifiers for data journalism. But there is an abundance of literature on the conflict between privacy and the right to know, one of the primary tensions directing the course of sunshine laws today.

Through my research I hope to arrive at practical findings of potential use to data journalists. I will study the real impact on data reporting of sunshine exemptions and restrictive court decisions. As such, my goal isn't so much to determine the causes and legal underpinnings of restrictions on access to personal identifiers but their impact on data journalists and methods of overcoming them. This will include personal identifiers in data subject to both the Freedom of Information Act and to individual state open records laws.

It will be useful to take a broad look at the legal origins of legislation and court decisions blocking access to personal identifiers. Although legal issues will not be the primary focus of my research, understanding them will offer the context necessary for a broader picture to emerge.

Surveying the literature, it becomes clear that although the Freedom of Information Act is an invaluable tool for investigating the executive branch of the federal government, court interpretations of some of its exemptions have taken a toll on freedom of access to information. There are two privacy exemptions to the FOIA – Exemptions 6 and 7(C). Exemption 6 allows agencies to withhold records if doing so

would “constitute a clearly unwarranted invasion of personal privacy” while 7(C) allows agencies to withhold law enforcement records if doing so “could reasonably be expected to constitute an unwarranted invasion of personal privacy” (Hoefges et al., 2003, p. 11).

The terms in these exemptions are broad and, as Hoefges et al. explain, it has been up to the courts to determine how they are to be applied.

Martin Halstuk writes that after the U.S. Supreme Court case *U.S. Dept. of the Air Force v. Rose (1976)*, the court has increasingly decided cases in favor of privacy over freedom of access to information (Halstuk, 2000, p. 126). *Rose* was prompted by the Air Force’s denial of a *New York University Law Review* FOIA request for summaries of official proceedings against Air Force Academy cadets accused of honor and ethics code violations. The journal requested the summaries with personally identifying information excluded. The request was denied, and the journal filed suit. It lost, and the appeals process took the case to the Supreme Court, which sided with the journal. The court ruled that privacy exemptions had to be interpreted narrowly (Halstuk, p. 126), meaning the bar for justifying use of a privacy exemption was high.

Since *Rose*, however, the court has “narrowly interpreted the public interest in disclosure and broadly construed the act’s privacy exemptions, thus allowing federal agencies to deny FOIA requests for access to government information that could be used to advance the public interest in important areas” (Halstuk, p. 126).

In *Department of State v. Washington Post company (1982)*, Supreme Court Chief Justice William Rehnquist suggested that Exemption 6 could be triggered with

even a minimal threat to a person's privacy. *The Washington Post* wanted to find out if two members of Iran's anti-American government held United States passports, but the State Department denied *The Post* that information. A district court ruled in the newspaper's favor, but the Supreme Court reversed that decision, determining that records do not need to have "intimate information" to trigger Exemption 6. Hoefges et al. wrote that that court case signaled a significant change in the interpretation of Exemption 6. In essence, it lowered the privacy threshold that would make information exempt.

Another case that shaped FOIA as it is today was *U.S. Department of Justice v. Reporters Committee for Freedom of the Press*. A CBS reporter requested a rap sheet from the FBI of someone allegedly involved in organized crime. The FBI refused, and the courts sided with them throughout the appeals process, all the way up to the Supreme Court (Hoefges et al., p. 21-23).

Charles Davis wrote that *Reporters Committee* "dramatically altered the treatment of federal Freedom of Information Act requests implicating privacy concerns." He explains that the Supreme Court's opinion concluded that FOIA's true purpose was to reveal government operations and that release of personal information usually does not further that goal (Davis, 2003, p.15).

The majority opinion established a litmus test for FOIA requests that involved private information: The records could be released only if they would help keep government open to public scrutiny. If the records do not meet that standard, then

disclosing them would constitute an unwarranted invasion of privacy (Hoefges et al., p.24).

Davis writes that that decision has been followed by numerous lower courts to close access to records with personally identifying information. And Davis does not understate the implications:

“... a prime purpose of the FOIA, the exposure of fraud, waste and abuse by government, has been tossed aside in favor of a reflexive privacy right that seemingly has no end in sight” (p. 18).

The fact that many records are now kept electronically, in databases, has underscored the challenge of preserving individuals’ privacy while maintaining access to information about the government. Privacy once almost entirely was ensured simply by the physical difficulty of accessing records. That is no longer the case. A Florida Judicial Management Council said it best in a report and recommendations submitted to the Florida Supreme Court in 2002:

The emergence of electronic information management technology and the Internet are causing widespread transformations in American society... Such changes do not always occur smoothly, and difficult issues have arisen. Among the most challenging is that of personal privacy. The balance between the free flow of information and the protection of personal privacy has been altered; institutions and individuals are now grappling with how a new balance should be struck (Report and Recommendations, 2002, p.1).

Institutions and individuals are grappling with this challenge just as much today as they were in 2002. One facet of this challenge is the fact that electronic data can provide in bulk what used to be available only in pieces, something that was also addressed in *Reporters Committee* (Chumbler, 2007).

Davis wrote that the *Reporters Committee* case created the concept of “practical obscurity.” This means that an individual’s privacy could be violated by the release of electronically compiled information that is already publicly available but in ways that make that information difficult to collect on one’s own. The information on the rap sheet of the criminal that the CBS reporter requested from the FBI was publicly available from different law enforcement agencies, but the release of that information in aggregated form would have violated his privacy (Davis, 16).

The cumulative effect of the *Washington Post* and *Reporters Committee* cases, and others, has been to tilt the balance in favor of privacy. The cases changed the “FOIA calculus,” making even a minimal risk of violation of privacy grounds to keep records closed (Hoefges et al., p. 56).

“Essentially, the Supreme Court has virtually ignored the notion that broad access to public records fulfills important societal functions beyond just informing the public about government operations” (Hoefges et al., pg. 57).

Journalists rarely argue in favor of restricting access to information and rarely try to examine the perspective from the other side of the issue. It will therefore be useful to take a look at some of the arguments in favor of limits to access to personal information.

Home addresses are among the things Grayson Barber wrote the government has a responsibility to protect. Citizens must provide an address to the government for basic services and should therefore expect that such information won't be distributed. Among the arguments is that the purpose of open records laws is to shed light on government, not to reveal information about specific individuals:

“Government agencies should not publish home addresses if the effect of the disclosure would be solely to disclose personal information about an individual, especially if the disclosure would shed no light on the conduct of a public agency or official or on other governmental matters of significance to the public” (Barber, 2006, p.88-89).

