# CHAPTER THIRTEEN
# The Quest for Electronic Data: Where Alice meets Monty Python meets Colonel Jessep

*Jim Rankin*

## Introduction

From a government perspective, releasing paper documents is about as painless and routine as going through an airport metal detector. Forking over raw electronic data to a journalist? Well, that's a full body X-ray, strip search *and* a finger up you know where. It's indeed the Full Monty of freedom of information requests, and things tend to get weird, fast. Also – as will be highlighted in this chapter in a few personal war stories and those of fellow Canadian journalists – data requests can be lengthy, surreal, absurd, obstructionist, secretive, patronizing, hilarious, infuriating and costly. By the time some of these requests result in disclosure, the Stanley Cup may have been decided five times, the Federal government may have changed three times and been prorogued twice, there may be a new Mayor, new Premier, and new Chief of Police, and there are several changes in top newsroom management. No wonder, then, that so few journalists use our access and freedom of information laws to go after government-held data.

There are, I estimate, fewer than a dozen Canadian media organizations that embrace what is known as computer-assisted reporting and employ journalists who regularly use access laws to get at government data – and know how to analyze it. It is a small club. We are masochists. We commiserate, yes, but we also cheer each other on whenever another's access or FOI odyssey – which begins with a letter and, depending on the jurisdiction,  a $5 cheque, but can take a lot more money and varying numbers of lawyers and years to complete – results in a blockbuster story or groundbreaking freedom of information decision that makes it easier for us all.

As flawed as the laws are, access requests for data do yield results and lead to institutional openness.

The CBC's David McKie fought and won access to a Federal database used to track adverse drug reactions and the CBC's investigative team questioned whether enough was being done to protect the health of Canadians (CBC News Online 2004). Not long afterward, Health Canada made the data available to the public on its web site. In a joint project, former *Hamilton Spectator* reporters Fred Vallance-Jones and Tamsin McMahon, and the *Toronto Star*'s Robert Cribb waged a successful battle for a Federal aviation database that tracks close calls, and that dataset, too, is now publicly available (Hamilton Spectator 2006). The *Vancouver Sun*'s Chad Skelton got his hands on Vancouver's parking ticket data and made all of it available on-line in 2009 in an interactive project (see Vancouver Sun 2009a) that, at time of writing, had generated more than 2 million hits. Another of his on public salaries (see Vancouver Sun 2009b) has hit the 2 million mark.

As for me, I've worked with some terrific *Star* colleagues on projects that began with requests I made for data. We've used inmate address data to illustrate where we are spending the most money to incarcerate citizens and question penal policy that will put more people from already socially-disadvantaged groups behind bars, for longer, for more crimes (Toronto Star 2008). We also obtained Toronto police data that showed how people were treated after arrest and who was more likely to be stopped for certain traffic offences (Rankin et al. 2002). That group effort led to a human rights inquiry into the impacts of racial profiling in Ontario (see Ontario Human Rights Commissioner 2003) and a sparked a sea change in how Canadian police view biased policing. Many police services are taking steps to deal with it.

In an age when governments are turning more and more to e-records and databases, and using hard data to make more and more policy decisions, such as anti-poverty strategies, police deployment choices and budget cuts that affect small and large segments of the general population, unfettered public access to e-records is of utmost importance. Similarly, institutions may be ignoring or missing out on stories buried within own databases. To deny access to e-records is to shut down the ability to scrutinize government decisions.

This chapter is about flawed, wacky systems, but ones that *can* be worked to expose broken systems, identify corruption and bias, broaden understanding of how our society works and hold governments accountable and result in positive social policy changes. On the pages that follow, I will, along with some of the best data-based investigative journalists in the country, flag problems and share lessons and practical advice on how to navigate the system. For information keepers, this chapter will provide a sense of what it is like on the requesting end, and what can be done to improve the process for both the requester and institution. I will also provide a toolkit for requesters, which, due to the constraints of this chapter, is far from complete. The process may seem intimidating to the uninitiated but it begins with a simple question, I encourage those who have no experience with requests to seek help from the journalists you will read about in the coming pages. They are a solid and sharing lot.

The chapter will also look at the perils, particularly for academics, of suing the government for raw electronic information. The chapter concludes with a brief look at the future of data access requests and investigative journalism.

## The Truth: Why Data Requests are Difficult

“You can’t handle the truth,” Colonel Jessep, Jack Nicholson’s character in the dramatic thriller *A Few Good Men*, shouted at the young naval attorney, played by Tom Cruise. He wanted the truth, but may as well have been asking for raw electronic data from government; the reaction often seems the same. Most institutions are naturally suspicious of access requests, and enormously more so when it involves raw data. Like Colonel Jessep, they *really* don't trust you, at least not initially.

Skillful negotiation and brokering for access can be fruitful in some cases, but the relationship between journalists and access coordinators can be adversarial from the get go.

"I think there's a great deal of concern within government about allowing internally maintained data from being placed in the hands of outsiders who can use it to draw specific conclusions according to their own analysis or even pair data with a different data set to find links," says Robert Cribb, an investigative journalist at the *Toronto Star* who has won many freedom of information battles for data and is a past president of the Canadian Association of Journalists. "It represents a loss of control far more profound than the simple release of documents, which are static and whose conclusions are plainly evident and easy to prepare for should they reach public attention. All of this is, of course, a natural reaction by government."[1]

In Canada, the advent in 1982 of the ATIA and a spirit of openness soon spawned efforts to hide or omit information and "message discipline," as explored by Alasdair Roberts (2006) in *Blacked Out: Government Secrecy in the Information Age*. Requests became more probing and the handling of them more political, which in turn caused more probing by requesters. What exactly were ministers and departments doing to manage the message? Through the 1990s, the Liberal government used litigation to try to block access to records held by in the offices of cabinet ministers and the prime minister's office. The government also attempted to hide information by shifting responsibilities to quasi-governmental bodies not bound by the act.

