Anticipatory Government: Integrating Big Data for Smaller Government

Michael E. Milakovich, Senior Professor
University of Miami, Coral Gables, Florida USA

Abstract: Translating the avalanche of newly-created data sources into actionable decision-making requires high-value analytical skills and forward-looking political leadership. Few public administrators, and ever fewer elected officials, have access to—or knowledge of—statistical skills and relevant case examples necessary to fully utilize information available from burgeoning mega-databases. This paper examines the anticipated merger of private sector methodology, public administrative expertise and political leadership needed to apply big datasets to public sector decision making processes. Case profiles are drawn from public and non-profit organizations in healthcare, information communication technologies, law enforcement, procurement, and virtual training, among others. Cases are selected for geographic as well as population diversity. Cooperation between appointed managers and elected politicians is essential to apply big datasets and improve decision making in public and non-profit organizations.

Challenges: Administrative and Political

All public enterprises face major fiscal, managerial, and technological challenges. Continuing budgetary pressures resulting from the mortgage meltdown and the lingering economic recession produce only gridlock and political discontent. Current political realities in the United States and Europe offer ample opportunities for public managers to find new ways to analyze metrics and maximize internal operational efficiencies. Appointed and elected officials are aware of the increasing volume of raw information from traditional sources as well as that generated by new social media outlets (aka ‘big data’). The U.S. federal government has devoted substantial resources to assess, research, and develop big data initiatives. Numerous ongoing projects have been initiated by U.S. federal agencies, including the Department of Defense, Intelligence Community, Homeland Security, Energy, Health and Human Services, NASA, and NSF—funding aimed at spurring scientific discovery and innovation. The White House announced funding for new research and development initiatives aimed at extracting and exploiting the “knowledge and insights from large and complex collections of digital data… to help solve some the Nation’s most pressing challenges.”

Translating the exponential growth of new data into actionable decision-making requires high-value analytical skills which exist only in few governments. Still fewer public officials have access to relevant case examples necessary to fully operationalize information available from these burgeoning mega-databases. Methodologists are practically non-existent. Moreover, public sector organizational structures, with their diverse and divisive political coalitions, further inhibit interoperability and cross-
functional exchange of information and communication technologies (ICTs).\(^3\) As the sheer volume of data increases, the capacity of organizations to analyze and utilize its contents declines.

Since the ability of effectively cope with big data lies at heart in any area of digital governance, it is difficult to make any further sustainable progress without better data management. Therefore, significant attention must be placed on data mining, data processing and data managing in order to better construct data-driven digital governance.

In addition, organizational and political barriers prevent broader utilization of big data for achieving digital governance. The digital revolution is not waiting for government reform: 90% of all data available for decision-making analysis has been created within just the past two years.\(^4\) New sources of data are emerging from digital sensors almost everywhere: biometrics, blog postings, e-file tax records, climate sensors, video images, “likes” posted on social media sites, digital pictures, transaction records of online purchases, cell phones and GPS signals--to name a few. McKinsey and Company has identified how big data creates value in the private sector.\(^5\) Analysis of big datasets makes information more transparent and usable at much higher frequencies; data can be created and stored in digital and therefore more accessible formats; sophisticated analysis can improve decision-making; and big data can be used to improve the development of the next generation of customer-focused networked public services.

Government can use the same analytical processes to integrate multiple sources, improve transactions and outcomes, reduce costs and enhance citizen satisfaction. However, the dearth of expertise in this area is becoming more apparent as an estimated 150,000 to 200,000 more workers with analytical expertise and 1.5 million more data-literate managers will be required in the future. This is creating a ‘knowledge gap’ between future public sector employment expectations and currently available personnel. Whether sufficient numbers will be retrained or newly hired, this dilemma presents unique personnel, managerial and political challenges for national as well as sub-national governments.

