

latimes.com/news/local/la-me-geocoding-errors5-2009apr05,0,5966285.story

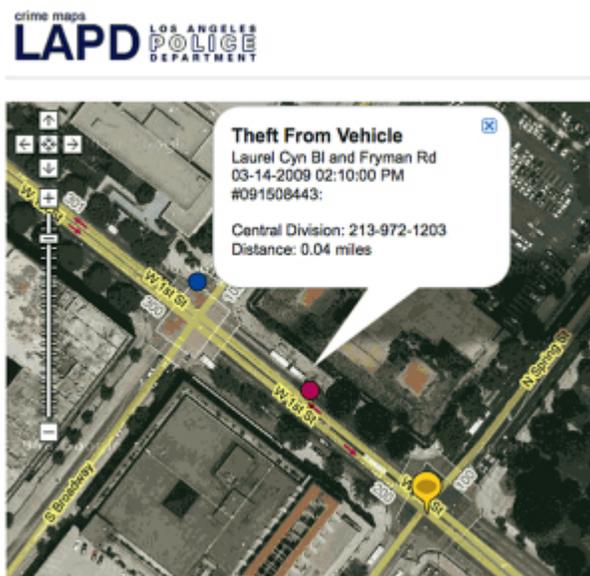
latimes.com

Highest crime rate in L.A.? No, just an LAPD map glitch

LAPD's online map showed big trouble downtown at a single point: the software's default address.

By Ben Welsh and Doug Smith

April 5, 2009



On Monday it was a grand theft auto and two robberies, on Tuesday two more robberies and four aggravated assaults. By Friday the toll had risen to 39 major crimes.

And, according to the Los Angeles Police Department's website, that week late last month was pretty typical of the mayhem around the corner from City Hall.

Since the inception of the LAPD's online crime map three years ago, the 200 block of West 1st Street has consistently shown up as the most likely place in Los Angeles to be victimized by crime.

But don't believe everything you read on the Internet. The spot, directly in front of the Los Angeles Times and a block from the new LAPD headquarters, is actually quite lawful.

Behind the apparent enigma is a case of virtual unreality. The crimes reported there were real, but they actually happened somewhere else. The only thing they had in common was an address that proved impossible for a computer to find.

The distortion -- which the LAPD was not aware of until alerted by The Times -- illustrates pitfalls in the growing number of products that depend on a computer process known as geocoding to convert written addresses into points on electronic maps.

In this instance, www.lapdcrimemaps.org is offered to the public as a way to track crimes near specific addresses in the city of Los Angeles. Most of the time that process worked fine. But when it failed, crimes were often shown miles from where they actually occurred.

Unable to parse the intersection of Paloma Street and Adams Boulevard, for instance, the computer used a default point for Los Angeles, roughly 1st and Spring streets.

Mistakes could have the effect of masking real crime spikes as well as creating false ones.

In the six months from last October through March, the LAPD placed 1,380 crimes -- 4% of all crimes mapped -- at the Civic Center point, a rate of nearly eight a day, a Times analysis found. The Times discovered the problems while developing its own crime website that will feature the LAPD data. After finding the Civic Center error and others, The Times is developing a strategy to geocode addresses with a higher degree of accuracy.

In the LAPD map, many of the crimes placed downtown actually occurred in the San Fernando Valley, the Westside or South L.A. Sometimes, L.A. crimes were placed outside the city, as far away as Lancaster and Catalina Island. In hundreds of cases, crimes that took place in South Los Angeles migrated north dozens of blocks.

Alerted to the findings, Lightray Productions, the contractor that designed the LAPD site at a cost of at least \$362,000, has promised to fix the problems.

As a first step, last week Lightray's geocoding subcontractor, PSOMAS, stopped plotting both past and future crimes to the default point at Civic Center. Instead, the system now codes those crimes with a latitude and longitude of zero, a point that actually exists about 7,800 miles away, off the coast of West Africa.

A spokesman for the LAPD said the department will add a disclaimer to its site once it's cleared by the city attorney.

Though it doesn't use the crime map to deploy police resources, the LAPD considers it a crime-fighting tool. Community-based officers encourage residents to search it for crime patterns, said Sgt. Frank Preciado, who oversees the LAPD's online unit.

