Weaving Together the Strands of Big Data Policy and Practice in Local Government

By Cheryl A. Leanza* and Joseph Van Eaton**

For more than two decades, beginning with the emergence in the mid-1990s of CompStat in the New York City Police Department,[1] local governments have been using data, combined with accountability and follow-through, to improve public services, leading to improvements in the quality of life for their residents. But in the last few years the acceleration toward more widespread use of complex data analytic techniques has been dizzying, with new programs and futuristic analyses seemingly appearing daily around the country. From local government experiments to university-based centers[2] and book-length treatments,[3] public entities everywhere are thinking about how to better obtain and use data. The White House undertook a 90-day review of big data in 2014 underscoring that "the power of big data ... will be equally transformational for states and municipalities,"[4] pointing to New York City's Office of Data Analytics and Chicago's SmartData project[5] as examples of some of the most innovative uses of big data to improve service delivery.

Actors in the corporate and public interest sectors are urging government to implement systems that allow local governments to take advantage of big data analytics. Private sector firms seek to become both technology contractors to local government and also potential consumers of the newly available data. Public interest advocates, such as those supporting better services and transparent government have different, often laudable, goals. While the array of potential applications is vast, the approach to utilizing these techniques has not always been systematic. In particular, those advocating for big data's use in one context are not always aware of countervailing considerations in another context, and often are not accustomed to considering these issues from the perspective of local governments.

> Locally elected officials are eager ... to reap the advantages [of] new technological tools.

As counsel to governments recognize, local governments face particular legal liabilities and constitutional constraints, and also possess their own policy agendas which are often quite different from the private sector or others who use and analyze this technology from different perspectives. Local elected officials are eager, like the private sector, to reap the advantages and efficiencies that can be generated from new technological tools. But at this time, some of the most important questions about this technology's use in the public sector has not even been identified, let alone answered. For local elected officials to fully take advantage of the opportunities offered by big data while avoiding the serious complications big data raises, they must develop a more systematic approach of best practices that take into consideration the unique legal obligations of local governments.

WHAT IS BIG DATA?

Statistics and the analysis of data are not new, but "big data," as a term, is subject to a variety of definitions. At its root, however, big data means bringing together a huge volume of digital data available from many disparate sources and using vast computational capacity to identify trends or anomalies. No longer are rationally-organized data or particularly structured queries necessary to produce actionable information. For example, many cities are experimenting with big data analytics in predictive policing tools to identify hot spots or likely crime victims and perpetrators.[6] In another example, the city of Austin, Texas combined local school student health data on "BMI [body mass index] and cardiovascular fitness scores that

are geo-tagged with social and economic information" to produce neighborhood maps of wellness to fight childhood obesity.[7]

PUBLIC DATA, OPEN SOURCE

Much local government data is publicly available, and the trend has been to increase access to this data—particularly promoted by the movement supporting "open data" or "open source data."[8] Publicly-available data can be beneficial for a variety of reasons. First, as a democratic principle, city leaders may believe that a city that is transparent can be more accountable and responsive to its citizens. And in California, open government laws require that public data be generally available to the public, and cities find that digital release of data can be an easy and productive way to make information available in accordance with the law.

Many cities have released data, using principles of "open source."

Many cities have released data, using principles of "open source," so that private citizens and private companies can help make information about city services available to their residents. Bloomberg Philanthropies launched "What Works Cities" in April 2015, aiming to train cities that apply in the best data techniques to "produce results for their residents."[9] Cities around the country have established open data portals to make their data available via interactive easy-to-use tools.[10] For example, complex data sets can enable private developers to make arrival times and bus routes easily available to riders, such as the newly-announced Santa Monica open source trip planning app that can be used by transit riders and cyclists.[11] California's SB272 requires local agencies to inventory and publish the list of IT platforms they use. But open source data is not a panacea. Easy access to data and computer analytics can allow unintended and

unforeseen invasions of privacy. For example, after the tragic elementary school shooting in Newtown, CT a local newspaper used publicly available data to develop an interactive map of containing the addresses of everyone with a firearm permit.[12] Moreover, sometimes data that has always been public, but was not practically available in the past, can become uncomfortably public once it is digitized in a government data set—such as campaign finance records being used to create an online interactive map of California's Proposition 8 financial supporters.[13]

BUT IS IT ANONYMOUS?