**Data Mining and Digital Governance**

Big data is more than just accessing streams of data faster from greater number of sources: it is coping with mega-bytes of information which cannot be analyzed with traditional processes or tools. Volume, variety, and velocity are three of its characteristics. The sheer volume of data being collected and stored grew by 62%
between 2008 and 2009 to nearly 800,000 petabytes (PB) and increased to another 1.2 million PB, or 1.2 zettabytes (ZB) in 2010; it is projected to compound exponentially to an estimated 35 ZB by 2020. The vast quantity of data presents new challenges for data centers trying to deal with its sources and variety. Variety characterizes all types of data—a fundamental shift in analysis requirements from traditional structured data to raw, semi-structured, and unstructured form as part of faster data-based management systems. Not only has the sheer volume and variety of data increased, but the velocity at which it is generated, its ‘shelf life’ or useful value requires immediate analysis.

Organizations of all types may be overwhelmed by vast volumes of different types of data. They need to build the capability for analyzing more and different kinds of data in order to take advantage of the big data opportunity. Moreover, the timeliness of processing unstructured sources is also an important factor. Coping with big data effectively requires performing data mining in real time rather than after it has been collected and stored.

There have been numerous data mining and information sharing applications in various segments of the public sector (e-government, e-healthcare, e-procurement, predicative criminology, cyber security, and virtual learning), but very few studies combine sophisticated methodological tools with substantive public policies. Profit-motivated private companies have had greater incentives to pioneer advanced analytical systems aimed at customer relationship management (CRM). Similarly, governments have begun to realize that services should be also oriented toward meeting citizens’ needs, rather than reinforcing existing administrative and bureaucratic hierarchies.6

**Decision Making and Policy Implementation** Data mining extracts valuable information from multiple sources. It is an important technical aspect of data interpretation----the collection, management and use of data. Its credibility depends on the quality of data collected as well as the integrity and transparency of methodology. Requirements and challenges of data mining include the ability to handle different types of data; graceful degeneration of data mining algorithms; valuable data mining results; representation of data mining requests; mining at different abstraction levels; mining information from different sources; and protection of privacy and data security.7

More and more organizations are recognizing data mining as an effective technique to significantly enhance decision-making and performance. For example, IBM developed its Hadoop-based platform (a distributed file system for storing and managing large data repositories) which is well-suited to deal with semi-structured and unstructured data, as well as situations in which a data discovery process is needed.
Data mining is and will continue to be one of the major competitive tools of any organization, including public agencies.

**Data Driven Digital Governance** Governments apply data mining technology in multiple arenas from education to tax collection in order to plan, implement and evaluate policies designed to meet the needs of particular subsets of the population. These may include disabled school children in need of special instruction or elderly citizens requiring more intensive health care. The quality of planning and decision-making in all these areas are dependent on the collection, analysis and interpretation of accurate objective data, making it possible to spot trends and optimize processes. Greater precision will occur if governments build strong capability to collect and analyze data from multiple and reliable sources. As a result, those problems easily solved can be dealt with quickly and those neglected in the past may be brought to the surface and exposed for scrutiny. Decision makers will be better equipped to deploy geographic distributions, find cause and effect relationships, and monitor trends over time. For example, in order to meet the requirements of the No Child Left Behind Act, many school districts in American states are using SPSS software to analyze students’ records and spot trends through data mining tools.

In Australia, Centrelink is one of the largest government systems to employ data mining. This benefit agency deals with more than 6 million claimants and carries out over 5 billion electronic transactions a year. The Job Seekers’ Classification Instrument is a predictive model used to evaluate benefit claimants and assess the risk that they will become unemployed long-term. The high-risk group is thus identified and given more help in job hunting. Centrelink also uses of data mining to identify fraudulent claimants and to identify those for whom further investigation is merited. The inspiration comes from insurance companies. Low-risk claims are paid quickly, and high-risk claims are investigated in detail further. Data-mining software is also used by Denmark’s National Board of Health, France’s benefits agency, the South African treasury and Belgium’s finance ministry for performance measurement and policy planning.\(^8\) As successful cases multiple and are available as benchmarks to other agencies, governments will become less reluctant to apply lessons learned to similar policy areas.

