



**Software Freedom
Law Center**

1995 Broadway, 17th Floor
New York, NY 10023-5882
tel +1-212-580-0800
fax +1-212-580-0898
www.softwarefreedom.org

**Testimony of James Vasile
before the
New York City Council
Committee on Technology in Government
in re Int. No. 991**

June 29, 2009

1 Introduction

Good afternoon, Chairperson Brewer, and thank you for this opportunity to testify before your committee today.

My name is James Vasile, and I am Legal Counsel at the Software Freedom Law Center. The Center, based here in New York, is a non-partisan, non-profit organization whose mission is to provide pro bono legal representation to protect and advance Free and Open Source Software.

This bill makes government data available to New Yorkers in two different ways. The first is a web portal, a central place for finding government records online. The second is with raw data files available for download. While the portal will be useful for simple searches, it is the raw data that holds the greatest promise of improving the lives of many New Yorkers.

In my remarks today, I'm first going to describe some of the ways this data can be assembled and turned into useful information and services for the people of New York. Second, I'm going to discuss who will do that assembly and why they'll do it at no cost to the city.

2 From Data to Information Services

This bill proposes to give New Yorkers a wealth of data that currently sits unaccessed in city files or is only provided in formats (usually PDF) that make automated processing impossible. Those records are a mountain of data, an almanac of the city that would be immensely useful if only people could get to it.

Data, and the search for isolated facts and records, though, are only the beginning. The real promise of this mountain of data is in mining for gold. §23-302 provides that "All public records shall . . . be made available in their raw or unprocessed form." This one line opens an entire future of possibility that can be had at zero

cost to the city because the exciting thing about this bill is not the city giving its people data. It is about what we, the people, are going to *do* with that data.

People are going to take the undifferentiated stream of isolated data points and aggregate it. They will download every single item of data you provide. Then they will cross-reference, map, and annotate it with semantic markup. They'll add metadata from other sources, combine it with online maps and make it searchable using tools even more sophisticated than Google's. In other words, they're going to take your data and spin it into useful information.

Services that combine data from Consumer Affairs, Police, the Health Department, Sanitation and non-governmental databases will enable people to answer questions like:

- “Where is the nearest place I can park without having to wake up for alternate side street cleaning tomorrow morning?”
- “Which restaurants in my neighborhood have clean health records and great reviews?”
- “Where should I put my electronics store so it will have the most foot traffic and the least competition?”
- “What is the average price of apartments similar to mine in a 5-block radius?”

This is not just information about government. It is information about New York. The interesting thing about a sidewalk-cafe permit is not that it was issued. What is interesting is that a restaurant is opening or expanding at that location. Someday soon, every time a new restaurant opens, all the locals will know because they signed up for neighborhood alerts at their block association website. That is how city data becomes useful information and, eventually, *news* that is relevant to the daily lives of New Yorkers.

3 New York vs. New Yorkers

Whether the city provides raw data feeds or just web portals makes a huge difference in the scope and quality of the resulting services. I mentioned above all the great things that can be done with raw data, and the current administration plans to implement some of those services. Unfortunately, the city alone cannot do it all, and it is only by providing raw data feeds that the rest of New York can pitch in to create targeted services that the city is unable to provide.

Private competition (both from business and not-for profits) has beaten the city's online information services time and time again. A perfect example of the government's inability to compete with an entire city's worth of information providers is HopStop¹ which serves the same purpose as MTA's Trip Planner². MTA's service will plot a subway and bus route between any two addresses in New York. HopStop does not restrict itself to city data and includes destinations reachable by Ferry, Amtrak, LIRR and MetroNorth. MTA's Trip Planner is not likely to ever include such information.

Earlier, I mentioned custom maps of electronics stores and foot traffic data. Such data is of rather esoteric utility, but is of immense value to the narrow segment of the population that can use it. Such microuses of data abound:

- Companies with large delivery operations need software that can allot packages among a fleet of vehicles, schedule the deliveries and plot best routes, taking into account traffic, construction and holidays. They need this functionality integrated with their normal software, not on a website.

¹<http://www.hopstop.com/?city=newyork>

²<http://tripplanner.mta.info>

- An event planner might combine data from Certificates of Occupancy and Catering Establishment licenses with subway maps to find a venue suitable for a large event.
- For marketing purposes, an elevator repair business would want daily reports on safety violations in elevators.
- Hospitals and medical offices could use injury and illness data to help understand which communities are underserved by medical treatment facilities.