Sensitive requests were also being flagged, or “amber-lighted,” as journalist Ann Rees discovered in an access request that led to the release of documents relating to how the government reacts internally. Requests from journalists and opposition parties received special and immediate attention, with the Minister’s Offices receiving notice of such requests within a day of receipt. Certain requests resulted in the pre-emptive creation of “media lines” and “house cards” that could be called upon by department spokespersons and ministers should the requests lift the lid on something the government would like to massage, such as stories about mismanagement.

These measures, noted Roberts (2006), resulted in delay of processing of sensitive requests, which provide an advantage to government. The government, in other words, is “on message” when the you-know-what hits the fan. This, for journalists, is to be expected. The subsequent filing of an access request on how a government department behaved behind the scenes in response to your request is not always newsworthy, but well worth the $5, if for nothing more than the entertainment value alone. To go back to the airport security analogy, it’s the fear. What might a third party find out about you that you might not know yourself? Or, maybe you’re hiding something, like love handles or exploding underwear.[2] Either way, scary – or potentially embarrassing. Possibly, game changing.

> ***Tip**: You may not want an institution to know why you want data but, whenever possible, be transparent. While not required, you may ease some of the suspicion by letting an institution know what you intend to do with the data.*

Another hurdle is the relative newness of data requests. It is worth noting here that, in my experience, despite the differences in the language describing the various acts and laws – “access to information” and “freedom of information” – data requests are, by default, less about openness and more about control. Also, many institutions still don't know how to handle them. So, they simply begin by saying no.

So, how to get to the data? Data requests, as Cribb points out, are more "surgical" than document requests, "largely because the bureaucratic default position is even more firmly switched to the 'denial' position." While documents, says Cribb, have been the historical "currency" of access disclosure, data release is still seen by many ministries as foreign, discretionary or beyond the scope of the Act. And, thus, are met with resistance. "This likely has much to do with the ability of data to be manipulated according to the wishes of the requester," says Cribb. "This brings with it a loss of control that many within government appear to find threatening. In any case, there's no question that a large data request is an exercise in patience and persistence." It is almost always a two-step request process rather than a single request, which is typical for documents. In seeking data, begin by seeking a record layout, sometimes informally, but often through a formal request, which tells you what is in the database and allows you to filter out sensitive data, such as personal information.

> ***Tip****: Access and freedom of information laws are complicated. School yourself on the act that applies to your request. Search privacy orders and decisions that apply to, and support, your arguments for the release of the data. Cite them in your requests. Always appeal a bad decision. Hold institutions to mandated deadlines. Typically, an institution is required to respond to a request within a month. Count the days. Some journalists use a spreadsheet to track their requests.*

Institutions sometimes feel that the data it collected was never meant to be analyzed or released in electronic format. "I think a lot of officials thought databases were internal resources and that only summaries from the data should be released externally," says former *Spectator* reporter

Fred Vallance-Jones, now a journalism professor at the University of King's College in Halifax. "So, right away there was a reluctance. This held across the board – from municipalities right up to federal departments." Case in point: In 2008, an annual audit of freedom of information laws (which Vallance-Jones is involved with) saw journalists from across the country file similar requests. The audit found institutions at all levels would release paper records only – when electronic records had been expressly requested (see Canadian Newspaper Association 2009).

Chad Skelton, a reporter with the Vancouver Sun, only recently started making requests for raw electronic data but he quickly had a number of successful hits. However, not all were without some fight involved or institutional holdback on electronic release. A request for British Columbia government salaries in electronic format was met with refusals by agencies that insisted instead on sending paper printouts from what were clearly Excel electronic spreadsheets. The government itself fought the electronic release, despite a privacy ruling in favour of the Sun that said agencies should release data electronically. "The argument I've gotten from some agencies is that they have a policy not to release records in a 'modifiable format,'" says Skelton. "In other words, if they give me a printed list of salaries, it's harder for me to fiddle with the numbers than an Excel spreadsheet. I frankly think this argument makes no sense."

> ***Tip**: Once you have the data, do not assume anything. Question the data, probe it for weaknesses and bulletproof your analysis. If you lack expertise in this area, visit your IT department. Chances are, there is someone there who can help and will be interested by the task at hand and want to do so. Librarians are also quite often versed in spreadsheet and database software. Keep an open line with the institution and ask follow-up questions.*

It's unfortunate that these fights continue to play out, despite decisions by many privacy czars and courts across the country that say institutions must provide requested records in electronic format. Like Bill Murray's TV weatherman character in the movie *Groundhog Day* – trapped in a loop where he wakes every day to the same day, and must cover the emergence of Punxsutawney Phil from a hole in the ground – journalists are forced to make the same arguments over and over again.

The result? Accessing electronically-held government data in a timely manner, typically, becomes impossible. In a news culture growing overly consumed by the feeding the maw of a 24/7 news cycle, your request for data will not be ready for the evening news. Try, instead, years from now. Governments gain time, as Roberts (2006) points out, to get its message straight and prepare for potential negative fallout. Some requesters simply abandon their efforts. Others carefully word their requests to make it clear they want the most up-to-date data at time of release, whenever that may be (this is not possible in all jurisdictions.) "I think the cultural resistance within most government departments is single largest obstacle to accessing data.," says Cribb. "Despite numerous rulings reaffirming the public's right to government data, departments again and again issue denials that force lengthy appeals that inevitably come to the same conclusion." In the meantime, says Cribb, months or years are lost with no public access to sometimes "vitally important information." A recent two-year fight of Cribb's to access private career college inspection data was denied on "tired, old arguments that have been roundly dismissed by the Ontario information commissioner in numerous other appeal cases, many of which I had fought myself. Those lengthy, hard-fought battles appear to have little impact beyond the specific release they trigger. And so, we are left to re-invent the wheel each time in the face of bureaucratic intransigence."