**Data Mining by Network** Government by network serves as an important policy implementation tool to link fragmented public agencies with private sector providers in national defense as well as other important policy areas. Through contracting and partnerships, governments create cross-functional networks with private sector businesses capable of utilizing modern data-mining techniques. The process of analyzing large databases aims at constructing information that is not immediately evident from raw data alone. This is an expensive and carefully guarded process, but
governments are finding important clues to support missions such as preventing terrorism. An early example of the use of data mining in prevent terrorism comes from the former West Germany in the 1970s. By mining private sector databases such as travel companies, utilities and even pension funds, the West German Federal Crime Office created prescient profiles of how terrorists would behave.

**Data Sharing Applications**

Efforts have been made to enhance communication and collaboration among governments at all levels. Single sources of data are no longer sufficient to cope with the increasingly complicated problems in many policy arenas. Linkages between different data sets are occurring will continue in the future. As such, new models and partnerships of multiple data pools are being established and traditional bureaucratic concepts of data ownership are being challenged. In the post 9/11/01 world, the sharing of electronic data among agencies has been improved in the name of homeland security. The U.S. Department of Homeland Security emphasizes interoperability, as well as technological tools to move data between municipal departments, communities and other agencies and entities, such as hospitals, blood banks, and human and animal shelters. With the extended use of ICTs, governments are able to communicate with others much faster and share data on a larger scale. In this section, data sharing applications are briefly discussed in the specific public administrative fields such as environmental protection, transportation, healthcare and public safety.

**Environmental Protection and Transportation** After 9/11/01, the U.S. Environment Protection Agency (EPA) developed a database of technologies in the fight against terrorism. Multiple agencies benefit from the new EPA database and information sharing capabilities. During the time the EPA centralized information into a database, governments from all the levels were building electronic databases and creating data-sharing initiatives to help respond to biological threats. The Customs Service issued a request for proposals from commercial, off-the-shelf databases or software housing biographical information about airline passengers, such as current addresses, mobile phone records, criminal history, social security and vehicle registration. The Transportation Department used existing technology to establish a network that links airline reservation systems to private and government databases to screen passengers. At the state level, the critical and relevant data collected through the intelligence network was filtered down to local police because of the pilot program allowed New York officials to exchange information electronically in real time. The Disaster Management interoperability Services (DMIS) offers federal, state and local emergency managers open online access to disaster-management information, planning and
response tools and also applies ICTs to enhance the ability of public, private and non-profit agencies to respond to emergencies.

Health Care Social and behavioral data which store patient histories and directly influence patient choices to participate at different levels of engagement in particular programs are often overlooked. Separating the data sources from the final product or services could threaten the quality of health care services. Because most health insurers and disease management organizations currently target and manage participation via dependence on data regarding the patients’ medical condition, utilizing big data effectively combines public data and medical histories to better determine treatment options. This connection can drive smarter predicative models that will more accurately assist doctors. Health care utilization databases have been expanded in recent years and are used in data sharing efforts nationwide supplement claims payment processes. Claims and discharge data are widely used due to the relative ease of collection and comparability across regional sites. These data lend themselves to linkages and are useful in evaluating patterns of care and documenting variations in practices as well as outcomes.

No health information initiative will survive without a market for information. Leading examples of models exemplifying progressive governance and ownership structures are Utah Health Information Network and the California Information Exchange (CIE). Utah Health Information Network exemplifies a shared model and implementing electronic data interchange among trading partners, competitors. After being established by California business, physicians, health plans, hospitals, and health care systems, CIE is developing consensus standards for basic database architectures. These and other organizations are charting new courses to compel re-engineering of local health information systems for the purposes of streamlining business and transaction costs. And at the same time, they help move public agencies further along the technical and information continuum.

Law Enforcement and Public Safety Among the past successes are 311/CzRM systems which promote changes within state and local government, potentially eliminating insular and unresponsive departmental ‘silos’. New York City has transformed early versions of CompStat and CitiStat from reactive systems responding to problems as they occur, to proactive systems which enable officials to integrate databases, discover and address problems before they happen. Residents may contact either 911 or 311 or submit photos or videos via smart phones to a call-center to record their complaint. Bicyclists can summon police directly by taking pictures of motorists blocking bike lanes. New York City uses comprehensive metrics to measure just about everything, from emergency assistance calls to bike paths; from tree
plantings to detailed, agency-specific indicators. The mayor's office has a large video
screen that scrolls through many of these metrics all day, allowing officials and visitors
alike to constantly monitor performance.