So far we have only examined case law impacting federal records. States have their own open records laws, with considerable variation among them. Some court decisions have mimicked *Reporters Committee for Freedom of the Press*, such as *Kestenbaum v. Michigan State University*. In that case, a student requested a computer tape with the names and addresses of Michigan State University students. The tape was used to make a publicly available university directory, but the court ruled in the university's favor, writing that the form of the information was just as important as the content in determining if it should be released. The student was not given the directory electronically (Bunker, 1993, p. 590).

Not all state courts have ruled in favor of privacy, however. In *Webb v. City of Shreveport*, a Louisiana appeals court granted the release of a computer tape with the names and addresses of public employees. The court determined that the fact that the

records were available in other forms meant that their release wouldn't violate the *Louisiana Constitution's* privacy provision (Bunker, p.592).

These court cases and numerous others demonstrate that the conflict between privacy and the need to keep government open does not have a simple solution. And although there are many nuances to the legal history of this conflict, they are not the focus of my research. Instead, the focus is on how restrictions on access to data affect investigative journalism.

One would assume that access to data increases the depth of news coverage, which is something that Raymond La Raja endeavored to determine quantitatively. La Raja analyzed the relationship between campaign contribution disclosure laws in the states and election coverage (La Raja, 2007).

Of La Raja's findings, the most pertinent is that horserace coverage decreased with more open "disclosure regimes," as the author calls disclosure laws in the various states. La Raja writes:

"The other striking finding is that horserace coverage predominates in poor disclosure regime states. This appears to support hypothesis 3 that inferior disclosure regimes lead reporters to rely on the horserace story because they have few other options. (La Raja, 2007, p. 243)"

Regarding the FOIA, Hoefges et al. corroborate La Raja's findings:

"In its FOIA privacy-exemption cases, the Supreme Court has paid little deference to the societal harm that can result from official government policy that favors secrecy over access. This harm ultimately includes the risks of a less-informed

electorate and weakened democratic participation and self-governance” (Hoefges et al, p.59).

It could be safely assumed that availability of data improves coverage of any subject, be it elections or otherwise. That’s the broad – and relatively self-evident – principle in favor of open data. Combined with a brief legal overview of freedom of information in the United States, this provides the background necessary to proceed to the core of my research – how lack of access to personal identifiers affects data reporters, and what they do to overcome closed access.

There is little academic research on this subject, unfortunately, but that’s not to say that it hasn’t been written about. The available evidence is mostly anecdotal, written by journalists for blogs and in newspapers.

An article on the Investigative Reporters and Editors website, written by Joe Yerardi, data editor for the *San Antonio Express-News*, presents one obstacle and a workaround to obtaining personal identifiers. Yerardi writes that he had requested payroll data from the City of San Antonio, and that the data he received did not include employee ID numbers. He said he needed those ID numbers because no other piece of information would uniquely identify an employee. The ID numbers were exempt under Texas sunshine laws (Yerardi, 2013).

But Yerardi found a workaround. He reached a compromise with the Finance Department whereby each employee was assigned a random but unique string of numbers. Still, Yerardi writes that that solution does not guarantee access to unique identifiers from records requests to other agencies:

“For the time being, journalists in Texas operate under the shadow of an Attorney General's ruling that, if aggressively implemented, will seriously damage the ability of the press to serve as watchdogs of the public's money and as advocates for open and effective government” (Yerardi, 2013).

Texas also happens to be one of a number of states to close access to public employee dates of birth. The Texas Supreme Court closed those records based on the fear that releasing them would constitute an unwarranted invasion of government employees’ right to privacy. Bob Mong, editor of the *Dallas Morning News*, said that the ruling makes it more difficult to hold public institutions accountable (Drew, 2010).

The article did not mention any ways to work around this obstacle. Judging by the experience of a few reporters at the *Philadelphia Inquirer*, the workaround to that problem could be rather unreliable (Snyder and Purcell, 2012).

Susan Snyder and Dylan Purcell, reporters for the *Philadelphia Inquirer*, wrote about their efforts to get school police officer employee records. They needed names and birth dates of school district police officers to check their criminal histories, but public employee birth dates are closed in Pennsylvania. Their solution to the problem? Getting the data from a “reliable source,” by which presumably they meant a confidential inside source (Snyder and Purcell, 2012).

Article

Data with personally identifying information are an invaluable tool for reporters nationwide. For beat reporters and veteran investigative journalists alike, information such as names, birth dates and addresses can make or break a story. But access to such

information isn't guaranteed, with laws that restrict the public's access a regular source of frustration. And bills adding to those restrictions are introduced regularly.

"There are constant changes," Kenneth Bunting, director of the National Freedom of Information Coalition, said. "It seems every year we're playing defense against bad new exemptions in every state legislature across the country."

But short of getting into politics to change the law, what can journalists do to get the data they request? What are some techniques and methods that journalists have used to negotiate successfully for data?

In attempt to answer that question, I asked hundreds of reporters across the country to recount how they negotiated for data with personally identifying information. I emailed reporters a questionnaire with questions about a records request for data with personally identifying information of their choosing. About fifty responded, all with valuable lessons on how to get the data that you need.

All but five of the 47 respondents work for newspapers, with an average of 13 years of experience as journalists. Reporters told about requests mostly for state or local data, with 17 and 21 requests, respectively. Law enforcement was the most popular category with 13 requests, followed by education, with six. The most common kind of request was for employee salary data, although data requested ran the gamut from hunting licenses to fire department log data.

A general trend that emerged in the responses was that persistence, ingenuity, and good relationships with records custodians can produce positive results. Of the 40 respondents that either got the data they requested or a version of it 17 were initially

denied their request. They succeeded using a number of techniques, including negotiation, compromise and finding other data sources.

Sometimes the only way to get the data you need is to find other avenues - often more time-consuming ones. Isaac Wolf of the Scripps Howard News Service wanted to see if owners of stores that had been banned from accepting food stamps were simply creating shell companies that bought the banned stores. To do that he requested a list of banned stores from the US Department of Agriculture, which gave him most of what he sought except for owner names.

