Canaries in the Coal Mine: Understanding and Acting on Data in Context

White Paper to Accompany the Keynote Presentation at the 2014 Lackawanna Luzerne Indicators Forum, Wilkes-Barre, Pennsylvania

Henry J. Amoroso
Associate Professor
Department of Economics and Legal Studies
Seton Hall University

As collecting, compiling, and cataloging large amounts of data grows ever easier with each passing year and technological advancement, governments have begun to slowly embrace a more data-oriented approach to their operations, budgeting, planning, and overall management strategy. Local governments, in particular, see data as a way to make the very visible delivery of services that residents and non-residents use every day more effective and efficient. The City of Buffalo, New York, for example, has begun to place GPS trackers on fleet vehicles to track mileage and cut down on fuel costs;[i] the City of San Francisco uses analytical tools to adjust metered parking prices according to variable rates of demand.[ii]

This revolution in governing comes at a vital time. Governments across the country, and particularly in the older cities of the Northeast and Midwest, face increasing financial headwinds, which have only been exacerbated by the collapse of the housing bubble and the consequences of the Great Recession.[iii] Pension costs continue to climb as more and more public employees enter retirement and begin to collect the benefits they're owed;[iv] a similar pattern has emerged for health care, which faces the compounded difficulty of rapidly rising medical costs nationally.[v] Debt taken on in better times can pose an additional crippling liability today.[vi] In short, governments are asked to maintain a level of expected services with less. Often they are asked to do more. Making better use of data is no doubt an essential part of the solution.

But there is a significant difference between *having* more data and *using* more data to solve the problems governments face. I hold that there are five fundamental ideas that must be accounted for when thinking about data's role in government:

1. Data-oriented governing is not new. In fact, the most important data a government typically has its budget. While governments have been able to identify remarkable efficiencies through advanced data analytics, bringing effectiveness and efficiency to government remains a task heavily focused around concepts that an accountant in 1914 would understand. It is very important not to limit one's idea of data to digital platforms or other forms of information derived from the newest technology; governments must *genuinely* take advantage of what they already have.

It is useful to consider the metaphor of a map to help draw out this point. We often take maps for granted given the age of the technology; but, in reality, they hold an incredible amount of data. As a thought experiment, imagine trying to convey all of the information held in a map in written text. Even a map of a small town would take a library's worth of volumes to contain (the road curves 5 millimeters to the right at 41°52′55″N 087°37′40″W and 6 millimeters to the left at 41°52′55″N 087°37′41″W). Today, an ever-greater number of people use digital maps—e.g., Google Maps and GIS programs—which simply make the prior technology more efficient and,

hopefully, more effective. However, one must remember that the digital overlays that the map provides would be useless without the map itself.

Cities would benefit from approaching governance according to this essential framework and avoid seeing data and other technology as a panacea for the ills they face. There is no doubt that digital technology can add utility, but it must serve to augment foundational technologies—namely, budgets. If that foundation is not present, then the addition of technology is essentially superficial and often amounts to a greater victory for public relations than good governance.

Two examples of best practices in this regard are so-called "open" budgeting and dynamic budgeting. The first major open budget project was in Oakland, California, though the practice has spread rapidly to other cities such as Jersey City, New Jersey. In essence, the open budget allows anyone with access to the internet to click through the City's budget and see expense and revenue items visually expressed. Various mechanisms to encourage discussion are also embedded within the software in order to gauge which items of the budget community members and other stakeholders hold to be among the most important and most relevant for their lives.[vii] Dynamic budgeting allows City staff to monitor revenues and expenses in real time. The budget, in essence, is a road map; quite often, revenues come in more or less than anticipated and the same is often true of expenses. Of course, a surplus does not pose any troubling challenge to the City, but, if budgeted revenues are coming in under budget, or expenses are coming in over budget, knowing this information in real time can help the City adjust and prevent cash flow or other financial issues. There are a number of software packages available that lend this service, though, at this stage, all are offered through private firms.

2. Data do not and cannot do anything. Because data is inert and requires people to make it useful, it is important to consider when data is a *substitute* for management rather than a management *enhancement*. Fleet tracking programs, for example, are most effective in environments where there is a high rate of fraud and abuse—e.g., using vehicles for non-governmental activities or filing up a personal vehicle with publicly funded fuel.[viii] Thus, a fleet tracking program can be an expensive solution for what essentially is a management problem. Of course, even the most sophisticated data platforms remain open for sometimes surprisingly simple 'work-arounds' and return management considerations to the fore. Data only helps if people want to use it or if the City has the capacity to use it.