In Miami-Dade County, Florida, citizens seeking information or filing complaints
against public agencies or private companies are directed to the appropriate agency
through a 311 call-center. The City of Miami utilizes a fully-online virtual procurement
system. The 311 system in Baltimore, Maryland is linked to the CitiStat database and
allows managers to better track the performance of city services such as responses to
calls for assistance, trash pick-up, street repairs or snow removal. The program also
records and plots data from residents’ service calls so public officials can identify trends
and respond appropriately. One-stop portals also provide citizens a single point-of-
contact to reach public officials when initiating routine actions such as moving, paying
taxes or applying for a driver's license, passport or social security number. Bling Nation
is a California-based service that allows users to pay utility bills and parking tickets
instantly with one swipe on their smart phones. These as well as many other cases
provide practical guidelines for concerned public officials who seek to save money,
ensure security, and better serve citizens by improving internal processes.

New media outlets such as blogs, social media, and voice over IP (VoIP) systems
add a face-to-face element to voice interactions, moving beyond what a simple phone
call provides. Multi-way video connections such as AIM’s Video-Chat, Google Plus
and Skype, can further enhance conversations and improve citizen-official contacts,
trust in government, national security, and virtual training. In addition, Google has
developed a new web-based email productivity suite tailored for governments to meet
more stringent federal IT security benchmarks.11

Predictive policing has been used by the City of Memphis, Tennessee to identify
crime-prone intersections and deploy law enforcement to prevent crimes from occurring.
The Virginia Dept. of Emergency Management implemented a GPS-based system to
handle highway emergencies.12 Researchers at the Centers for Disease Control and
Prevention (CDC) in Washington think responders can reduce fatalities by receiving this
sort of information before arriving at an accident, especially if the telematics computers
could route alerts directly to 911 centers, enabling responders to reach accident-scenes
more quickly. The computer sends the details, referred to as Automated Collision
Notification (ACN), to the telematics company which then places a call to the
appropriate 911 center. This system has already been installed in millions of new
automobiles and is widely available.

Data Security, Politics and Digital Governance
The purpose of data protection is to ensure information privacy and security: this is a particularly difficult challenge for the public sector because government agencies must implement policy changes that deal with threats in real time. There are several major threats, especially in the data mining stage. Despite the good intentions of data mining as an effort to improve the delivery of public services, the collection and analysis of personal data by governments inevitably raises concerns about civil liberties. The same sets of concerns have been raised about commercial social media websites such as Facebook which can also misuse personal data. With the quickening pace of technological development, more individuals and organizations are likely to experience both benefits as well as threats inherent in big data collection.

Concerns about online privacy are among the most pressing issues. Media’s attention on high-profile security has raised growing public awareness and apprehension about ‘information sharing’ with any organization. The public is becoming increasingly aware of how pervasively data is used by e-commerce traders for “spamming” citizens. Many citizens are losing their trust in the government partially because of the distrust of government agencies that seem to have too much access and too little protection of personal data. In addition, “apps” and mobile access provides governments with opportunities for hidden decision-making processes and also gives agencies more ways to gauge what citizens expect in terms of digital governance. However, there are obviously drawbacks to this tactic as it creates more opportunities for those with malicious intent to create apps that might include malware.

Only about one in five (20%) of non-e-commerce users trust that the government will keep records confidential. Deeper confidentiality and ethical problems are triggered by the evolution of GIS, social networks, mobile and wireless technologies. The ethical and ideological implications of added reliance on big data collection have not been adequately addressed by the research literature. Without complementary improvements to protect individual privacy and data security, it is possible for outside users to obtain access by using someone else’s computer address for themselves or, the worst scenario, stealing data, personal information or even identities. Data mining would be counterproductive in such situations. Income and social inequalities may be worsened by lack of knowledge of ‘safe’ internet sites. Those at the more vulnerable end of the social economic scale are likely to stop seeking advice and help if they suspect that the information will be noted and generally available.

