NASCIO Representing Chief Information Officers of the States

NASCIO Series on Analytics

Is Big Data a Big Deal for State Governments?

The Big Data Revolution – Impacts for State Government – Timing is Everything

NASCIO Staff Contact:
Eric Sweden, MSIH MBA
Program Director,
Enterprise Architecture & Governance
NASCIO

NASCIO represents state chief information officers and information technology executives and managers from state governments across the United States. For more information visit www.nascio.org.

201 East Main Street, Suite 1405
Lexington, KY 40507
Phone: (859) 514-9153
Fax: (859) 514-9166
NASCIO@AMRms.com
www.NASCIO.org

Introduction

"Big Data" is in the press and on everyone's mind. In this issue brief, the universe of "big data" will be explored in order to:

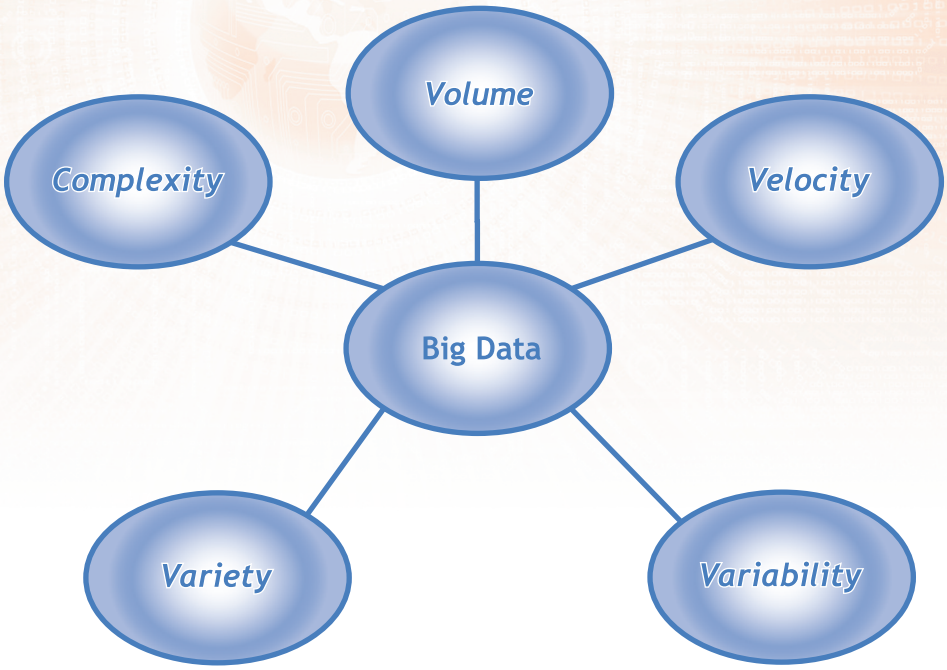
- Create a foundation preliminary to further description and exploration in future briefs, conference sessions and innovations forums.
- Set common characteristics of big data versus simply lots of data.
- Emphasize the necessity of data governance and data management within a broader state government enterprise architecture.
- Present some early recommendations for state government regarding big data.

What is Big Data?

There has been significant press on the idea of "big data." The term has a variety of definitions and has been used in a variety of contexts. To facilitate further discussion it is necessary to be clear on what we mean when we use that term. The common themes for big data are Volume – Variety – Velocity – Complexity – Variability. It is important to include all of these characteristics because otherwise the term "big data" may continue to be applied to a variety of circumstances diluting the real meaning. The use of the term "big data" should carry *all* of these characteristics. Nevertheless, definition still remains an issue – and a big enough issue that establishing a clear definition is one of the priorities of the TechAmerica Big Data Commission. In the meantime, NASCIO will use these five characteristics that help to shape the definition of big data. Further explanation of these themes is presented in Appendix A.

Big data is relevant to state government due to the vast amount, variety, complexity and variability of data produced by state government. State government may be described as an enormous *data generation engine*. And the sky is the limit in terms future data generation based on the growth in mobile applications, sensors, cloud services and the growing public-private partner-

ships that must be monitored for performance and service levels. The challenge is that many state government agencies are still being run as islands of information versus members of a single state government enterprise. The result is state government is not fully exploiting the data it has at hand.



Characteristics of Big Data

"Big data" and "big economics"

State governments, particularly state CIOs, need to care about the growing activity around big data. There is a driving force for better decisions based on analytics, business intelligence AND big data. The value of big data lies in its use in managing the enterprise, improving operations and supporting inter-enterprise value chains. This holds true whether the enterprise is a for-profit entity, a government, an academic institution, a non-profit, or an inter-governmental collaborative. Data by itself is of little value until it is turned into information, generates knowledge, and enables wisdom in making decisions which in turn results in outcomes. In the case of state government, most important are citizen outcomes.