New York City, for example, introduced a data-driven pothole registry and reporting program early in the Bloomberg administration. The time between receipt and repair was carefully monitored and entered into a dynamic database that would adjust according to the status of the repair. 1.25 million potholes were filled between 2002 and 2007 and 99 percent were filled in 30 days or less (vs. 65% before the institution of the registry and 311 reporting).[ix] However, New York City has both staffing and funding in excess of all American cities and instituting a similar program in a small or mid-sized city could be ineffective or even counter productive. Technology of this sort is expensive and can take a considerable amount of staffing and financial resources to maintain. And, while the platform might make the process more efficient, it still relies on the City's ability to send people to fill the potholes and acquire the materials to do so.

One solution for smaller and mid-sized cities is to engage the population directly in some of the upkeep of government services. The City of Boston, for example, has over 13,000 fire hydrants. Snowy winters and generally older infrastructure make keeping fire hydrants accessible a

priority—and quite expensive. Through a digital mapping tool, City residents can choose to "Adopt-A-Hydrant," maintaining access to the hydrant and reporting to the City when more serious issues arise. This saves the city money and, more importantly, allows the Fire Department to spend more time on fighting fires and establishing fire prevention strategies.[x] In addition, often-thorny political issues, such as private ownership or outsourcing the management of public assets, can also be avoided, which makes execution easier. Of course, fire hydrants are just one example. In general, this principle could be extended to many aspects of City government—perhaps even potholes and other infrastructure.

3. All forms of data are always situated in a particular context. Although numbers "don't lie," their meaning must be interpreted in light of their position in time and space. The first major oil strike, for example, was in 1859. In 1879, experts predicted that world oil production had peaked. According to the data available at the time—which necessarily limited itself to known oil fields (Northwestern Pennsylvania) and known methods—this prediction was correct.[xi] Governments must, therefore, be careful to understand the limits of data and continually assess what phenomena might emerge to disrupt current models and modes of approaching problems.

Detroit is one the most notable example of failing to adjust to changing market trends and, instead, continuing practices that do not conform to the realities of a new context. When times were good, the City of Detroit tried to mirror the grandeur of the auto industry, such as by granting substantial pensions and benefits and embarking on large infrastructure projects. However, as the auto industry—and thus the City's financial base—declined, the City's liabilities remained in place and it did not begin to take corrective action at a necessary scale or in a timely manner.[xii]

The best way to understand context is to study it and to make those findings actionable. Bodies as diverse as the CIA to Royal Dutch Shell to the City of Austin, Texas have found enormous utility in developing advanced "scenarios," which carefully measure current trends and make various detailed projections regarding possible future possible outcomes. Of course, all—or nearly all—of the scenarios will not come to fruition; but by having a better understanding of what *might* occur, cities or other organizations that undertake this type of strategic planning will have a far greater capacity to act and to deal with difficulties as they arise. Shell has performed some scenario building for cities—and even Detroit in particular.[xiii] Austin, TX, for example, has an entire Department devoted to both planning and strategic planning initiatives and crafts some of the richer reports available among municipalities.[xiv]

4. Data do not exist unless we *know* data exists. Many municipalities are now required by law to post their municipal budgets and other important documents on their websites. But many of the links to budgets, while posted, are broken—i.e., they are not actually available for download by the public. This pattern repeats itself within municipal government where immense amounts of data are collected but left underused or unused. Data simply does not matter unless it is made widely available *across* government departments and, vitally, accessible and transparent.

Boston, Chicago, Los Angeles, New York City, Philadelphia, San Francisco and Seattle (in a formation called the G7 Initiative) have joined in an effort to share the data they collect and better their predictive capabilities by increasing the sample size and by generally sharing best

practices in this regard.[xv] Chicago, for example, has been able to accurately predict when a rodent outbreak is likely to occur using predictive modeling.[xvi]

But small and mid-sized cities that have excellent data programs—e.g., Dubuque, IA—have small sample sizes and limited financial capacity. Regional and national sharing platforms have been proposed to help cities like Dubuqe mirror the work of the G7 and bring these advancements to a broader cross section of the population.[xvii] In another effort to help combat this problem, the City of Boston has made its "Citizen Connect" app available to 54 other communities around Massachusetts, which allows citizens to report issues via their smartphone.[xviii]

Another, and rather simple, step is to require that government agencies share their data. Too often information, even if it is collected, is left siloed in a department and not made available either within or outside of the government entity in question. Mayor Rahm Emmanuel of Chicago, in an attempt to tackle this problem, issued an Executive Order "mandating that city agencies publish public data sets under their control as well as update them on a regular basis."[xix]