Without the names, Wolf had no easy way to check whether people were using shell companies to be able to accept food stamps again. So instead of doing a comprehensive check on all banned stores, Wolf took a handful and tracked their ownership using corporate records, health inspections, liquor licenses, and more. Although time-consuming and less comprehensive, his method was sufficient to prove his hypothesis: Some people owning stores that had lost the right to participate in the food stamp program simply created new store names for the store and continued to participate in the program. The first story based on the data was published in February 2012, available here: <http://projects.scrippsnews.com/story/exclusive-store-owners-banned-taking-food-stamps-still-do/>.

A major project Chad Day of the Arkansas Democrat-Gazette worked on required similar creativity. Day and his colleagues wanted to find out how many Arkansans charged with murder or manslaughter had committed crimes while on parole. Virtually all personally identifying information for parolees was exempt in Arkansas, so Day and

his colleagues requested a court database with case information for everyone in Arkansas charged with murder or manslaughter, which they then joined with a prison database. The court data allowed Chad to find parolees with pending criminal cases, as opposed to only those already convicted. It also had the arrest date for each offender, crucial for tracking parolee's path through the criminal justice system.

Among other things, the data revealed that an Arkansas Department of Community Correction policy released parolees off of state supervision after their sentence is finished, even if they had previously stopped reporting to their parole officers. Some of those parolees went on to commit murders and other crimes. Articles based on the data started running in the summer of 2013, one of which can be read here, with a subscription: <http://m.arkansasonline.com/news/2013/aug/11/parole-skipper-killer-2012-20130811/>.

Sometimes there is no way to get the data you need, in which case reporters might have to compromise. While Andy Boyle was an intern at the Arkansas Democrat-Gazette, he requested hunting license data from the Arkansas Department of Fish and Wildlife, but was denied licensee dates of birth. He asked for month and year of birth instead, which the agency decided he could have. That was sufficient for him to join that table with a table of felons. He discovered that there were convicted felons in Arkansas with hunting licenses. The paper published Boyle's article based on the data in August 2008. The article is available on Boyle's website: <http://www.andymboyle.com/portfolio/felons-hunting-with-guns-slip-past-state-law/>.

Informal agreements can also help reporters get personally identifying information. Linda Johnson of the Lexington World-Herald requested an employee salary database for everyone working at the University of Kentucky. Though employee race and gender are exempt in Kentucky, the university agreed to give the World-Herald employee gender and race information, only on the condition that the paper would only publish this information in aggregate.

In another example of compromise, while at the San Antonio News-Express Joe Yerardi requested a database of city employees' salaries over a number of years. He asked for the employee ID number as well to accurately track people as they changed positions or got married. But the city declined and said that employees used their ID numbers to buy gasoline, and the Texas Attorney General had already determined that those IDs were not a matter of public record. Yerardi then asked the city to generate a random and unique string for each employee, which they did. This allowed him to track employees through the system while allaying the city's fears that their ID numbers would be used to buy gasoline. He didn't end up using the data for any stories, though he did have it on hand if needed.

When such workarounds are impossible, the best thing to do can be to rely on data obtained before laws restricting access were passed. A respondent who didn't want to be named for fear of compromising his investigation was denied the dates of birth of election candidates in his state. To get them, he joined an older table of voter records with a list of the names of election candidates.

Personally identifying information is an invaluable resource for journalists, and problems getting it aren't going away any time soon. Even though almost every records battle is unique, as is each agency and records custodian, there are numerous techniques and strategies which, in combination with persistence, allow journalists to keep fighting the good fight.

Works Cited

- Akpan, C., Ifeanyi, A., Martin, O., Alexander, O., & Uchenna, A. (2012). Rethinking objectivity in news reporting in the digital age. *Alternative Perspectives in the Social Sciences*, 4(4), 711-729.
- Barber, G. (n.d.). Personal information in government records: Protecting the public interest in privacy. *Saint Louis University Public Law Review*, 25, 63:121.
- Chumbler, H. C. (2007). Addressing privacy in the digital age: The loss of practical obscurity. In H. Chumbler (Ed.), *Access to Government in the Computer Age: An Examination of State Public Records Laws* (1 ed., p. 71-91). Chicago: American Bar Association Publishing.
- Davis, C. (2003). Electronic access to information and the privacy paradox: Rethinking practical obscurity and its impact on electronic freedom of information. *Social Science Computer Review*, 21(1), 15-25.
- Drew, J. (2010, December 04). Texas supreme court rules state workers' birth dates aren't public. *Dallas Morning News*. Retrieved from <http://www.dallasnews.com/news/politics/texas-legislature/headlines/20101204-texas-supreme-court-rules-state-workers-birth-dates-aren-t-public.ece>
- Halstuk, M. (2000). Blurred vision: The supreme court. In C. Davis & S. Splichal (Eds.), *Access Denied: Freedom of Information in the Information Age* (1 ed., p. 121-153). Ames: Iowa State University Press.
- Hoefges, M., Halstuk, M., & Chamberlin, B. (2003). Privacy rights versus foia disclosure policy: the "uses and effects" double standard in access to personally-identifiable information in government records. *William and Mary Bill of Rights Journal*, 12(1), 1-65.

- La Raja, R. (2007). Sunshine laws and the press: The effect of campaign disclosure on news reporting in the american states. *Election Law Journal*, 6(3), 236-250.
- Locke, J. (1689). *An essay concerning human understanding* . ILT Digital Classics. Retrieved from http://www.ilt.columbia.edu/publications/locke_understanding.html
- Macias, W., Springton, J., Lariscy, R., & Neustifter, B. (2008). A 13-year content analysis of survey methodology in communication related journals. *Journal of Current Issues and Research in Advertising*, 30(1), 79-94.
- Milton, J. (1644). *Areopagitica a speech for the liberty of unlicensed printing to parliament of england*. The Project Gutenberg. Retrieved from <http://www.gutenberg.org/files/608/608.txt>
- Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, 73(1), 74-97.
- Ramaprasad, J., & Rahman, S. (2006). Tradition with a twist : A survey of bangladeshi journalists. *International Communication Gazette*, 68(148), 148-165.
- Siebert, F., Peterson, T., & Schramm, W. (1963). *Four theories of the press*. Urbana and Chicago: University of Illinois Press.
- Snyder, S., & Purcell, D. (2012, March 30). [Web log message]. Retrieved from <http://ire.org/blog/ire-journal/2012/03/30/crimes-classroom/>
- Wimmer, R., & Dominick, J. (2006). *Mass media research: An introduction*. (8 ed.). Boston: Wadsworth.
- Yerardi, J. (2013, January 17). [Web log message]. Retrieved from <http://www.ire.org/blog/transparency-watch/2013/01/17/san-antonio-withholds-employee-identifiers-backing/>
- Yu, J. & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, 10(1), 35-44.
- In re* Report and Recommendations of the Judicial Management Council of Florida on Privacy and Electronic Access to Court Records, 832 So. 2d 712, 714 (Fla. 2002) http://www.floridasupremecourt.org/pub_info/summaries/briefs/02/02-659/Filed_12-17-2001_ReportRecommendations.pdf

Appendix

Project Proposal

Project Proposal

Impact of Privacy Exemptions in Open Records Laws on Computer-Assisted Reporting

By Fedor Zarkhin

Project Committee:

David Herzog, Chair

Scott Swafford

Mark Horvit

Summer, 2013

Missouri School of Journalism

Introduction

I came to the University of Missouri's School of Journalism with vague and idealistic goals of changing the world for the better. I couldn't think of a career path more ideal for satisfying my desire to learn, write, and be useful in some way to humanity.