The city will never be able to produce services that answer all of these questions. Fortunately, if it provides raw data, it does not have to.

Simply put, the city cannot predict and then fill the needs of all New Yorkers. Many of those needs are not yet known, are not fully articulated, or are too narrow to be addressable by city resources. The only way to serve those needs is to give New Yorkers the raw data that enables them to help themselves.

4 Zero Cost Services

I mentioned earlier that if you provide the data we will see these great services arise at zero cost to the city and the New Yorkers accessing these services. Now I'd like to describe who will make these services and why they will do so.

4.1 Volunteer Librarians

New York is full of people who care about information. They collect it, organize it, and share it with others. There are 73,000 pages on the Wikipedia website³ providing information about New York City. Every single one was created and improved by volunteers. That's a lot of people producing a lot of information. The question is what happens when you give people like that access to government data. The answer is they quickly build free services we did not know we needed but now suddenly cannot live without.

There is a searchable database of SEC filings available online. It's a free service run by the SEC called "EDGAR". When this service went online in 1994, nobody knew how badly we needed it. Today, EDGAR is vital to the daily business of New York's finance and legal industries. What a lot of people forget about EDGAR is that it was started by a private citizen. He ran EDGAR at his own expense for two full years, simply because he wanted to make this information available to others who might find it useful.⁴

I have a colleague who had trouble finding schedules for bus service to his Prospect Heights neighborhood. He tried to write a computer program to collect bus schedule information from MTA's website and deliver that information, on demand, to people's cell phones. He was stymied by the data not being released in a standard format that permitted automatic processing. Instead, all he could find was a PDF. If the data had been available, he would have made sure that any New Yorker could have used her cell phone to find an up-to-date schedule for any bus line in the city.

Two friends who met while studying Urban Planning at New York University have created a website, Ride-TheCity.com. Users can enter two addresses, and the website will suggest the safest route for a bicyclist traveling between the two points. The site accomplishes this simply by favoring streets with bicycle paths. It would be much more effective if it took street-by-street bicycle accident statistics into account.

³See <http://en.wikipedia.org/w/index.php?ns0=1&search=new+york+city> &title=Special:Search&fulltext=Advanced+search.

⁴It should be noted that when the SEC took control of the project, they received a world-class website with a track record of proven success and two years of free development, testing and operation. The website is still in operation today.

These three examples show the promise of open government data. If you make it available in raw form suitable for automated processing, you've completed the task because the public will take it from there. There is ample evidence that once people get this kind of data, they immediately improve it and share it with everybody else. They build new tools to share information and then they share the tools too.⁵

4.2 Digital Civics Groups

Anticipating the trend in open government, teams of technologists have already started forming non-profit organizations to build portable tools that vacuum up the information and make it programmatically searchable in sophisticated ways.

These digital civics groups are building tools to collect, analyze and distribute government data.⁶ Their pitch to any government official who will listen is simple: give us your data and we will spin it into information. These volunteer, grant and donation-funded organizations are scanning the city's paper documents, scraping data from public web sites, and building software tools to effectively search and analyze what they find. All they need are the public records and any other metadata the city can add. They will do the rest.

The people that volunteer their time for these efforts have some core values in common. They believe information is power that should be distributed widely and freely throughout society. They see the enormous opportunities offered by access to government data, and they fear the prospect of that data being locked up with exclusive city contracts and then sold back to the New Yorkers who owned it in the first place.⁷ These people fight proprietization of public data by giving it away before it can be sold.

5 Conclusion

New York City is big and complex. Our lives are intertwined in great and awesome ways. The only entity large enough to have a perspective on the whole city is our government. New Yorkers know this. They are hungry for government data and information about government services.⁸ They want to share in the power of the government's ability to aggregate an entire city's worth of data. This bill gives that power to every New Yorker with a computer and, as computers move off our laps and into our pockets, to every New Yorker with a cell phone.

⁵Many of the same people who care about open government are members of the Free Software movement. They will release their software under Free Software licenses that *further* enable other citizens to provide yet more services using public data.

⁶See, e.g., <http://civx.us>.

⁷For example, a company in San Francisco is currently asserting an exclusive right to publish Muni train arrival times.

⁸Half of all calls to 311 (nine million calls) result in an immediate answer to a question by operators who search a database of 6000 pieces of information about the city. See http://www.nyc.gov/html/ops/html/311/311_vol_perf_levels_mar_09.shtml and http://www.911dispatch.com/info/new_york311.html.