When the LAPD launched the mapping site in March 2006, it was portrayed as a publicly accessible version of Chief William J. Bratton's vaunted CompStat system. CompStat is a computer-powered tracking process first developed under Bratton at the New York Police Department -- and since featured in police dramas such as HBO's "The Wire." An LAPD official told The Times that CompStat uses a different procedure to map crimes internally and does not provide the latitudes and longitudes used to create the public map.

Preciado said the LAPD website, on which the crime map is one of the most popular features, brings in 4 million to 7 million page views a month. Preciado said he had an inkling of a problem when a Studio City resident complained that the map didn't show a crime in which he was a victim. But Preciado said he assumed that the map point had merely expired.

One reason the errors were not caught earlier may be that the LAPD site retains crimes for only six months and allows viewers to see only a seven-day period at a time. The presentation makes some trends, such as the large accumulation of crimes mapped at Civic Center, more difficult to spot.

The mistakes spread on the Internet, often compounding the distortion.

In competition for Internet viewers, newspaper websites and online-only publications such as EveryBlock.com are using commercial geocoding to map large bodies of data including crime reports, traffic accidents, pothole locations, liquor licenses and, in some cases, the homes of registered sex offenders.

Using the LAPD data, EveryBlock has consistently ranked 90012, one of downtown's ZIP Codes, the most dangerous in the city, positioning a large, foreboding orange cluster over the Civic Center with the number of crimes regularly updated.

EveryBlock, an enterprise that specializes in pulling together local data from many sources, has been praised as a model for the future of journalism, but unlike traditional publications, the site takes no responsibility for the accuracy of its aggregated data.

"Any mistakes found on EveryBlock should be sent to sources that provide the original data," a disclaimer states.

EveryBlock founder Adrian Holovaty said that the LAPD data received a "human sniff test" before being published on his website but that it did not expose the errors. He noted that "every database is flawed." EveryBlock, he said, takes "a very experimental stance" in the developing field of online database publication, choosing to publish first and investigate problems once they are brought to the staff's attention.

"We have to assume at some fundamental level that the governments aren't feeding us data that is complete garbage," he said.

But the newfound ease provided by online services can create a false sense of confidence in the computer's matching ability. Behind the scenes, an algorithm tries to translate unruly street addresses, often drawn from handwritten forms, into the precision of decimal degrees.

"Most spatial data are inaccurate," said Paul Zandbergen, a professor of geography at the University of New Mexico who studies the quality of online maps. "It's much easier to go the path of 'let's build it and not worry about the quality.' And that's been the trend since we started doing mapping."

In some instances, mapping errors could affect policy decisions. Zandbergen, for example, studied the possible use of geocoding to create zones around schools and other areas where registered sex offenders would be legally barred from living. He concluded that the tool was unreliable.

Small changes in how an address is typed -- for instance, "68th St." instead of "W. 68th St." -- can put the point in the wrong neighborhood or even the wrong city. Adding or removing the name of the city or a ZIP Code can lead to differing results. Because of the way they break up the parts of an address, Google and Yahoo sometimes return different locations for addresses typed exactly the same way.

Los Angeles streets offer their own set of challenges: Valley streets may cause trouble if identified as being in Los Angeles, rather than with a postal address like Reseda or North Hollywood; identical addresses appear both downtown and in San Pedro; and the avenues of Northeast L.A. cause havoc in some systems.

Considering those roadblocks and the current state of geocoding technology, no mass-generated crime map will be 100% accurate. LAPD spokeswoman Mary Grady said the department will work with its contractor to make the map as accurate as current technology allows.

"We are offering a service to the community we hope is helpful," Grady said. "It's not perfect. We do the best we can with the software available."

Getting it right, in fact, is one of the central tenets of Bratton's system.

"The operative word in this process is accuracy and follows the garbage-in, garbage-out principle," Det. Jeff Godown, head of the LAPD's CompStat unit, wrote in January 2007. "In order to create the best crime reduction strategies, those strategies must be based on an accurate crime picture."

That may be a reasonable goal, but for mass geocoded data it's not today's reality, Zandbergen said.

"The field is happening, but it needs to mature."

Copyright © 2009, [The Los Angeles Times](#)