Two years into the program, I have a more practical understanding of journalism, but that doesn't mean I've strayed from my original goals. The initial idealism remains, but it has been complemented by a set of skills I plan on using throughout my career. The skillset I least expected to acquire and value as much as I do is that of computer-assisted reporting. Taking David Herzog's data journalism class, working at the National Institute for Computer-Assisted Reporting, and working with too many datasets to count – all these experiences have solidified my love for numbers, spreadsheets, and SQL queries.

I am confident in my skills as a data reporter and equally confident in my desire to develop them further while participating in investigative projects. I am now ready to take what I have learned and get to work in the "real world."

The *Kyiv Post*, meanwhile, showed me what kind of media environment can exist when data is almost never available; some of my colleagues didn't even have a conception of what a data request was.

My career goals are not quite clear to me yet, although I will certainly incorporate CAR in my career. My tentative plan is to gain experience as a CAR and also

as a beat reporter in the United States, and to then take the skills I learn and use them at a publication in Russia. The Russian media have a lot of progress to make if they are to adequately serve the watchdog role in Russian society. I want to be part of that progress.

Professional Practice

Fortunately, I will spend the next year in an environment very much opposite to the one I found in Kyiv: Since April 8, I have been working as a computer-assisted reporting intern at the *Palm Beach Post*. I will spend a year there, and I will use 14 weeks of that year for the professional component of this project. Those 14 weeks will start upon my committee's approval of my project proposal, which I hope will be around May 16. My supervisor will be Joel Engelhardt, investigative editor at the *Post*. My duties will include helping reporters with data, independently finding and reporting investigative stories, working with the investigative team on long-term projects, and covering breaking news on a rotational basis. My supervisor's explanation of my duties can be found in Appendix A.

The *Palm Beach Post*, a newspaper owned by Cox Media Group, has a circulation of about 80,000 per day. It is the primary newspaper in Palm Beach County.

While at the *Post*, I will be the go-to-man for all things CAR. The staff's experience level isn't particularly high, although a few people have attended the NICAR boot camp and the environmental beat reporter has experience. I will have two primary roles: First, I will work with all the reporters on stories they have that involve data. If the

reporter on the education beat gets a database of salaries for the school district's employees, for example, it'll be my job to see if there's a story in there somewhere. Second, I will be working with the *Post's* investigative reporters on long-term projects.

The *Post* recently got court data for Palm Beach County dating back to the 1980s. It's my job to scour that data and find stories. Engelhardt said that project is wide open: I will work with the data and tell him what the stories are. I also will be looking into voting data and the sheriff's department's vehicle purchase data.

But those are only the first projects I will take on; the exact direction my professional practice will take is hard to predict; that's just the nature of the newsroom. What is certain is that my CAR skills will be put to extensive use: I will be one of only two people at the *Post* with any experience with SQL, Python and Django. I must stress, however, that my role at the *Post* will be fluid. Engelhardt has mentioned that I will also report and write stories for the paper.

While at the *Post*, I will submit weekly notes to my committee to demonstrate the work I am doing. For abundant physical evidence, I will provide the following: logs of the work I do on databases; articles for which I analyze data; and articles I report and write myself.

Analysis Component

Data journalism hinges on access to official records. Sure, reporters can compile their own databases in some cases, but that cannot compare with the wealth of stories to be found in the data stored by government agencies. Getting access to that data isn't

always easy, however, with legislatures passing restrictive exemptions and agencies stonewalling journalists. Accessing the records necessary to do stories – or just to find them, for that matter – is a never-ending battle for data journalists nationwide.

Restrictions on access to personal identifiers are a particularly damaging category of exemption, passed in various forms in numerous states. Without personal identifiers such as dates of birth and addresses, it becomes virtually impossible to match one dataset of people to another. Numerous investigations become difficult, if not impossible, to do.

When personal identifiers become closed to the public, is that the end of the investigation? Fortunately for investigative journalism and for American society, that does not always have to be the case. For my professional analysis, I will do two things. Through surveys and interviews, I will first find out how investigative projects are affected by lack of access to personal identifiers. Second, I will find out what techniques data reporters use to overcome those obstacles.

This professional analysis will fit neatly with my professional skills component. As a data reporting intern at the *Post*, I am likely to encounter problems similar to those I will be researching. Public employee dates of birth have recently become closed in Florida, for example, so I will almost certainly have the chance to fight the same battles as my research subjects, as well as to try some of their techniques.

My goal is to produce an article suitable for publication in a magazine such as the *IRE Journal* or *Uplink*.

Theoretical Framework

I will use both political and media theories for my professional analysis. More specifically, my analysis will stem from libertarian theory of the public's right to know about the workings of government, and the idea that fulfilling that right is vital to ensuring a just and representative democracy. I will also use social responsibility theory, which states that the media have a number of responsibilities to society, one of which is monitoring the government.

John Milton sparked libertarian thought in his 1644 essay *Aeropagitica*, in which he protested English book licensing laws that stymied dissemination of his work. In the essay, Milton argues that the truth emerges in an open marketplace of ideas. Ideas should not, therefore, be censored. The idea that government records must be available to the public is a natural extension of this idea (Milton, 1644).

John Locke, another originator of libertarian thought, argued in *An Essay on Human Understanding* for openness of information. If all people have different opinions, and nobody accepts a universal higher authority, the best course of action is for information and ideas to be available to all, he wrote (Locke, 1689).