Analytics is driving new capabilities for competitiveness and effectiveness. Data must have the currency and reliability applicable to the particular analysis being undertaken. Those analytics must be applied within the context of the specific issue or decision under consideration, the needs and intentions of the analysis, and the associated risks and opportunities related to the decisions that will rely on that analysis. Data becomes relevant in the greater context of identifying and understanding the full portfolio of issues facing state government. Not all data must be of the highest quality. The quality of the data will depend on: the purpose of its use; the magnitude of related outcomes and potential resource investment to achieve those out-

comes; time-to-effect or time-to-impact of an issue; and the level of reliance and criticality to government strategy and operations.

Owing to the availability of new data sources, analytics is able to accelerate and become increasingly comprehensive. It is the drive for analytics that is creating the seeming frenzy for big data. MIT and IBM conducted surveys in 2010 and 2011 with the objective of gaining perspectives on the burgeoning demand for analytics. There was a 50% increase in the number of respondents to the 2011 survey (4,500 executives and analysts). Moreover, there was a 57% jump in the number of respondents that believe analytics provide "substantial" or "significant" contribution to effectiveness. This translates to competitive advantage for for-profit organizations and effective operational advantage for for-profit and non-profit organizations alike. Organizations with experience in analytics are expanding those capabilities.[1] They are applying analytics in more diverse lines of business and functions. In essence, analytics is permeating the enterprise and impacting more and more decisions.[2] Even now, while we're still at the shoreline for analytics and big data, the IBM / MIT surveys present a compelling message. *If that same study is done in 10 years – how will it compare with 2010?* There is an amazing journey ahead.

MIT's Center for Digital Business analyzed the financial performance of 179 large publicly-traded firms. They found that those firms that adopted a data driven approach to decision making are about 5% more productive and profitable than their competitors.[3] That can amount to 100's of millions of dollars per year. This research essentially makes the point that proper use of big data and analytics can significantly improve enterprise performance.

The United Kingdom anticipates significant economic stimulus from organizations that invest in big data technologies. There is the potential to generate £216 billion for the UK economy and create 58,000 new jobs over the next five years, according to a report published by The Centre for Economics and Business Research.[4] Based on currency exchange rates on June 27, 2012, £216 billion translates to 336 billion U.S. dollars.

Wikibon estimates that the current "big data market" for related services and products is over $5 billion in 2012 and will rise to $50 billion globally within the next five years.[5] See Appendix B for a graphing of this forecast. Additional forecasts regarding economic potential are presented by the McKinsey Global Institute 2011 report *Big data: The next frontier for innovation, competition, and productivity*[6].

- $ $300 billion in value generated through efficiencies in healthcare. 0.7 percent annual productivity growth. *Contribution – Demand for people, skills, and technology.*
- $ Shortage of analytical and managerial skills and talent. Shortage of 140,000 to 190,000 people with sophisticated analytical skills; Shortage of 1.5 million managers and analysts to analyze data and make decisions based on analytics. *Contribution to the U.S. economy – Demand for training services from private industry and academia.*
- $ Global demand for personal location data anticipated to drive over $100 billion in revenue for service providers and almost $700 billion in value derived by end users. *Contribution to the U.S. economy – Demand for location services.*

- The World Economic Forum convened in Switzerland in January of 2012 highlighted big data as a *new economic asset* comparable to currency and gold[7].

Enterprise Architecture: The need for "big" data management discipline

To prevent big data from becoming a "big quest" there must be rational, defensible reasons for pursuing it, and there must be a discipline for governing and managing the associated investment of people, finances, and technology.[8] Big data must be managed within a formal data management discipline within a formal enterprise architecture program. A formal data management discipline includes *data governance* – a topic explored in previous NASCIO issue briefs[9].

Enterprise architecture becomes even more important as the organizational structure of government encompasses more public-private partnering. In essence, the government enterprise is expanding to include *other* enterprises. External service providers will continue to become necessary partners in delivering government services. Those services will be employed through service level agreements, cloud services, mobile services, and off-site hosted solutions, all of which must be continually monitored and formally audited to measure demand and performance requirements through *inter-enterprise* data streams. These data streams will be used not only for ensuring service commitments, but also for calculating variable cost elements used in billing state government for services employed. This adds to the ever-expanding government data generation and the necessity for managing data and organizational complexity through data management and enterprise architecture.