It shouldn't, of course, be this way. In the case of the electronic release of data, any resistance should have melted away following an Ontario court ruling in 2002 that ended a five year fight by *Toronto Star* reporter Phinjo Gombu for electronic City of Toronto campaign contribution data. Gombu wanted to analyze the contributions for possible contribution breaches. Although the contributions were public records that document names and amounts contributed, and could be viewed by any member of the public in paper format, the city refused Gombu the electronic version, saying it had already been published in paper format. Ontario's Information and Privacy Commissioner found that argument to ring hollow but denied Gombu access because the data contained "personal information," such as phone numbers, even though they could be viewed on the paper documents by anyone (see Gombu v. City of Toronto 2000). Gombu took the case to divisional court, and Mr. Justice J. David McCombs, writing for a three-judge panel, set things straight, finding that it would be in the public interest to disclose the electronic data.

McCombs accepted the *Star*'s position, saying "that the only way that (Gombu) can meaningfully scrutinize the information about campaign contributions is through the electronic database." Public interest in disclosure, wrote the judge, outweighed other considerations, including the telephone number information, which the judge noted was already required to be public any way (see Gombu v. Mitchinson 2002). The judge ordered the city to hand over the data. Gombu recalls that the Ontario information and privacy commissioner was set to appeal but suddenly dropped the case. He believes the commissioner was looking for an easier fight on the personal information angle.

***Tip**: Negotiate with the institution. An open dialogue goes a long way, provided, that is, the institution is willing to talk.*

Journalists familiar with data requests have learned to make a few things straight in their initial requests that help speed things along. Although Gombu received personal information, the information came from publicly available records. In most every other instance, names, addresses and other personal information are non-starters, as are free-form text entry fields that could contain names and other personal information. If you don't want this information, make it clear from the beginning that you do not want it.

In the following example, government data keepers sought to end a request by releasing a watered down version of aviation safety data the journalists were seeking.

**Case Study: Collision Course**

Cribb, Robert; Vallance-Jone, Fred; McMahon, Tamsin (2006) "Collision Course." *Hamilton Spectator, Kitchener-Waterloo Record, Toronto Star* – June

**Act**: Federal Access to Information and Privacy Act

**Start to finish**: About five years

Transport Canada tracks aviation close calls in a massive database called the Civil Aviation Daily Occurrence Reporting System, or CADORS, and in 2001, Fred Vallance-Jones, then with the *Hamilton Spectator*, decided he wanted all of it. He made a formal request that set in motion an Alice in Wonderland journey that he learned more about in a follow-up request for internal correspondence that dealt with his initial request. Federal bureaucrats had routinely released paper reports on incidents, most often to aviation industry insiders and the odd journalist. They initially were perplexed by the thought of releasing the whole thing, electronically. The document trail uncovered by Vallance-Jones in the subsequent request provides a glimpse of life inside the information and privacy rabbit hole.

His request resulted in a slew of internal e-mails, set off a series of meetings, led to the production of a PowerPoint presentation, a review of who did receive occurrence reports and whether anyone should get them, period, and sparked debate over whether third party information, such as names of major airlines, would be released. It was thought the data could be "injurious to third parties," such as WestJet and Air Canada, due to a large number of incidents. The data included aircraft make, locations, registration numbers, airline names, narratives of the incident, pilot information. Was this private information or commercially sensitive? The debate went on and on internally. It is also clear the department was misinterpreting the request, and at one point would not release "software." Vallance-Jones had never asked for that, just the data. They also claimed he had amended his request, which he hadn't. "I thought the issue of the database was resolved. Does the requester still want the database?" reads one internal document.[3]

"I don't know whether this wasn't some tactic they were using to slow it down, buy themselves more time," says Vallance-Jones. "That was weird. I never got an explanation for that." Another party involved in the internal debate suggested the data was intellectual property and perhaps could be bought "through proper channels," since the system cost $250,000 to create.

A year after the request, in 2001, a meeting was held. It included ATIP people and database specialists. The seemingly stunned department determined the data was not transferable electronically and just decided to say no. Vallance-Jones appealed to the Office of the Information Commissioner, complaining that the information requested was being shared with stakeholders, so, why not with then public as well? An investigator unfamiliar with database requests was assigned to look into the roadblock but he and the

Commissioner's office went to school, and hired a computer expert to help. This moved things along. But internally, the department was still thinking paper. A draft estimate shows they were considering an electronic release but were also going to ask for $18,000 to cover the cost of manually severing information they didn't intend to share.

"I actually laugh more now when I look back at it, but it was initially very frustrating," says Vallance-Jones. At one point, the department released a stripped down electronic version, which was useless. Vallance-Jones felt they were now simply stonewalling. So, the Information Commissioner went to Federal court. In mid-2005, in the midst of the court challenge, Transport Canada suddenly forked over the commercial data but withheld private aircraft data.

Vallance-Jones and Cribb had previously talked about doing a joint airline project and the pair eventually teamed up with McMahon from the Record. The reporters went to work. To make more sense of the data, Vallance-Jones, Cribb and McMahon poured over the narratives in the data, and classified the incidents to check for patterns of pilot and air traffic controller error. They interviewed officials, insiders, pilots, accident victims and others. Pilot fatigue, among other problems, was flagged as an issue.