“For where is the man that has incontestable evidence of the truth of all that he holds, or of the falsehood of all he condemns; or can say that he has examined to the bottom all his own, or other men's opinions?” (Locke, 1689, p. 4).

Both Milton and Locke argue that the best society is one in which information is free and open. Their libertarian ideals are rather abstract, though. Social responsibility

theory, meanwhile, homes in on the media's role in society, a vital component to the theoretical backdrop for my research.

As Theodore Peterson writes in *Four Theories of the Press*, the media do not function in a vacuum. Instead, the freedom they have in democratic societies obliges them to use that freedom in the interests of society. Peterson lists six press functions, the most pertinent of which is the third: "Safeguarding the rights of the individual by serving as watchdog against the government" (Siebert, Peterson and Schramm, 1963, p. 74).

The difference between libertarian and social responsibility theory is the role both theories assume government has in society. While the former theory states that the government shouldn't interfere in the public's quest for information, the latter states that the government should play an active role in disseminating that information. In other words, the difference is one of positive versus negative freedom (Siebert, Peterson and Schramm, 1963).

It is appropriate to use both these theories in my research because they are similar yet rather different. Both theories underscore freedom of access to information, but they also propose different methods for achieving that.

Jyotika Ramaprasad and Shafiqur Rahman used libertarian theory to frame their research questions in their survey of Bangladeshi journalists. Their study sought to compare and contrast libertarian media ideals with how they are put into practice by journalists in Bangladesh. In establishing a theoretical framework, they built on the theory as described by Siebert in *Four Theories of the Press* by making it more empirical.

They also sought to make it more applicable to third-world countries (Ramaprasad and Rahman, 2006, p. 150).

Akpan et al., meanwhile, used social responsibility theory for their research on objectivity in online news reporting. They chose this theory because objectivity is a key component of journalistic social responsibility. The authors write that online journalism poses a threat to objectivity and, by extension, social responsibility (Akpan et al., 2012, p. 721).

Literature Review

Little research has been done on the importance of personal identifiers for data journalism. But there is an abundance of literature on the conflict between privacy and the right to know, one of the primary tensions directing the course of sunshine laws today.

Through my research I hope to arrive at practical findings of potential use to data journalists. I will study the real impact on data reporting of sunshine exemptions and restrictive court decisions. As such, my goal isn't so much to determine the causes and legal underpinnings of restrictions on access to personal identifiers, but their impact on data journalists and methods of overcoming them. This will include personal identifiers in data subject to both the Freedom of Information Act and to individual state open records laws.

It will be useful to take a broad look at the legal origins of legislation and court decisions blocking access to personal identifiers. Although legal issues will not be the

primary focus of my research, understanding them will offer the context necessary for a broader picture to emerge.

Surveying the literature, it becomes clear that though the Freedom of Information Act is an invaluable tool for investigating the executive branch of the federal government, court interpretations of some of its exemptions have taken a toll on freedom of access to information. There are two privacy exemptions to the FOIA – Exemptions 6 and 7(C). Exemption 6 allows agencies to withhold records if doing so would “constitute a clearly unwarranted invasion of personal privacy” while 7(C) allows agencies to withhold law enforcement records if doing so “could reasonably be expected to constitute an unwarranted invasion of personal privacy” (Hoefges et al., 2003, p. 11).

The terms in these exemptions are broad and, as Hoefges et al. explain, it has been up to the courts to determine how they are to be applied.

Martin Halstuk writes that after the U.S. Supreme Court case *U.S. Dept. of the Air Force v. Rose*(1976), the court has increasingly decided cases in favor of privacy over freedom of access to information (Halstuk, 2000, p. 126). *Rose* was prompted by the Air Force’s denial of a *New York University Law Review* FOIA request for summaries of official proceedings against Air Force Academy cadets accused of honor and ethics code violations. The journal requested the summaries with personally identifying information excluded. The request was denied, and the journal filed suit. It lost, and the appeals process took the case to the Supreme Court, which sided with the journal. The court

ruled that privacy exemptions had to be interpreted narrowly (Halstuk, p. 126), meaning the bar for justifying use of a privacy exemption was high.

Since *Rose*, however, the court has “narrowly interpreted the public interest in disclosure and broadly construed the act’s privacy exemptions, thus allowing federal agencies to deny FOIA requests for access to government information that could be used to advance the public interest in important areas” (Halstuk, p. 126).

In *Department of State v. Washington Post co. (1982)*, Supreme Court Chief Justice William Rehnquist suggested that Exemption 6 could be triggered with even a minimal threat to a person’s privacy. *The Washington Post* wanted to find out if two members of Iran’s anti-American government held United States passports, but the State Department denied *The Post* that information. A district court ruled in the newspaper’s favor, but the Supreme Court reversed that decision, determining that records do not need to have “intimate information” to trigger Exemption 6. Hoefges et al. wrote that that court case signaled a significant change in the interpretation of Exemption 6. In essence, it lowered the privacy threshold that would make information exempt.

Another case that shaped FOIA as it is today was *U.S. Department of Justice v. Reporters Committee for Freedom of the Press*. A CBS reporter asked requested a rap sheet from the FBI of someone allegedly involved in organized crime. The FBI refused, and the courts sided with them throughout the appeals process, all the way up to the Supreme Court (Hoefges et al., p. 21-23).

Charles Davis wrote that *Reporters Committee* “dramatically altered the treatment of federal Freedom of Information Act requests implicating privacy concerns.” He explains that Supreme Court’s opinion concluded that FOIA’s true purpose was to reveal government operations, and that release of personal information usually does not further that goal (Davis, 2003, p.15).

The majority opinion established a litmus test for FOIA requests that involved private information: The records could be released only if they would help keep government open to public scrutiny. If the records do not meet that standard, then disclosing them would constitute an unwarranted invasion of privacy (Hoefges et al., p.24).

Charles Davis writes that that decision has been followed by numerous lower courts to close access to records with personally identifying information. And Davis does not understate the implications:

“... a prime purpose of the FOIA, the exposure of fraud, waste and abuse by government, has been tossed aside in favor of a reflexive privacy right that seemingly has no end in sight” (p. 18).

The fact that many records are now kept electronically, in databases, has underscored the challenge of preserving individuals’ privacy while maintaining access to information about the government. Privacy once almost entirely was ensured simply by the physical difficulty of accessing records. That is no longer the case. A Florida Judicial Management Council said it best in a report and recommendations submitted to the Florida Supreme Court in 2002:

“The emergence of electronic information management technology and the Internet are causing widespread transformations in American society... Such changes do not always occur smoothly, and difficult issues have arisen. Among the most challenging is that of personal privacy. The balance between the free flow of information and the protection of personal privacy has been altered; institutions and individuals are now grappling with how a new balance should be struck” (Report and Recommendations, 2002, p.1).