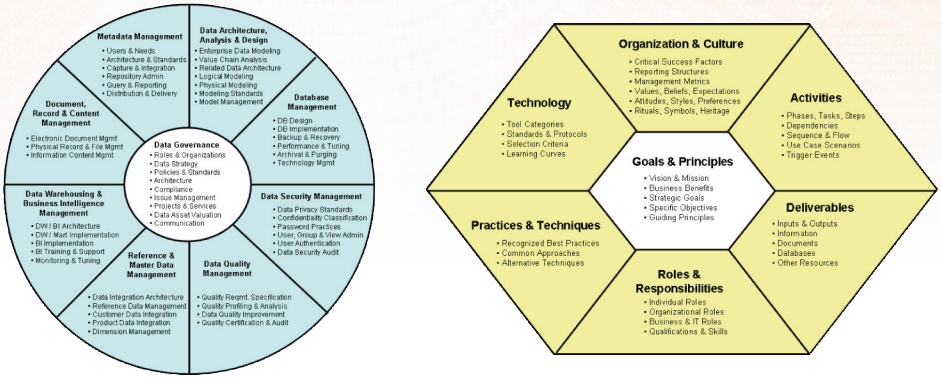
No matter what the form of data, it must be managed as a state government information asset. Data governance, possibly the most critical function of data management, increases in importance with the growing diversity in data sources, data partners, and value chains. The same *data management principles and techniques* apply to all data regardless of its form though the approach, methods and procedures will depend on the type of data and information. For further information see NASCIO's series on data governance – www.nascio.org/publications.

Big data does not replace more traditional data base design and implementation. Rather it supplements the state government portfolio of information assets by providing important additional data types and data streams captured real time.[10] Like more traditional data and information assets, big data sources and volumes must be *evaluated, classified, inventoried, valued and secured*. Further, the data generated from the analytics *applied to* any data must also be *inventoried, valued and secured*. As with any other data and information asset, big data must be managed within a records management lifecycle and archiving must be established. However, much of the *timing and future relevance* of big data are yet to be determined.

One of the defining characteristics of big data is that it affords the discovery of new patterns, new correlations[11], and new insights, in turn creating even more data. Data lifecycle will continue to be an area of exploration. When do certain types of big data become irrelevant to the state enterprise? That depends on the type of data, and what future value can be anticipated. So,

it will be difficult if not impossible to say when data become irrelevant to future issues and decisions. NASCIO's previous publications on records management described a case where data over a century old was still relevant to property line decisions.

As the volume of data continues to grow at a rate never before considered conceivable, data management strategy and governance within the greater context of enterprise architecture becomes even more critical. DAMA International has published a set of frameworks for data management that present the necessary functions and elements that must be included in a comprehensive, mature data management discipline. That same discipline must be applied to realm of big data.



The DAMA Functional and Environmental Frameworks from the DAMA Data Management Body of Knowledge (DMBOK)[12]

Data should be *evaluated* as it is created to estimate its quality as well as it utility. Analytics applied to the data will inherit the quality of the data. Big data gives rise to a new emphasis on *data lineage* – the path that describes where data was created and by whom; how it is transformed; how it flows; and how it was combined with other data. Data lineage can be very complex and must involve additional disciplines such as meta data management (i.e., data about the data); intellectual property; stewardship and ownership; security; authentication, authorization, and assurance. Data lineage must be managed to ensure the quality of the data is at the appropriate level when the data is created, and is not diluted as it moves through the path of capture, translation, aggregation, calculation and analysis. Every step adds to or erodes the ultimate quality of the data used for decision making. The higher the quality of the data, the more powerful are the conclusions drawn from the data analytics. Higher levels of data quality will be more expensive in terms of the resources needed for collection, security, authentication and authorization. The ultimate analysis that will use the data will determine what level of quality is necessary and the level of importance of the issue(s) that will be managed with that *data and analysis*. If the needs of that analysis change, so may the necessary quality of the data.

In government as well as in business, analysis will be centered on *issues*. That is, understanding issues and the impact of those issues on the enterprise. Issues can be presented as opportunities or threats, and some particular issues may present both. Effective analytics requires compiling a full

issues portfolio, known sources of business intelligence and any need to *discover* new, heretofore *unknown*, sources. A rendering of an *issues oriented approach* to analytics and business intelligence is presented in Appendix C. This approach will be modified and refined over time.

Investment in big data skills and technology *competes* with other investment pressures facing state CIOs. In general, states are still working on many other high priorities such as legacy modernization, consolidation and shared services, cloud computing, mobile services, and cyber security as presented in NASCIO's 2012 Top Ten Priorities[13]. Investment in big data is competing for attention and resources in this larger portfolio and requires an economic business case just as any other investment. That business case must start with a purpose, the rationale, and the need for investing now. It may still be too early for significant investment. To build the capability to exploit big data, as well as other data resources, state government must mature its data management discipline within its enterprise architecture program to prepare for the governance, management and harvesting of big data information assets. This *is* the right time to be exploring big data and running small initiatives to experiment and learn. Even small initiatives must be supported by a business case. NASCIO has made the case for including analytics, data and big data requirements analysis when state government considers legacy replacement or investment in capabilities in any government line of business. It is important that the state CIO and the state chief enterprise architect maintain a focus on further development of data management and analytics capabilities.