5. **Data must be taught to speak.** One of the most exciting things about 'big data' is that it is big—i.e., it is filled with information. But this information can often be buried in thousands and thousands of pages, lines of code, or Excel columns. In 2014, every day 2.5 *quintillion* (25,000,000,000,000,000,000) bytes of data were created, with 90% of the world's data, at that time, created since 2012 alone.[xx] Government related operations are not an insignificant portion of that amount. The British government's tax collection arm alone holds 80 times more data than the British Library, the largest library by catalog in the world (and the UK's population is about a fifth of America's).[xxi] Governments need to establish ways to extract valuable information from data by quickly defining key metrics and inputting the information into easily accessible and dynamic dashboards. That being said, a far larger problem is helping data speak for itself even when it comes from a more traditional source like a budget, departmental report, or a financial statement. Essentially, this means putting data in context and explaining why it matters.

Chicago's "SmartData Platform" and "Data Dictionary" offer ways to navigate and make use of truly large amounts of municipal data. The aggregate total of Chicago's data collection efforts is 7 million rows per day. But this data exists in dozens of separate databases across units of government, making it difficult for synergies to be identified (e.g., the Department of Public works could redirect other City vehicles, such as street sweepers and refuse vehicles, away from street repairs). The SmartData platform allows users to quickly query the database for vital information across City departments. Moreover, it also uses collected data to predict future needs and makes suggestions to users.[xxii] The Data Dictionary provides another invaluable tool for the non-computer scientist or trained data analyst to understand and make use of the complex data indicators that might be found on the SmartData platform. It further explains what data the City has, where it may be found, and what formats it is available in.[xxiii]

The most important use of data, in itself, is to provoke reactions that lead to actions. Commonly, because of the structure of government and the common distribution of data, only insiders are given a chance to react—and, often, they choose to not act. A business does not usually have this capacity. While a business may put off tough decisions when presented with challenging data, the pressure from the business' internal stakeholders is greater and the threat of failure more immediate, thus making action more likely. Governments, on the other hand, have a greater

capacity and incentive to brush aside key indicators and keep pushing the date of action further into the future—even when the "data on the wall" is clear to see.

That's where we see the role of "canaries" come in. A canary's job is to have a keen grasp of the vital indicators *in their context* and be able to relay these indicators to *all* stakeholders. They must interpret data and be able to explain their significance both inside and outside of the government apparatus—particularly to the public. Even if the dangers facing a government have been mounting for some time, action may still be limited or not taken at all. To combat this behavior, a canary *must* be equipped to inspire decision makers to move beyond reaction and urge them to take genuine action.

A canary may hold a formal position within government, be any other stakeholder, or simply a member of the community. Who they are is much less important than what they do and how they do it. In short, it is a role that grows ever more valuable as governments absorb more data and face greater challenges.

Former Speaker of the House Tip O'Neill liked to say "all politics is local." We believe governing is too. Without having a deep and qualitatively based understanding of the context of a community, even the most comprehensive data set will remain inert, impenetrable, and ineffective—and the consequences may be inescapable.[xxiv]

[i] NexTraq, "Fleet Tracking Benefits for the Government Sector: How GPS Fleet Tracking Can Help You Balance Your Budget," 19-Oct-2012; GovernmentFleet, "City of Buffalo Installs GPS Units on Snowplow Fleet," 01-Nov-2010. [Online]. Available: http://www.government-fleet.com/channel/gps-telematics/news/story/2010/11/city-of-buffalo-installs-gps-units-on-snowplow-fleet.aspx. [Accessed: 02-Jun-2014].

[ii] Jaime Ballesteros and Alex Demisch, "SF Park: Overview Presentation," 04-Mar-2014; San Francisco Municipal Transportation Authority, "About – SFpark," 19-Dec-2013.

[iii] Tom Ferrick Jr., "Recession Over, Cities Still Struggle," Pew Charitable Trusts Magazine, Spring-2014; Ann Owens and Robert J. Sampson, "Community Well-Being and the Great Recession," Stanford Center on Poverty and Inequality, Stanford, CA, May 2013; Robert I. Lerman and Sisi Zhang, "Coping with the Great Recession: Disparate Impacts on Economic Well-Being in Poor Neighborhoods," Urban Institute, 2012.

[iv] Thomas J. Healey, Carl Hess, and Kevin Nicholson, "The Public Pensions Crisis," The Milken Institute Review, Jan. 2013; Dean Baker, "The Origins and Severity of the Public Pension Crisis,"
Center for Economic and Policy Research, Feb. 2011.

[v] Government Finance Officers Association, "Containing Health Care Costs: Proven Strategies for Success in the Public Sector." 2011; United States General Accountability Office, "State and Local Retiree Health Benefits: Liabilities Are Largely Unfunded, but Some Governments Are Taking Action," Nov. 2009.