The resulting series won the Canadian Association of Journalists' top award for investigative journalism in Canada, the Don McGillivray Award for Investigative Journalism, the CCNMatthews/CAJ Computer-Assisted Reporting (CAR) award, earned Vallance-Jones and McMahon Journalist of the Year honours at the Ontario Newspaper Awards, and received a citation of merit in the Michener Awards for public service journalism.

In 2006, Vallance-Jones received the private aircraft data, and the five-year battle was over. “This wasn’t my worst experience. But it was a horrible sort of long grinding process.”

The internal ATIP documents illustrate how the department shifted its thinking on the release of electronic data, and most satisfying for Vallance-Jones: today, it is available online. The safety issues raised by the journalists continue to reverberate through the aviation community.

**Curiouser and Curiouser: You do have some Cheese here?**

Just when you think you’ve stumbled across an institution that understands requests for raw data – and there are some out there – things turn absurd. Fee estimates for programming can be comically astronomical, well-meaning coordinators dig in, common sense takes a vacation and the spirit of disclosure intended in access and freedom of information acts is all but exorcised.

On fees, my personal favourite is a request I made in May 2003 for a copy of the database used to document Canada’s criminals. It’s part of the Criminal Police Information Centre database maintained by the Royal Canadian Mounted Police. They quickly connected me with their “geek” who knew how the database worked, and what was in it. This hardly ever happens, and if it happens at all, it is late in the process. Getting to the analyst is like finding cheese in the famous Monty Python cheese shop skit, where an obstructionist cheese shop owner frustrates a customer looking for cheese. So, we spoke geek to geek, and in short order it was agreed that the data could be cleansed of personal information that potentially might allow me to identify individuals. Fantastic. I modified my request to eliminate problematic fields.

***Tip**: Speak “geek” to “geek.” Connect as soon as possible with an institution’s database analyst. FOI coordinators are notoriously ill-versed in databases, the software required to extract data and the process of electronically redacting personal information. A good*

> *coordinator will connect you with an analyst who can explain the data and how it works. This speeds the process and helps tailor the request to get what you want. If you don't speak "geek" find someone – a librarian, one of us journalists or an IT guru – to help you.*

And then came the official decision from the Mounties' access coordinator. Yes, this was doable. I got a receipt for my $5 cheque and a request for a deposit on estimated programming fees that would need to be paid to cover the cost of extracting the data. I flipped to page 2 of the decision, and read this: "A total cost of $1,599,840.00 has been assessed for the processing of this request …. If you wish to proceed with the processing of your request, forward a deposit" in that amount (see RCMP 2003).

The letter went on to assure me that this was just an estimate and that the final total may not be as high, and also that "should you continue your request, please note that there are no guarantees that any part of the information will be released, bearing in mind that the material may qualify for exemption under the Act." So, they wanted $1.6 million for the cheese, and there was no assurance there would even be cheese. As Alice in Wonderland remarked, things were getting "curiouser and curiouser."

The fee estimate was based on antiquated regulations that mention costs associated with duplicating "microfilm" and "magnetic tape-to-tape duplication," which I suspected and pointed out in my complaint letter. I also cited a 2002 report to Parliament by the Information Commissioner, in which the office had flagged these outdated fees (see Reid 2002). The Mounties said it would take 200 days of "central processor" time to complete the request. That costs $16.50 per minute, or at least it did when they had machines like that.

I complained to the Information Commissioner. At the time, the Commissioner was reviewing these old regulations, due in part to requests by other journalists for other federal databases. In December 2005, a letter and a CD arrived. It was, mostly, what I'd asked for – electronic summaries of 2.9 million criminal records. The letter suggests it took less than five hours of an analyst's time to extract the data, and stated "we have waived all processing and reproduction fees" (RCMP 2005). From $1.6 million to free. Cheese, after all.

The database, along with two other inmate datasets I obtained from through additional requests, served as the foundation for *Crime & Punishment*, an eight-part series that examined the wisdom of Canadian penal policy and the financial and societal costs of jailing people for problems that would be cheaper to deal with at an earlier stage (see Toronto Star 2008).

> ***Tip**: Challenge programming and other fees. Often, the initial estimates are grossly more than what it will actually cost to extract and release the data. If you feel the dissemination of the data is in the public interest, request a fee waiver or fee reduction. Be prepared to argue why. Don't know how? Get help.*

Often the tone of how a data request will proceed is set by the co-ordinator handling the request, and there are good ones and not-so-good ones. Some are well schooled on the act and seek to use it fairly. Others see this as war and are prepared to use every weapon available to thwart a request. Yet another group is motivated, I think, by fear of reprisals from higher-ups, and this is understandable and justified.

To be fair, we journalists, as former Bruce Mann (1986), former Assistant Information Commissioner of Canada put it, do indeed like to poke sticks at the federal information coordinator, and the same is true of their counterparts at the provincial and municipal levels. Just three years into the fledgling existence of the ATIA, Mann likened the federal coordinator to the meat in the sandwich, which gets worse when you don't know who will take the next bite.

Coordinators are seen as obstructionist and the face of government by journalists, a "thorn in the side" by department colleagues who must take time to discuss matters of disclosure, a threat to third parties with corporate interests (such as airlines in the case of Vallance-Jones's request) and a "consummate slave-driver" to support staff for trying to adhere to deadlines and the letter of the law. As for the Information Commissioner, any investigation could result in an over-ruling of the coordinator. And what of the boss? What if the release of information causes embarrassment or scandal to government?

Tough job, and one that varies in its demands, according to the department and number of requests.

The *Star*'s Robert Cribb, who started making data requests in 1997, says he can tell five minutes into a telephone call with a coordinator which camp they fall into.