Institutions and individuals are grappling with this challenge just as much today as they were in 2002. One facet of this challenge is the fact that electronic data can provide in bulk what used to only be available in pieces, something that was also addressed in *Reporters Committee* (Chumbler, 2007).

Charles Davis wrote that the *Reporters Committee* case created the concept of “practical obscurity.” This means that an individual’s privacy could be violated by the release of electronically compiled information that is already publicly available, but in ways that make that information difficult to collect on one’s own. The information on the rap sheet of the criminal that the CBS reporter requested from the FBI was publicly available from different law enforcement agencies, but the release of that information in aggregated form would have violated his privacy (Davis, 16).

The cumulative effect of the *Washington Post* and *Reporters Committee for Freedom of the Press* cases, and others, has been such as to tilt the balance in favor of privacy. The cases changed the “FOIA calculus,” making even a minimal risk of violation of privacy grounds to keep records closed (Hoefges et al., p. 56).

“Essentially, the Supreme Court has virtually ignored the notion that broad access to public records fulfills important societal functions beyond just informing the public about government operations” (Hoefges et al., pg. 57).

Journalists rarely argue in favor of restricting access to information and rarely try to examine the perspective from the other side of the issue. It will therefore be useful to take a look at some of the arguments in favor of limits to access to personal information.

Among the things Grayson Barber wrote the government has a responsibility to protect are home addresses. Citizens must provide an address to the government for basic services, and should therefore expect that that information won't be distributed. Among the arguments is that the purpose of open records laws is to shed light on government, not to reveal information about specific individuals:

“Government agencies should not publish home addresses if the effect of the disclosure would be solely to disclose personal information about an individual, especially if the disclosure would shed no light on the conduct of a public agency or official or on other governmental matters of significance to the public” (Barber, 2006, p.88-89).

So far we have only examined case law impacting federal records. States have their own open records laws, with considerable variation among them. Some court decisions have mimicked *Reporters Committee for Freedom of the Press*, such as *Kestenbaum v. Michigan State University*. In that case, a student requested a computer tape with the names and addresses of Michigan State University students. The tape was

used to make a publicly-available university directory, but the court ruled in the university's favor, writing that the form of the information was just as important in determining if it should be released as the content (Bunker, 1993, p. 590).

Not all state courts have ruled in favor of privacy, however. In *Webb v. City of Shreveport*, a Louisiana appeals court granted the release of a computer tape with the names and addresses of public employees. The court determined that the fact that the records were available in other forms meant that their release wouldn't violate the *Louisiana Constitution's* privacy provision (Bunker, p.592).

These court cases and numerous others demonstrate that the conflict between privacy and the need to keep government open does not have a simple solution. And though there are many nuances to the legal history of this conflict, they are not the focus of my research. Instead, the focus is on how restrictions on access to data affect investigative journalism.

It seems intuitively true that access to data increases the depth of news coverage, which is something that Raymond La Raja endeavored to determine quantitatively. La Raja analyzed the relationship between campaign contribution disclosure laws in the states and election coverage (La Raja, 2007).

Of La Raja's findings, the most pertinent to my research is that horserace coverage decreased with more open "disclosure regimes," as the author calls disclosure laws in the various states. La Raja writes:

"The other striking finding is that horserace coverage predominates in poor disclosure regime states. This appears to support hypothesis 3 that inferior disclosure

regimes lead reporters to rely on the horserace story because they have few other options. (La Raja, 2007, p. 243)”

Regarding the FOIA, Hoefges et al. corroborate La Raja’s findings:

“In its FOIA privacy-exemption cases, the Supreme Court has paid little deference to the societal harm that can result from official government policy that favors secrecy over access. This harm ultimately includes the risks of a less-informed electorate and weakened democratic participation and self-governance” (Hoefges et al, p.59).

It could be safely assumed that availability of data improves coverage of any subject, be it elections or otherwise. That’s the broad – and relatively self-evident – principle in favor of open data. Combined with a brief legal overview of freedom of information in the United States, this provides the background necessary to proceed to the core of my research – how lack of access to personal identifiers affects data reporters, and what they do to overcome closed access.

There is little academic research on this subject, unfortunately, but that’s not to say that it hasn’t been written about. The available evidence is mostly anecdotal, written by journalists for blogs and in newspapers.

An article on the Investigative Reporters and Editors website, written by Joe Yerardi, data editor for the *San Antonio Express-News*, presents one obstacle and a workaround to obtaining personal identifiers. Yerardi writes that he had requested payroll data from the City of San Antonio, and that the data he received did not include employee ID numbers. He said he needed those ID numbers because no other piece of

information would uniquely identify an employee. The ID numbers were exempt under Texas sunshine laws (Yerardi, 2013).

But Yerardi – a recent University of Missouri graduate – found a workaround. He reached a compromise with the Finance Department whereby each employee was assigned a random but unique string of numbers. Still, Yerardi writes that that solution does not guarantee access to unique identifiers from records requests to other agencies:

“For the time being, journalists in Texas operate under the shadow of an Attorney General's ruling that, if aggressively implemented, will seriously damage the ability of the press to serve as watchdogs of the public's money and as advocates for open and effective government” (Yerardi, 2013).

Texas also happens to be one of a number of states to close access to public employee dates of birth. The Texas Supreme Court closed those records based on the fear that releasing them would constitute an unwarranted invasion of government employees’ right to privacy. Bob Mong, editor of the *Dallas Morning News*, said that the ruling makes it more difficult to hold public institutions accountable (Drew, 2010).

The article did not mention any ways to work around this obstacle. Judging by the experience of a few reporters at the *Philadelphia Inquirer*, the workaround to that problem could be rather unreliable (Snyder and Purcell, 2012).

Susan Snyder and Dylan Purcell, reporters for the *Philadelphia Inquirer*, wrote about their efforts to get school police officer employee records. They needed names and birth dates of school district police officers to check their criminal histories, but

public employee birth dates are closed in Pennsylvania. Their solution to the problem? Getting the data from a “reliable source,” by which presumably they meant a confidential inside source (Snyder and Purcell, 2012).