At the same time that cities are rushing toward ever more transparent and open data, big data analytics are making de-anonymization of data ever more feasible, indeed inevitable. The research of Dr. LaTonya Sweeney, Director and Founder of Harvard University's Data Privacy Lab, and former Chief Technology Officer to the U.S. Federal Trade Commission, has demonstrated how easily supposedly anonymous data can be quickly traced back to individuals. As early as the mid-1990s, Dr. Sweeney used public, supposedly anonymous data, to isolate and identify the health records of the then-Governor of Massachusetts, including his diagnoses and prescriptions.[14] In a much more recent example, New York City officials inadvertently revealed detailed individual trip information from over 173 million taxi trips, because their attempts to anonymize the data were inadequate and could be easily reverse engineered.[15] Although the federal Privacy Act prohibits federal agencies from disclosing identifiable data, it is unclear in the era of big data what kind of protections are sufficient.[16] One scholar has pointed out that traditionally privacy law protects personally identifiable information" (PII), but "the trouble is … PII is an ever-expanding category" because in the past some information would never have been considered private, but with new re-identification techniques, seemingly innocuous data can be quickly connected down to surprising detail.[17] Localities subject to state and local privacy obligations must identify the ways in which open data policies—or responses to requests for records—could inadvertently violate privacy protections, and determine what their obligations are if data sets, considered individually, do not violate privacy obligations, but data sets, analyzed collectively, would reveal that information.[18]

RESPONSIVE TO WHOM?

Many municipal big data efforts focus on obtaining citizen input to improve city services or respond to needs. Citizens can report potholes, inferior city services, or in Los Angeles, participate in an effort to identify important historical landmarks through online surveys and submission of historical resources.[19] These efforts, however, only sometimes take into account the underlying disparities in technology access. Low income people have less access to technology and often these digital divides are even worse for historically disadvantaged groups, such as African Americans or Spanish-speaking households. For example, these communities subscribe at lower rates to broadband services,[20] or may not have the time or capacity to report data via city portals. Thus, a data-driven initiative that depends on citizens to interact digitally will automatically miss many people of color or low-income families.

If city resources are allocated according to these data, this allocation could intensify disparities instead of closing them. Thus, the city of Boston uses a smartphone app to help detect potholes on its streets, but must work with academics to supplement smartphone data with projections that are designed to remedy the data disparity.[21] Similarly, one writer describes the mapping of social media data after Superstorm Sandy: most activity originated in Manhattan, which was less profoundly impacted by the storm than the lower-income lower-density neighborhoods such as Breezy Point and Rockaway that produced less social media output.[22] While the "discourse of open data is often infused with the idea that data is objective," in fact data is as much the product of the society that creates it as the people who analyze and interpret it.[23] A locality that wishes to rely on data to target problems will need to consider what steps it should take to ensure that a "data divide" does not translate into an infrastructure or resource divide, and consider whether policies that target resources based on social interactions create liabilities under applicable law.

DATA & CIVIL RIGHTS—AN OBLIGATION TO ANALYZE?

The White House Big Data Report recommended that lead civil rights and consumer protection agencies expand their technical expertise so that they can identify big data analytic practices with discriminatory impacts on protected classes, and that they use that new expertise to investigate and resolve violations of law.[24] This tool could be powerful as well in the hands of city enforcement agencies—from identifying code violations to promoting public safety.

> [Public entities] should proactively analyze ... data to ensure its algorithms are not inadvertently biased.

Another idea gaining traction is that entities, such as corporations, that use big data techniques should proactively analyze that data to ensure its algorithms are not inadvertently biased against historically disadvantaged groups or violate other company principles.[25] In the area of constitutional law, such a concept might take on particular significance in cities. If a city has the big data available to analyze its transportation network to improve traffic flow and retime traffic lights, what are its liabilities if analysis of this public data also reveals that it is systematically underserving its Hispanic or Black residents? What is the legal obligation of the city? Such an analysis might be even more critical if—as in the case of at least one city—the city has turned over public functions to private sector companies using big data. If Altamonte Springs, Florida subsidizes Uber rides instead of purchasing city buses,[26] what is the city's obligation if Uber is demonstrably underserving that city's African American or Hispanic residents?[27]

The potential impact of new technologies on civil rights—and the data generated through use of new technologies—is being examined in particular cases, most notably with respect to use of body cameras. Like all use of data, the technology can either advance or impede civil rights, depending on how they are implemented. Several civil rights organizations have proposed best practices for development of policies regarding public safety body cameras and the use of data generated through those devices.[28] The increased use of intelligent traffic systems and electronic payment systems for public transit (to the exclusion of cash payments) may raise equally troubling privacy and civil rights issues that need to be confronted proactively.

WEAVING THE STRANDS

The various streams of big data use in U.S. cities is only increasing, but, as-yet, cities have largely been able to focus on pieces of this landscape; few if any have integrated the disparate elements of law and policy that bear on new data uses. A sound political directive to maximize open source data could well conflict with privacy law. An innovative use of big data to identify better ways to serve citizens could exacerbate existing class

inequalities. Data that can be used proactively to protect historically disadvantaged groups could also be used to pinpoint when the city itself may be violating constitutional rights. The legal and policy implications of these programs are just beginning to be understood. Cities require a more systematic approach to these issues, complete with the hard work of identifying principles, best practices and legal analysis. California cities—with their combination of strong support for open government, deep belief in the promise of technology, commitment to the protection of civil rights and civil liberties—should start getting ahead of the curve on a comprehensive approach.