"Good FOI coordinators do exist; they are fair, open and feel vested in playing an advocacy role for the public in mediating between requesters and departments," says Cribb. "The good ones ask questions, listen intently and respond in a timely way, not simply with blanket denials or excuses, but also suggestions and alternatives that move the process forward. In other words, they mediate, problem-solve and seek a solution that ultimately serves the spirit of the legislation under which they work." More typically, says Cribb, access coordinators are firmly entrenched defenders of their department, seemingly anxious to thwart any efforts that could bring them – or their bosses – any headaches in the press."

***Tip**: Coordinators unschooled in databases often cite the size of a database in what feels like an effort to persuade you to tighten the scope of a request or suggest it would be a*

> *burden to respond to it. Size, however, matters little. It takes a programmer the same amount of time and effort to write queries to extract data from a database containing 100 records, as it does 2 million. Don't buy the volume argument. It's a red herring.*

There is, of course, a balance that must be struck between disclosure and protection of privacy, and even us journalists understand this. But sometimes, particularly in data requests, well-intentioned coordinators become hyper vigilant about the possibility of determining who people are, and in some instances hamper the ability to meaningfully interpret and study government-held data. Imagine a dataset that has had names replaced with unique, randomly-generated numbers, exact dates replaced with month and year, or year only, and exact addresses are removed. This dataset on its own does not allow one to determine individual identities. But connect it to another database and theoretically, it may be possible to determine who some of the individuals are – damn hard, and costly to do so, but theoretically, in some cases, possible. Known as the "mosaic" or "matrix" effect, some institutions are using the argument that the release of their dataset may allow one to connect dots that are in other domains and beyond their control, and determine who people are.

I've run up against this a number of times, most notably with Ontario's Ministry of Community Safety and Correctional Services in a quest to obtain the full home postal codes of inmates who have been sentenced to a term of less than two years. I asked for a one-day snapshot of who was sitting in jail, and only for two bits of information about each inmate: the length in days of their sentence; and full postal code of the last address they had before entering jail. I asked that the snapshot come from any day of the ministry's liking in 2007. I did not ask for the day. I did not ask for the crime for which the inmate was sentenced. I did not ask for age, gender, or any other pieces of information. Our intended use of the postal code and sentence lengths was to calculate and map high incarceration pockets of Toronto. With full postal code, we could then look at underlying demographic and socio-economic factors at play in these areas, and also put a cost to what we spend in these areas to jail people. The ministry denied full postal code, saying it was personal information that could be used to identify inmates. Instead, we got the first three digits. I appealed and won access to the full postal code data. We produced the map we wanted to, and it revealed a clearer pattern of costly jail areas and underlying social conditions (see Toronto Star 2009).

> ***Tip****: More and more institutions are allowing the public to access data online through one-off search queries but do not make the entire dataset available for download. Rather than making a formal request for the data, it may be possible to create an automated script that lifts – or "scrapes" – the entire dataset from the institution's web site. Don't know how? Find a geek.*

Skelton sees the "mosaic" effect as a real threat to future data releases. In his words, it drives him "bonkers. I've had it applied in some pretty ridiculous cases. For example, awhile back I asked for disciplinary records from a bunch of professional associations. I knew I wouldn't get the disciplined members names, so I asked for the details on their offence and punishment with the names redacted. A few of the associations balked. They argued that their membership was so small that even releasing something generic like 'Massage therapist B sexually assaulted his patient' could be an invasion of privacy. Their reasoning, says Skelton, is that even without other identifying information – like where the person worked, or even what city they were in – if the patient or someone familiar with the case read the story they'd probably be able to identify who the case referred to.

Programming fees are another area of concern. While the access and freedom of information acts are being modernized to reflect the modern day ability to extract and modify data with off-the-shelf software and a laptop, costs do add up and fee estimates remain out of reach for most private citizens, and are becoming an issue with news organizations as well as budgets shrink. Requesters can claim financial hardship and public interest overrides, but these arguments are routinely dismissed. The existing laws, says Cribb, are "entirely ill-equipped" and open to the "whims of the bureaucracy in conjuring up extortionate fees and expansive timelines can become authoritative policies, beyond questioning."

Kevin Donovan, a *Toronto Star* journalist and the paper's investigations editor, has overseen several successful fee appeals and waivers, but has grown frustrated by the "Groundhog Day" phenomena to the point he avoids formal requests all together. "On a recent story I convinced a government official to give me data. It took a lot of work, but ultimately much easier than getting it (through the act). I also try and find data that I can prove is releasable without an FOI or ATIP request. Charity is a good example. The tax returns of all 83,000 charities are in electronic form and each year I get a CD with all the data."

Andrew Bailey, the *Star*'s database specialist feels the same way, and has taken to web scraping as a way to bypass the formalities. "Rather than try to dance my way to the whole," he says, "I grab each record one at a time and rebuild them into a whole for analysis."

### Case Study: Race & Crime

Rankin, Jim; Simmie, Scott; Shephard, Michelle; Duncanson, John; Quinn, Jennifer (2002) "Race & Crime." *Toronto Star* – October 19, 20, A1. October 26, 27, A1

**Act**: Ontario's Municipal Freedom of Information and Protection of Privacy Act

**Start to finish**: Initial request: 2.25 years; Follow-up request: 7 years

The genesis of this investigative series on race, policing and crime in Toronto was a routine police blotter item in 1999, about a routine crime and an unusual description of a suspect's skin colour: yellow. As it turned out, yellow was code for Asian, pulled, embarrassingly so for Toronto police, from an internal database that tracks people who had been booked and fingerprinted for previous crimes. That got me thinking about what police document in terms of race and ethnicity. A 1989 policy forbids Toronto police from analyzing race-based data out of fear it might be used to stigmatize communities, but that didn't mean that they didn't collect it, and it didn't mean we couldn't have a look at it.

