



Can data help break through Mexico's information blackout?

By: Joshua Keating – March 12, 2013

Last weekend, a newspaper in Mexico's northern Coahuila state became the latest outlet in the country to announce that it would no longer report on the country's drug violence after receiving death threats from the Zetas cartel. The decision by the editors of *Zocalo* follows a paper in Nuevo Laredo which made a similar decision after a series of grenade attacks last year.

A panel at the Cato Institute today, co-hosted by Google Ideas, addressed the question of how data and social media can be effectively deployed to gather information on the drug war, at a time when traditional outlets of information are increasingly self-censoring. As Google Ideas Director Jared Cohen explained in introducing the panel, Mexico is a rare case of a democratic country that has effective media censorship, not due to laws but due to threats of violence posed by non-state actors.

With reporters often unable to file on violent incidents, social media is now frequently filling the gaps explains *New York Times* reporter Karla Zabłudovsky, who is based in Mexico city. "This is becoming the norm in many parts of Mexico," said. "Unconfirmed information from Twitter and blogs are increasingly replacing traditional media in the critical first hours [after an incident] and sometimes well beyond that."

But Zabłudovsky argues this reliance on social media comes with a price. "Anonymity means it is nearly impossible to verify who the sources are, if they are indeed at the scene they are reporting from, and what their agenda is," she says.

Mexico is certainly a vibrant social media environment. Internet usage has doubled in the last 10 years to at least 36 percent of the population -- likely higher. 60 percent of internet users are active on social media and 20 percent of those are on Twitter, making Mexico the fifth largest Twitter community in the world.

As Microsoft researcher Andrés Monroy-Hernández explained, drug violence is one of the most topics on Mexican twitter, with terms like *balacera* -- shoot-out -- among

the most frequently used. Users can track violent incidents near their houses -- often referred to euphemistically as "risky situations" using organically-developed city-specific hashtags such #mtyfollow for the city of Monterrey.

Monroy-Hernandez also compared Mexican Twitter use to the United States, noting that Mexicans are more active retweeters, and tend to use the site more to disseminate information rather than interact with each other. The information broadcasted over Twitter often contrasts sharply with what's available through official channels, he argued:

[In the city of Nuevo Laredo] the mayor mysteriously disappears for days and refuses to discuss drug violence, the military general who presides over the city doesn't hold any news conferences or issue any statements to the media. So you can imagine if you are a citizen in this kind of place, you are seeing all sorts of violence and you turn on the TV and all you see is soap operas.

Javier Osorio, a PhD. candidate at Notre Dame, still sees utility in traditional media sources but takes a Big Data approach to them. Osorio has developed linguistic analysis software -- the first of its kind for Spanish, he says -- which he used to comb through media, police, and government reports to build a geo-referenced database of drug violence which notes who is committing the violence, who is the victim, and the nature of the event. (This contrasts with the official data released by the Mexican governments which tends to just list numbers of deaths without much context.)

Osorio's data, for instance, shows that the vast majority of drug-related fatalities are cartel-on-cartel violence, without involvement of law enforcement.

I was curious, however, about how reliable this data would continue to be if drug violence is increasingly going unreported. "Sometimes when there is violence against journalists in an area, they still report—they just send it to the national newspaper and report it as "staff." Since I have so many sources I can cross-validate," Osorio replied.

Monroy-Hernandez suggested that Osorio's dataset could be improved with the inclusion of information from social media:

They are quoting directly from social media without the name of the journalists involved. You'll see articles like "Last night, on Twitter, people were reporting XYZ happening." There's a lot of information showing up on blogs and social media that doesn't show up in traditional media and it would be really interesting to augment some of the data with data from blogs.

Of course, with Mexican bloggers increasingly targeted by cartel violence as well, and Twitter users facing heavy-handed legal penalties, even social media comes with risks.