*Cheryl A. Leanza, Esq. is the President of her consulting firm, A Learned Hand, LLC, and has a 20-year career working on cutting-edge telecommunications and media policy advocacy that benefits the public interest.

**Joseph Van Eaton is a partner in Best Best & Krieger LLP's Municipal Law practice group in the firm's Washington, D.C. office and represents government agencies on a broad range of communications issues in federal and state courts, before federal and state agencies, and at the negotiating table.

ENDNOTES

1 Bureau of Justice Assistance, Police Executive Research Forum, *CompStat: Its Origins, Evolution, and Future in Law Enforcement Agencies* (2013) available at: http://www.policeforum.org/assets/docs/Free_Online_Documents/Compstat/compstat%20-%20its%20origins%20evolution%20and%20future%20in%20law%20enforcement%20agencies%202013.pdf

2 See, for example, Ash Center for Democratic Governance and Innovation's Data-Smart City Solutions Project at Harvard's Kennedy School, http://datasmart.ash.harvard.edu/

3 Anthony Townsend, Smart Cities; Susan Crawford and Steven Goldsmith, Responsive Cities (2014).

4 *Government Technology*, 7 Things to Know about the White House Big Data Report (May 2014), available at: http://www.govtech.com/data/7-Things-to-Know-About-the-White-House-Big-Data-Report.html

5 Note the City of Chicago project is a startup founded through a partnership among the city of Chicago, the John D. and Catherine T. MacArthur Foundation, and The Chicago Community Trust. See http://www.smartchicagocollaborative.org/.

6 John Eligon and Timothy Williams, Police Program Aims to Pinpoint Those Most Likely to Commit Crimes, New York Times (Sept. 24, 2015) available at: http://www.nytimes.com/2015/09/25/us/police-program-aims-to-pinpoint-those-most-likely-to-commit-crimes.html?_r=0; see also danah boyd, 33rd Annual Parker Lecture, (Oct. 20, 2015) available at: http://www.uccmediajustice.org/o/6587/p/salsa/web/blog/public/?blog_entry_KEY=7648 (describing ethical dilemma for use of big data information about individuals).

7 Stephen Goldsmith, "Austin Targets Youth Obesity With Neighborhood-Level Data," Government Technology Sept. 3, 2014) available at: http://www.govtech.com/data/Austin-Targets-Youth-Obesity-With-Neighborhood-Level-Data.html

8 "Open data" is information that is accessible to everyone, machine readable, offered online at zero cost, and has no limits on reuse and redistribution." U.S. DEP'T OF COMMERCE, FOSTERING INNOVATION, CREATING JOBS, DRIVING BETTER DECISIONS: THE VALUE OF GOVERNMENT DATA 28 (2014). A literature review prepared by the Berkeley Technology Law Journal on Open Data is a particularly valuable source on open data, its advantages and potential dangers. Emmie Tran and Ginny Scholtes, OPEN DATA LITERATURE REVIEW, (prepared for the 19th Annual Berkeley Center for Law & Technology/Berkeley Technology Law Journal Symposium: Open Data: Addressing Privacy, Security, and Civil Rights Challenges) available at https://www.law.berkeley.edu/wp-content/uploads/2015/04/Final_OpenDataLitReview_2015-04-14_1.1.pdf.

9 Tanvi Misra, "3 Cities Using Open Data in Creative Ways to Solve Problems," *CityLab from The Atlantic* (April 22, 2015) available at: http://www.citylab.com/cityfixer/2015/04/3-cities-using-open-data-in-creative-ways-to-solve-problems/391035/; see also About What Works Cities at http://whatworkscities.bloomberg.org/about/

10 See, e.g., Jason Hibbets, The Foundation for an Open Source City, (2013) (describing Open Raleigh's efforts).

11 Colin Wood, Jason Shueh, "Santa Clara, Calif.'s Multimodal Trip Planner Showcases the Power of the Cloud," Government Technology (April 12, 2016) available at

12 Open Data Research Network, Open Data & Privacy discussion notes; Newspaper and guns: If data is available should it always be published? Gigaom (Dec. 27, 2012).

13 Jesse McKinley, "Marriage Ban Donors Feel Exposed by List," New York Times (Jan. 18, 2009).

14 Nate Anderson, "Anonymized" data really isn't—and here's why not, Ars Technica (Sept. 8, 2009) available at: http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/

15 Dan Goodin, Poorly anonymized logs reveal NYC cab drivers' detailed whereabouts, ArsTEchnica (June 23, 2014) available at: http://arstechnica.