[vi] Thomas P. DiNapoli, "Debt Impact Study An Analysis of New York State's Debt Burden," Office of Budget and Policy Analysis, Albany, NY, Jan. 2013.

[vii] Adam Stiles, Andrea Moed, Dave Guarino, Jeff French, and Mohamed Gulaid, "Open Budget Oakland," 2014. [Online]. Available: http://openbudgetoakland.org/;Akisia Grigsby, Valerio Luccio,et al., "Open Budget Jersey City," 2014. [Online]. Available: http://openjerseycity.org/JerseyCityBudget/index.html?sq=JCBudget.

[viii] Cf. GovernmentFleet, "Man Sentenced in GSA Motor Pool Credit Card Fraud Scheme," 28-Sep-2011. [Online]. Available: http://www.government-fleet.com/channel/fuel-management/news/story/2011/09/man-sentenced-in-gsa-motor-pool-credit-card-fraud-scheme.aspx. [Accessed: 02-Jun-2014]; Nicole Perez, "Ex-county employee charged with fraud," 17-Mar-2014. [Online]. Available: http://www.abqjournal.com/369928/news/excounty-employee-charged-with-fraud.html. [Accessed: 02-Jun-2014]]Terry L. Jones, "Brusly police chief accused of theft," 24-Jan-2014. [Online]. Available: http://theadvocate.com/home/8180878-125/brusly-police-chief-arrested-accused. [Accessed: 02-Jun-2014].

[ix] Accenture, "Transforming Customer Services to Support High Performance in New York City Government," Accenture, 2013; Staff Reporter of the New York Sun, "Mayor Returns to the City's Streets, Literally," The New York Sun, 18-Dec-2007.

[x] Joel Mahoney, Karla Macedo, Talin Salway, and Erik Michaels Ober, "Adopt-a-Hydrant," 14-Feb-2014. [Online]. Available: http://codeforamerica.org/apps/adopt-a-hydrant/.

[xi] Ron Chernow. "Titan." 1998. Vintage Books

[xii] Nathan Bomey and John Gallagher, "How Detroit went broke: The answers may surprise you — and don't blame Coleman Young," Detroit Free Press. [Online]. Available:
http://www.freep.com/interactive/article/20130915/NEWS01/130801004/Detroit-Bankruptcy-history-1950-debt-pension-revenue. [Accessed: 16-Dec-2013]; Kevyn Orr, "City of Detroit: Proposal for Creditors," Office of the Emergency Manager, Detroit, MI, Jun. 2013.

[xiii] Shell Global, "New Lenses on Future Cities," Centre for Livable Cities, Singapore, Jun. 2014.

[xiv] City of Austin, "Imagine Austin Comprehensive Plan," Austin, TX, Jun. 2012.

[xv] Tod Newcombe, "Cities Ramp Up Data Projects," 09-May-2013. [Online]. Available: http://www.governing.com/columns/urban-notebook/col-cities-ramp-up-data-projects.html. [Accessed: 21-May-2014].

[xvi] Ash Center Mayors Challenge Research Team, "Chicago's SmartData Platform," Data-Smart City Solutions, 08-Jan-2014. [Online]. Available: http://datasmart.ash.harvard.edu/news/article/chicago-mayors-challenge-367. [Accessed: 23-May-2014].

[xvii] Tod Newcombe, "Is Big Data Just for Big Cities?," Jul-2013. [Online]. Available: http://www.governing.com/columns/urban-notebook/col-big-data-just-for-big-cities.html. [Accessed: 21-May-2014].

[xviii] Jane E. Fountain, "Connecting Technologies to Citizenship," presented at the Urban Forum 2013, University of Illinois - Chicago, 05-Dec-2013.

[xix] City of Chicago, "Mayor Emanuel Expands Open Data on City Portal with Executive Order," 10-Dec-2012.

[xx] IBM, "IBM - What is big data?," 15-May-2014. [Online]. Available: http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html. [Accessed: 03-Jun-2014].

[xxi] Chris Yiu, "The Big Data aOpportunity Making government faster, smarter and more personal," Policy Exchange, London, UK, 2012.

[xxii] *Op. Cit.,* Ash Center Mayors Challenge Research Team

[xxiii] Chapin Hall at the University of Chicago, "City of Chicago Data Dictionary," 02-Jul-2013. [Online]. Available: http://datadictionary.cityofchicago.org/.

[xxiv] Martin Tolchin, "Thomas P. O'Neill Jr., a Democratic Power in the House for Decades, Dies at 81," The New York Times, 07-Jan-1994.