Methodology

I will use answer my research questions through surveys, and I will some color and detail for the final product through telephone interviews. I will email a Google web form to at least 200 data editors and reporters in news organizations in each of the 50 states. Their answers will automatically appear in a spreadsheet on my Google Drive once they are filled out. Second, I will select some of the respondents for more in-depth telephone surveys. For my sample I will select the news outlets with the highest circulations or page-views in their state. I will do this based on the assumption that the highest-circulation news outlets also will have the largest amount of resources to overcome obstacles to access.

Online surveys are a good method for this analysis for two reasons. First, I would like to get as broad a sample of the challenges faced by data reporters across the country as possible, and surveys will allow me to collect a larger sample than interviews. Second, an online survey will allow the respondents some time to think about their answers. I will ask them for details, and details aren’t always easy to recall when one is put on the spot, as they would be with an interview.

The following is a tentative list of questions I will ask journalists to respond to on the web survey:

1. Have you ever been denied personally identifying information by an agency you requested data from?
 - a. What was the agency?
 - b. What were the records you requested?
 - c. What was the agency's rationale for denying the information?
2. Was the conflict resolved?
 - a. Did you get the records?
 - b. If yes, how?
 - c. If you didn't get the records, did you find a work-around that met your needs?
3. If you neither got the records nor found a compromise with the agency, how did this affect your project?
 - a. Did you achieve less than you had initially planned? Please explain.
4. By the end of the project, did you achieve your original goals? Please explain.

After collecting the answers to these questions, I will select reporters or editors whose experiences are most representative of all the responses for telephone interviews. If I find, for example, that many of the survey responses talk about difficulty with access to medical information, I will call the reporter whose experience will be most compelling for my final article. The goal of the telephone interviews will be to get more detailed information from the respondents. In other words, the vast majority of my research will be completed through the survey, while the interviews will add color for the final product, and will not play a key role in the research itself. I am not providing

examples of interview questions here because they will be entirely dependent on the survey results.

Although survey research involves some risk, it is outweighed by the benefits. Roger Wimmer and Joseph Dominick list a few: low cost; allows for research in realistic settings; allows for a large amount of data to be collected (Wimmer & Dominick, 2006, p.180).

There are a number of risks that come with doing survey research en-masse. Wimmer and Dominick write that different kinds of wording or question placement can influence responses (2006, p. 196). I do not believe this will affect my research, however. The purpose of my survey will not be to reveal the hidden meaning behind the respondents' answers; instead, I want to learn about their experiences in their own words. The only analysis I will do will be to find patterns among all the responses, not to dissect individual ones.

The greatest challenge in survey research is getting people to respond. Yu and Cooper found that there are a number of techniques that increase the likelihood that people will respond to surveys, and I will employ some of these techniques myself (Yu and Cooper).

Among other things, Yu and Cooper's study found that it can be useful to stress the social utility of a study. My research certainly has the potential to be useful to journalists across the country, and I will stress this when contacting subjects (Yu and Cooper, p. 39).

Wendy Macias et al. analyzed research in communication that involved surveys, taking a look at how methodology impacted response rate. Macias et al. studied about 13 years' worth of articles, with 565 surveys used for articles published in 46 journals.

Among their findings was that financial incentives increased response rates.

Unfortunately I cannot do this, for obvious ethical and financial reasons. They also found that following up after sending a survey increases response rates, and that is something that I can do. Specifically, I will send a follow-up email to all my survey subjects who have not responded two weeks after sending the survey (Macias, 2003, p. 80-93).

The design of the survey will have to be such as to ensure that those who start the survey also finish it. Andy Peytchev examined the reasons behind what he calls "survey breakoff," and I will use some of his findings to increase the likelihood that respondents will finish my surveys.

Among other things, Peytchev found that long questions and a large number of questions per page increased the likelihood of survey breakoff. Overall, I must make the survey as easy to complete as possible. Peytchev also recommends thinking through the survey design features. It is important to not make the survey seem long, for example. The study also demonstrated that a higher education level was associated with a lower risk of survey breakoff, and I expect that to work in my favor (Peytchev, 2009, p. 93).

Appendix A

Mr. Herzog,

My name is Joel Engelhardt. I oversee the investigative team at The Palm Beach Post and Fedor Zarkhin reports to me.

I will be happy to oversee his work for this project and submit an evaluation of his work at its conclusion.

Fedor's duties are broad and still being refined but they can include all or any of the following:

- Report and write investigative news stories.
- Seek out databases from public sources, clean and analyze them and suggest stories based on findings. Depending on the findings, be involved in either writing the story himself or working directly with the reporter assigned to write the story.
- Assist reporters from throughout the newsroom in identifying databases to secure. Helping them clean and analyze data.
- General assistance with common spreadsheet tasks for reporters newsroom wide.
- Coverage of weekend police and events as part of a rotation of reporters assigned to weekend duty.
- Work with web producers to brainstorm and develop innovative web projects to illustrate investigative stories.

-- Help newsroom response to major breaking news, as occurred after Boston Marathon bombings.

In his brief time here, Fedor has shown himself to be an eager worker, pursuing news stories in many unique and interesting ways. He also has reported and written his first 1A story. I look forward to working with him on his professional project and I am happy to answer any questions that you have.

Sincerely,

Joel Engelhardt

Investigative Team Editor

The Palm Beach Post

Works Cited

- Akpan, C., Ifeanyi, A., Martin, O., Alexander, O., & Uchenna, A. (2012). Rethinking objectivity in news reporting in the digital age. *Alternative Perspectives in the Social Sciences*, 4(4), 711-729.
- Barber, G. (n.d.). Personal information in government records: Protecting the public interest in privacy. *Saint Louis University Public Law Review*, 25, 63:121.
- Chumbler, H. C. (2007). Addressing privacy in the digital age: The loss of practical obscurity. In H. Chumbler (Ed.), *Access to Government in the Computer Age: An Examination of State Public Records Laws* (1 ed., p. 71-91). Chicago: American Bar Association Publishing.
- Davis, C. (2003). Electronic access to information and the privacy paradox: Rethinking practical obscurity and its impact on electronic freedom of information. *Social Science Computer Review*, 21(1), 15-25.
- Drew, J. (2010, December 04). Texas supreme court rules state workers' birth dates aren't public. *Dallas Morning News*. Retrieved from <http://www.dallasnews.com/news/politics/texas-legislature/headlines/20101204-texas-supreme-court-rules-state-workers-birth-dates-aren-t-public.ece>
- Halstuk, M. (2000). Blurred vision: The supreme court. In C. Davis & S. Splichal (Eds.), *Access Denied: Freedom of Information in the Information Age* (1 ed., p. 121-153). Ames: Iowa State University Press.
- Hoefges, M., Halstuk, M., & Chamberlin, B. (2003). Privacy rights versus foia disclosure policy: the "uses and effects" double standard in access to personally-identifiable information in government records. *William and Mary Bill of Rights Journal*, 12(1), 1-65.
- La Raja, R. (2007). Sunshine laws and the press: The effect of campaign disclosure on news reporting in the american states. *Election Law Journal*, 6(3), 236-250.
- Locke, J. (1689). *An essay concerning human understanding* . ILT Digital Classics. Retrieved from http://www.ilt.columbia.edu/publications/locke_understanding.html
- Macias, W., Springton, J., Lariscy, R., & Neustifter, B. (2008). A 13-year content analysis of survey methodology in communication related journals. *Journal of Current Issues and Research in Advertising*, 30(1), 79-94.