After informal attempts to obtain police data failed, we made a formal request in March, 2000, asking for record layouts for two internal police databases, and the data itself. Police denied both requests, and, in the case of one of the databases, which tracked police contacts with the public We appealed the denials and focused our efforts on one of the two datasets that details arrests and charges and certain non-criminal offences, such as traffic tickets.

In 2001, another IPC ruling in our favour led to mediation, which proved most productive. Working together, the police FOI coordinator, the police analyst in charge of the database, an IPC mediator and myself found solutions to privacy concerns. In June, 2002, police handed over the data on a single CD, and asked for $800 for programming, an amount they said they reduced because of the length of time the request had taken. They would never be that charitable with me again. It's also worth noting that we did not need legal help in this request.

We spent several months “interviewing” the data. Almost immediately we saw patterns in terms of arrest outcomes for black people facing certain drug and other offences. We also looked at people ticketed for a driving offence that we had coded as “non-moving.” These were offences that an officer would discover after a traffic stop, such as not carrying a licence, or having no insurance. We isolated those facing only one such offence in a single incident, and found that black motorists were ticketed at a rate three times higher than the proportion of blacks in Toronto’s population. We looked deep into the data to see if other factors were at play, and there were indeed. A previous record, for example, affected whether you were released at the scene or held for bail. But, nothing in the data could make the race factor go away. Prior to publication, we had a statistical expert from York University go through our work and replicate the analysis. A team of five reporters spent a month reporting and writing, and brought the numbers to life.

Over two weekends in October, 2002, we published the series in paper and online. Reaction was huge, mixed and heated. Toronto’s black communities embraced the series as affirmation of what they anecdotally had been complaining of for decades. They also acknowledged another part of our analysis that showed blacks were being disproportionately charged for violent crimes and renewed calls for social policy changes that would help youth in at-risk neighbourhoods. Police, however, denied there was a problem. The Toronto Police Association first called for a boycott of the *Star*, and followed that up with a class action libel suit, alleging the series had branded every member of the service as racist. They sought $2.7 billion in damages (the action was dismissed by two levels of court, and died in 2004 when the Supreme Court of Canada refused to hear the case, thus affirming the lower court decisions). There were calls for a race relations summit and the Ontario Human Rights Commissioner announced an inquiry into the impacts of racial profiling across all of society (see Ontario Human Rights Commissioner 2003). Meanwhile, the police hired an academic and a lawyer to analyze our analysis, and a sideshow of dueling statisticians began. The police experts called the *Star* analysis “bogus” and “junk science” (see Gold et al. 2003) . A later academic study of the police experts’ analysis of our analysis found fundamental flaws in that one (see Wortley and Tanner 2003).

The series was lauded by many, and in 2003, won a National Newspaper Award for investigation, the Canadian Association of Journalists’ CCNMatthews/CAJ Computer-Assisted Reporting, and the Governor General’s Michener Award for public service journalism, the highest honour in Canadian journalism. Attitudes have since changed dramatically. Police across the country acknowledge that racial bias is a problem, as it is in any aspect of society. Toronto police partnered with the human rights commission to find ways to improve hiring, promotion and retention of minority officers, and at ways to improve how they police.

I made a follow up request in 2003 for updated arrest and offence data, and renewed the request for the database that tracks who police stop and choose to document, in mostly non-criminal encounters. In early 2010, nearly seven years and many appeals and two court challenges later (both the Canadian Civil Liberties Association and Ontario Information and Privacy Commissioner were interveners at the latter Court of Appeal stage), Toronto police handed over the datasets.

> We needed legal help this time. The final Court of Appeal decision (see Toronto Police Services Board v. Ontario Information and Privacy Commissioner 2009) was a win for journalists and others who wish to obtain electronic data. It clarified what an institution must do to extract electronic data and corrected a lower court decision (see Toronto Police Services Board v. Ontario Information And Privacy Commissioner 2007) that would have hamstrung future data requests. Police were ordered to pay our legal costs, and, perhaps was why this time they were not cutting any deals on programming fees. They billed us for $12,000. The *Star* paid the fee and obtained the data and at time of writing was appealing the fee.
>
> We published *Race Matters*, the follow-up series in February, 2010. The lead piece, based on our analysis of police stops in mostly non-criminal situations, showed that black people in Toronto were 2.5 times more likely than whites to be stopped, questioned and documented by officers (see Rankin et al. 2010).

So, we've now heard of some victories and the possible application of data analysis is obvious for researchers who might be refused informal requests for raw data. But many of these victories by journalists well-versed in the access and freedom of information laws came following expensive legal maneuverings prohibitive to most, academics included.

### To Sue or not to Sue? *Send Lawyers, Guns and Money*

The late singer-songwriter Warren Zevon wrote of a guy in deep trouble following gambling problems in Havana, unspecific girl trouble involving Russians and, as a result, was in hiding in Honduras – in need of lawyers, guns and money. Unhappy with an information and privacy ruling, one can go to court. But if unprepared to self-represent and argue in court about the intricacies of access laws you will require both lawyers and money.

There is in the access request arsenal the option of going to court. In federal, provincial and municipal acts, a requester – or government department – can take a beef with a decision to either deny or release information to a “higher” level. This has mixed results, and will lead to victories and defeats All of this can cost a lot of money, as Matthew Yeager (this volume), an academic, illustrates in his study of two cases he was intimately involved in – one in the United States and one in Canada. As a researcher, he waged a 17-year-long battle for raw Drug Enforcement Agency data, which ultimately did not result in disclosure and left behind a horrible precedent – agencies were not required to use electronic redaction techniques, such as eliminating fields that were deemed “sensitive” or containing personal information – in responding to a request for government data. This is similar to the Ontario Divisional Court decision I was involved in during the *Star*’s successful bid to get police arrest and stop and search data. In my case, a higher court over-turned that decision. In Yeager’s fight, he lost but years later the FOIA was amended and electronic redaction was normalized.