Another concern about data sharing is that although automatic tracking systems were established by using data mining to assign a risk score to anyone entering the United States, there is no mechanism to find out the reason and challenge the high-risk recognition. There is also skepticism about government data mining as a threat to social cohesion. Paul Henman raises a rather more philosophical objection to government data-mining: that the technology starts to transform the nature of government itself, so
that the population is seen as a disaggregated set of sub-populations with different risk profiles rather than a single social body. Public administrators may be tempted to treat recipients in a detached way as a disaggregated class, rather than the traditional relationship of public agencies serving individuals in need of services.

There are various ways in which governments can improve their transparency and create a cost-effective balance using online platforms including social media and networks to promote positive citizen interactions. Guidelines for those agencies to reduce the risk of online mistakes or attacks are known, including applying malware filters, setting up intrusion detection programs, and using URL shortening preview tools and content moderation, which can help to limit potential damage. Three critical policies help to establish public trust: the establishment of a privacy policy for personal and proprietary information; information misuse safeguards and online security; and particularly stringent regulations for financial transactions and online payments. In order to ensure the restoration of public trust in government and increase the rate of adoption of data-driven methodologies, concerns over data security must be addressed. Although there are unresolved technical issues about data security and wrangles over data collection in specific fields, digital governance has a promising future with a comprehensive developmental strategy.

**Data Wrangling** Data may seem to be a nonpartisan term but, in fact, it is hardly neutral when it comes to impacts of decisions about what type of information should be collected, interpreted and released for use during policy-making processes. Rarely can data itself resolve a problem. In the political sphere, data ownership is a source of power. Battles over the control of raw data, how and when it will be released can become contentious political issues. And there is always the political reality that ideological bias clouds objective fact-finding and guides decision making processes.

One typical example is the frequent and fierce disputes over the collection and release of data in the field of environment policy. Although politicians generally agree that more environmental data is better, how much data collection to fund and for what purpose often starts another debate. Politicians like to talk about the need for more data, but it is rarely anyone’s top priority. Even when instrumentation is regularly funded, money is often lacking to maintain the data or to make them sufficiently accessible or digestible. And, if data collection and processing were to be institutionalized, another ongoing debate would emerge—how insulated the operation should be from politics. What’s more, industry has also been concerned about alleged inaccuracies in data on government websites. The Bush administration’s so-called Data Quality Act of 2001 (Section 515 of the Consolidated Appropriations Act (Pub.L.106-554) requires federal agencies to enable private parties to challenge the accuracy of information being disseminated by government. Anti-EPA business groups have used
this provision to delay regulatory compliance; the law has been an anathema to environmental groups, which view the provision as a way to stymie necessary regulation.

Conclusions and Future Prospects

Data-driven digital governance will continue to achieve technological breakthroughs in the future when human resources, intergovernmental cooperation and regulatory policies are coordinated to reach common ground to seek common goals.

Trained and skilled personnel are crucial elements of a good data management. Effective data management needs to be linked with both committed political leaders and skilled public employees. Data alone will never change behavior or improve performance. In the case of Baltimore’s CitiSat model mentioned above, top deputies are placed to be in charge of presiding over data review sessions. At the same time, employees are trained to be more proficient at entering data to GIS, conduct sophisticated data mining, and effectively analyze data. Case studies confirm that government personnel can be trained to be equipped with necessary skills to apply data to policy and management decisions, especially at the time of initial implementation.

The collaboration of multiple agencies for achieving the goals of digital governance is another important factor. Data security cannot be assured by any one agency alone, so there is a need for public, private and nonprofit sectors to share data and come up with solutions jointly. Governments generally fall behind in managing performance because they rely on “lagging indicators” to evaluate agency performance. Future decisions are based on data collected in the past. This often results in a “too little too late” abdication of oversight responsibilities. To help improve agency effectiveness, managers need to up-to-the-moment data. Moreover, governments at all levels need to rely more on continuous real-time data review, which allows them to perceive problems immediately and take actions in time to prevent them from becoming unmanageable.