- Milton, J. (1644). *Areopagitica a speech for the liberty of unlicensed printing to parliament of england*. The Project Gutenberg. Retrieved from <http://www.gutenberg.org/files/608/608.txt>
- Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, 73(1), 74-97.
- Ramaprasad, J., & Rahman, S. (2006). Tradition with a twist : A survey of bangladeshi journalists. *International Communication Gazette*, 68(148), 148-165.
- Siebert, F., Peterson, T., & Schramm, W. (1963). *Four theories of the press*. Urbana and Chicago: University of Illinois Press.
- Snyder, S., & Purcell, D. (2012, March 30). [Web log message]. Retrieved from <http://ire.org/blog/ire-journal/2012/03/30/crimes-classroom/>
- Wimmer, R., & Dominick, J. (2006). *Mass media research: An introduction*. (8 ed.). Boston: Wadsworth.
- Yerardi, J. (2013, January 17). [Web log message]. Retrieved from <http://www.ire.org/blog/transparency-watch/2013/01/17/san-antonio-withholds-employee-identifiers-backing/>
- Yu, J. & Cooper, H. (1983). A quantitative review of research design effects on response rates to questionnaires. *Journal of Marketing Research*, 10(1), 35-44.
- In re* Report and Recommendations of the Judicial Management Council of Florida on Privacy and Electronic Access to Court Records, 832 So. 2d 712, 714 (Fla. 2002) http://www.floridasupremecourt.org/pub_info/summaries/briefs/02/02-659/Filed_12-17-2001_ReportRecommendations.pdf

Questionnaire that was emailed to reporters

Data and Open Records Questionnaire

Being denied data because of the fear of privacy violations, both real and imagined, is a common experience for reporters. These denials can have a disastrous impact on a reporter's ability to write stories and hold government accountable.

What's been your experience?

My name is Fedor Zarkhin, and I'm a Master's student doing research on data and open records for my professional project at the University of Missouri School of Journalism.

Through crowdsourcing I aim to find both the trends in these denials and successful ways to overcome them. An analysis of the results will be pitched to a trade journal such as the IRE Journal or Uplink.

Please tell me about a records request you filed for data that included personally identifying information (PII). PII includes but is not limited to birth dates, addresses, employee numbers and phone numbers.

Choose either the most memorable and challenging request or, if nothing comes to mind, the most recent one.

If you provide personal information, it will not be shared or published without your permission.

Thank you for your time. Please email me or my supervisors if you have any questions.

Fedor Zarkhin

Palm Beach Post

fedorzarkhin@gmail.com

Supervisors:

David Herzog – Associate professor at the University of Missouri:

herzogd@rjionline.com

Scott Swafford – Associate professor at the University of Missouri:

swaffords@missouri.edu

Mark Horvit – Executive Director of Investigative Reporters and Editors:

mhorvit@ire.org

Joel Engelhardt – Investigative Editor at the Palm Beach Post: jengelhardt@pbpost.com

The Basics

General information about you and your news outlet ***If you want to go to a previous page and change your answers, DO NOT click "Back" in your browser. Click "Back" at the bottom of the questionnaire, instead***

Would you be willing to be contacted for follow-up questions? *

You can change this later if you change your mind.

This is a required question

First name

Last name

Age

Gender Male Female

Title

Years as a journalist

Email

News outlet's state *

News outlet name

Medium *

Check all that apply Television Newspaper Magazine Online Other:

Circulation or viewership

The Data Request

For these questions tell me about a DATA request involving personally identifying information (PII) that you submitted. It doesn't matter if the request was granted or denied. Choose either the most memorable and challenging request or, if nothing comes to mind, the most recent one. If you are still working on stories based on this request, please tell about the most recent such request that would not reveal what your paper is working on.

Why did you choose this request for the survey? Most recent request Most memorable or challenging Other:

Date of request

This can be approximate

What kind of agency did you request the data from? Local State Federal Other:

Agency category Law

enforcement Education Military Corrections Transportation Environment Business Fire H
ealth Courts Other:

Agency name

What data did you request and why?

Be as specific as possible.

What personally identifying information (PII) was in the data? Names Addresses Birth
dates Social Security Numbers Phone numbers Juvenile records Other:

Did you need the PII or just the data kept in the same database or table as the PII?

PII – Personally identifying information Needed the PII Did not need the PII Other:

Did you end up getting the data? * Yes No I got a version of the data, but not exactly
what I requested

Data request details

Last page!

How did the agency justify the denial?

List all the reasons, even if the final reason wasn't the same as the first. List statutes, if
available.

How did you negotiate for the data?

Please give a narrative of the negotiation and an explanation of why you did not
succeed.

If you got a version of what you wanted, what did you get?

How was what you got different from what you wanted?

When did you make the last attempt to get the data?

This can be approximate.

What other requests did you file to get closer to your goal? Data layout Data

sample Data from other agencies Other data from the same agency None Other:

How did the denial, in whole or in part, affect your project?

Did you have to drop it? Settle for something less?

Who helped you with the request? Editors Lawyers Data reporters Open records

specialists Nobody Other:

How did these people get involved?

Please copy-paste a copy of your records request.

What stories did the data help produce?

Please provide a link to one or more stories.

Additional comments or observations

Data request details

When did you get the data?

How did you succeed? *

Please describe how you negotiated for the data.

If the agency tried to deny the data, what were its reasons?

List all the reasons they gave throughout the negotiation process.