In Yeager’s Canadian request, he lost a bid for Correctional Services of Canada inmate and parole data that went as high as the Federal Court of Appeal, and was left personally holding the bag for a $21,000 legal bill.  Perhaps most troubling for Yeager, who was then a graduate student at Carleton University – and for academics toying with the idea of suing for data – Yeager wasn’t able to get financial help from “university-connected, non-profit agencies,” nor was Carleton offering help. While the university was “silent” on the issue, Yeager wrote that he learned that the institution was “hostile to the project.” Taking on the government, a source of university funding, can make one a “pariah within the academy,” wrote Yeager. Suing the hand that feeds, in other words, is frowned upon.

As Yeager concluded, suing the government for information “clearly represents an extreme form of applied ‘conflict’ theory.” Governments oppose the release of data and don’t want outside researchers “rummaging” through raw data and “elite state interests are supported by the courts, whose members are appointed by those same elite interests.” As a methodological tool, however, suing for data is a must, wrote Yeager (this volume), since it is the government that “largely controls the data, the funds, and the research agenda.”

As highlighted by the experiences of Yeager and Canadian journalists, taking matters to court often requires money and expertise, and even then, the courts can get it absolutely wrong. The government, on the other hand, has plenty of money and expertise.

There is also a fear in academy that using access requests and other social research tools such as surveys and interviews, as employed by journalists, may result in an “ethics creep,” where researchers become more “journalist” than academic, as highlighted by Kevin Haggerty (2004), an associate professor of criminology and sociology at the University of Alberta. Journalists may pursue similar information but the means and goals may differ, and the work of journalists are not scrutinized to the same degree as academics. Academics tend to draw broader conclusions than journalists, writes Haggerty, and “employ more sophisticated statistical tools,” yet use similar methodologies. “There is a heightened concern in the academy about the ethical implications of forms of knowledge production that, when performed by journalists, raise few, if any, ethical concerns,” writes Haggerty. “Ultimately, this raises the provocative question of whether university ethical protocols are making it easier to produce certain forms of knowledge as a journalist rather than as a university affiliated researcher.”

Haggerty argues for a reform of academic research standards and protocols that would allow researchers to dig deeper, and he calls for a reconsideration of whose interests are being served by the current protocols and at what is “being lost as a result.” “However, my suspicion is that the systematic creep of the ethical structure will continue its expansionist dynamic and the bureaucracy will become larger, more formal, and more rigid,” he writes. “The more ethical roadblocks (that) are installed for innovative and critical research, the more we risk homogenizing inquiry and narrowing vision, as scholars start to follow what they perceive to be the path of least institutional resistance.”

Journalists are indeed not bound by the same standards. We are not academics, but are increasingly employing similar social science tools. With the exception of publicly-funded news operations, we work for employers looking to turn a profit. We do things differently. We may go undercover. We do stake-outs on unwilling subjects. Our rules of engagement, as Haggerty points out, are different than an academic seeking to do social research. Nonetheless, investigative journalism is about telling stories that will be consumed by the general public and, due to space and time restraints, the stories must get to the point. The stories themselves – and how journalists go about collecting them – are subject to libel and defamation laws. We do adhere to ethics codes. We also tend to go hard after institutions, and part of that is through seeking raw government data. And we don’t typically care if the process or the end result pisses off the government or embarrasses it into fixing something that is broken. That is the point of it all.

## Conclusion

While the frustration level and costs can be high – and the access and freedom of information acts are wanting when it comes to requests for raw electronic data – journalists have been successful in using these tools to pinpoint and expose flaws and cause change. The examples in this chapter prove that the laws can be successfully navigated. But, what does the future hold in store for such requests by media?

Advertisers are abandoning traditional media. Newsroom budgets continue to shrink. Desks sit empty. Computer-assisted reporting is viewed by some journalists and bosses as a time suck and expensive. What this means for the future of data requests and investigative journalism in general is unclear.

In a January 2009 lecture at Carleton University on the future of newspapers, John Honderich, publisher of the *Toronto Star* in 2002 when the series on Race & Crime was published and now chairman of parent company Torstar, lamented the impact the Web has had on newspapers and questioned whether serious journalism, which requires money and resources, would survive. But he also pointed to other models worth watching – foundation-funded projects and not-for-profit journalism among them. On the Race & Crime series, he had this to say: "It took fortitude, patience and hundreds of thousands of dollars. Was it worth it? You bet. We nudged the world a little bit" (see Reid 2009).

Other traditional news media outlets engaged in serious journalism are facing the same uncertain future. Yet, with more and more people turning to the Web and digital devices for news and information, digital applications for data are proving to be a reason for optimism. The *Vancouver Sun*'s Skelton points to two recent projects he was involved in as proof. The online hits are in the millions and continue to accumulate on the parking ticket and salary disclosure data projects he was involved in. Another important things to note, once you have won the data, requests can be made for updated data with a fraction of the hassle and cost. In other words, the shelf life of these database projects is infinitely longer than a series published in print over several days. "If done right – database journalism projects can drive much more traffic to news organizations' websites than a typical series of stories," says Skelton. "I don't think we've ever had a story on our website – even if it's about Jon & Kate – that got 2 million hits. So these projects hold out the possibility of a big bang for your buck."

The dynamite question, of course, is whether hits can be translated into profit that will support journalism on the Web, or on other content delivery mediums, such as smart phones and tablet computers.