Finally, continued and improving regulatory enforcement should be carried out at the same time. The adoption of e-rulemaking by the U.S. federal government is an example of the use of digital technology to enable government agencies to manage the rulemaking process more productively; it could also expand and enhance the public’s involvement in this process. In addition to the potential improvement of general public services leading to decreased administrative costs, the adoption of e-rulemaking can also promote data mining, processing and management capabilities.

Modern ICTs are the technical building blocks for successful digital governance. With these systems in place, data collection and analysis will be less time consuming
and more cost-effective. Data today are much more easily collected, transferred, and disseminated. Governments worldwide are under pressure to gather various data such as spending patterns, population growth and demographic migration. Confronted with a rapid expansion of data, governments are adopting new technologies to solve potential problems. Both private and public sectors face difficulties in being able to create, manipulate, and manage big data. Since big data comes from different sources, it is more challenging to analyze with standard tools and procedures. Increasing dependence on diverse sources of data reinforce government incentives to set up accurate and reliable data collection systems. Taking the value of big data used in private counterparts as an example, governments are beginning to follow similar processes to not only implement policy but also enhance citizens’ satisfaction with public service delivery in a cost-effective way. Moreover, data sharing is also a critical part of data-driven digital governance. After 9/11/01, as one form of G2G interaction, data sharing among government agencies and departments at different levels in the United States was reinforced, which led to numerous demonstrable improvements in national defense and emergency management.

Government officials acknowledge the need to reconfigure service delivery systems with big data capabilities to anticipate and respond to citizen needs: few are equipped with the skill sets or benchmarks needed to fully integrate and exploit information contained in exponentially-growing databases. Many public agencies are emerging from the early phases of what former Indianapolis Mayor Stephen Goldsmith labeled the era of "preemptive government." Some cities and states are pioneering innovations based on the analysis of big data from multiple sources. These metrics have their limits: They tend to measure activity more than outcomes; they are often ‘rear view mirrors’ which look back at data rather than focus forward; and, they too often concentrate on the actions of a single agency. But they provide a critical first step toward a new and broader effort to utilize cross-agency analytics to preempt problems before they occur. Without capable personnel to operate advanced data collection systems, however, there are limits on the extent government can merge customer service goals using big datasets. Enhanced performance metrics and digital warehouses make up the building blocks for the future expansion of digital governance.

References


**Author Bio:** Michael E. Milakovich (Ph.D. Indiana; B.A. University of California-Santa Barbara) is a professor of political science and public administration in the College Arts and Sciences at the University of Miami, Coral Gables, Florida. He served as MPA Director, University of North Carolina-Greensboro, and Project Director for the University of Miami Institute for Quality in Manufacturing and Service. His areas of specialization are: public management and productivity improvement, electronic and digital governance, business-government relations, administration of justice, and healthcare management. He has authored or co-authored 15 books or monographs and published numerous articles, chapters, and reviews in referred journals such as the *International Public Management Review, Public Performance and Management Review, American Review of Public Administration, American Political Science Review, Health Care Management REVIEW, Journal of Electronic Democracy, National Productivity Review, National Civic Review, Public Administration, Quality Assurance and Utilization Review, and The Journal of Health and Human Resource Administration*. He is a Vietnam era U.S. Army veteran; Life Member of the American Society for Public Administration (ASPA); serves as an expert witness in state and federal courts; advises numerous public, private, non-
profit and international organizations on policy analysis and quality improvement strategies; and is a member of several editorial boards. He has conducted global interdisciplinary research, consulted and lectured with numerous businesses, governments, and health care organizations in the United States as well as Austria, the Bahamas, Czech Republic, Denmark, Egypt, Great Britain, Hungary, Germany, Japan, Martinique, Turkey, United Arab Emirates and Venezuela. (Full cv available at: http://www.as.miami.edu/mpa/Michael_Milakovich_CV.pdf )