Cribb, too, sees a future for data-based journalism, albeit a contracted one. "It's obviously tough times with shrinking resources and less patience for long-term projects, all of which might suggest the future for computer-assisted reporting is dim. That said, the commoditization of news has made investigative, contextualized reporting a method of differentiation for big-picture news organizations looking to distinguish themselves from the competition. Exclusive stories based on hard-fought data and analytical skills has the power to set the public agenda, change public policies and laws and bring distinction to a news organization with unmatched force. It's harder to do this work in tough times. But tough times likely represent the best opportunity."

Data requests represent an opportunity, and more sectors could and should be using them. It is, as illustrated in this chapter, often a battle to free raw electronic data. While there are indeed elements of Alice, Kafka and Monty Python, it is, above all else, a war – but a tool that should be used and employed to get something otherwise unattainable.

Brokering for access is the theme of this volume, and it indeed is a vital element in requests for raw data as well. However, unlike requests for paper documents, a battle is also to be expected. On a positive note, data requests under the acts, as flawed as they are, have led to more proactive disclosure on the part of institutions, which has brought greater transparency and enhanced accountability to the public.

Outside of journalism, the use of access and freedom of information acts to obtain and analyze raw electronic data presents near limitless areas of study for graduate students, human rights groups, lawyers, criminologists and other academics – provided, that is, that they have a masochistic bent.

**Notes**

[1] Unless noted otherwise, all quotes are from interviews by the author or from e-mail responses to the author's questions.

[2] A reference to the December 2009 arrest of Nigerian Umar Farouk Abdulmutallab, the so-called "underpants bomber" charged with trying to detonate explosives in his underwear aboard a U.S.-bound airline flight.

[3] Internal correspondence provided to the author by Fred Vallance-Jones. Includes copies of e-mail, handwritten notes and letters.

**References**

Canadian Newspaper Association. 2009. "Annual Freedom of Information Audit." can-acj.ca – January 10, retrieved from http://www.cna-acj.ca/en/news/public-affairs/cna-releases-4th-annual-freedom-information-audit

CBC News Online. 2004. "Faint Warning." CBC.ca – February 17, retrieved from http://www.cbc.ca/news/adr/index.html

Gold, A., E. Harvey, and R. Liu. 2003. "An Independent Review of the Toronto Star Analysis." TorontoPolice.on.ca – March. http://www.torontopolice.on.ca/publications/files/reports/harveyreport.pdf and http://www.torontopolice.on.ca/publications/2003.02.20-review/presentationsummary.pdf

Gombu v. City of Toronto. 2000. Order MO-1366, Decision of the Ontario Information and Privacy Commissioner, November 23

Gombu v. Mitchinson. 2002. 59 O.R. 3rd 773, 214 D.L.R. 4th 163 Div.Ct. freedom of information - elections - public interest, leave to appeal abandoned, [2002] O.J. No. 3309 C.A.

Haggerty, Kevin. 2004. "Ethics Creep: Governing Social Science Research in the Name of Ethics." Qualitative Sociology, 27(4):

Hamilton Spectator, Toronto Star, Kitchener-Waterloo Record 2006 “Collision Course.” TheSpec.com – June 3, retrieved from http://thespec.com/specialsections/section/CollisionCourse

Mann, Bruce. 1986. “The Federal Information Coordinator as Meat in the Sandwich.” Canadian Public Administration, 29(4): 579-582.

Ontario Human Rights Commissioner. 2003. “Paying the Price: The Human Cost of Racial Profiling.” October 21, Toronto.

Rankin, Jim, S. Simmie, M. Shephard, J. Duncanson, and J. Quinn. 2002. “Race & Crime.” Toronto Star – October 19, 20, A1. October 26, 27, A1.

Rankin, Jim; A. Bailey, D. Bruser, M. Welsh, B. Popplewell, M. Henry, and D. Brazao. 2010. “Race Matters.” Toronto Star – February 6, 7, 8, A1

Reid, Daniel. 2009. “Print journalism’s days numbered: former Toronto Star publisher.” now.carleton.ca – February 8, retrieved from http://www.now.carleton.ca/2009-02/2209.htm

Reid, John. 2002. “The Information Commissioner’s Response to the Report of the Access to Information Review Task Force.” Ottawa.

Roberts, Alasdair. 2006. Blacked Out: Government Secrecy in the Information Age Cambridge: Cambridge University Press.

Royal Canadian Mounted Police. 2005. RCMP Correspondence to Jim Rankin Regarding Release of Criminal Record Data Signed in Ottawa, December 1, 2005, RCMP 03-ATIP-21960

Royal Canadian Mounted Police. 2003. RCMP Correspondence to Jim Rankin Regarding Programming Fee Estimate Signed in Ottawa, July 8, 2003, RCMP 03-ATIP-21960.

Toronto Police Services Board v. Ontario Information and Privacy Commissioner. 2009. ONCA 2.

Toronto Police Services Board v. Ontario Information And Privacy Commissioner. 2007. ON S.C.D.C.

Toronto Star. 2009. “Toronto’s Provincial Inmates.” TheStar.com – July 18, retrieved from http://www3.thestar.com/static/Flash/inmates/index.html

Toronto Star. 2008. “Crime & Punishment.” TheStar.com – July 19-27, retrieved from http://www.thestar.com/crimepunish

Vancouver Sun. 2009a. “Parking Secrets.” VancouverSun.com – December 14, retrieved from http://www.vancouversun.com/news/parking/index.html

Vancouver Sun. 2009b. “Public Sector Salaries.” VancouverSun.com – December 19, retrieved from http://www.vancouversun.com/news/public-sector-salaries/index.html

Wortley, S. and J. Tanner. 2003. “Data, Denials, and Confusion: The Racial Profiling Debate in Toronto.” Canadian Journal of Criminology and Criminal Justice, 45(3): 367-390.