Big data means big analytics

As state government explores big data, it must include a move to develop data management practices and discipline specific to big data. What drives big data is the need for *analysis* of that data motivated by two factors: questions which cannot be answered adequately, or with enough confidence from existing information; or suspicions that relevant circumstances exist that are not being discovered through more traditional analysis. Seeking answers to those *questions* becomes part of enterprise strategic intent, leading to the launching of management initiatives, programs and projects. That means investment, the expenditure of capital. In times when state government capital is extremely limited, analytics, and *potentially analytics based on big data*, will contribute powerful insight and discrimination in the decision process for spending, investing, and cutting of state budgetary funds. *Big data initiatives should not be a solution looking for a problem*. Rather, there must be a defensible business objective in exploring big data, and certainly for investing in it. That said, big data is currently in a *discovery phase* for many organizations including state government. Big data concepts must be explored to assess their value now and in the future.

Anticipate that a next surge in analytics that will go beyond lines of business and functions to encompass enterprise level application, coordination, and orchestration to optimize the enterprise even while potentially "suboptimizing" certain functions, processes, and lines of business. This new generation

"This is that level of importance, the future of computing is not just big iron. It's big data."

Tom Kalil, Deputy Director White House Office of Science and Technology Policy[18]

of analytics, integrated with operations, strategic planning and enterprise architecture, may mature the demand for big data – or certain big data – as well as "just plain old data" (JPOD). Plus it will drive the need for data management discipline and technologies for proactively managing big data. As mentioned above and explained in NASCIO's series on analytics - *analytics inherits the quality of the data it uses*. Reciprocally, as more organizations mature their analytics capabilities, and people become more sophisticated in their use of and assessment of both analysis and the underlying data that the analysis relies on, data quality will continue to improve.

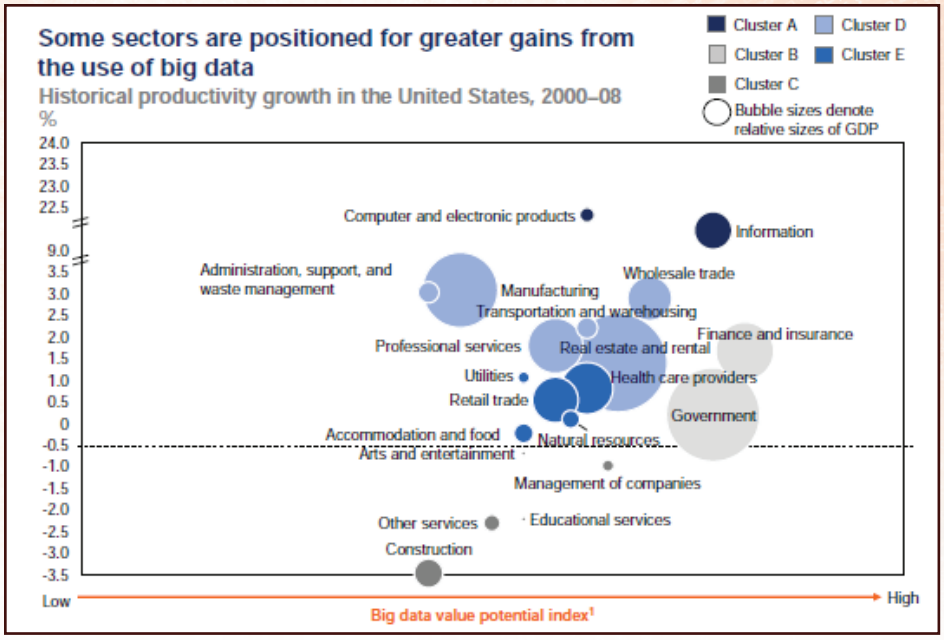
The amount of data in the world is exploding –a large portion of this comes from the interactions over mobile devices being used by people in the developing world - people whose needs and habits have been poorly understood until now. Researchers and policymakers are beginning to realize the potential for channeling these torrents of data into actionable information that can be used to identify needs & provide services for the benefit of low-income populations. This discussion note is a Call-to-action for stakeholders for concerted action to ensure that this data helps the individuals and communities who create it.

World Economic Forum

"The world's data is doubling every 18 months, presenting government and industry with new opportunities to transform information into insight. New database technologies and applications, coupled with real-time analysis of big data, will help business and government run better and ultimately improve the well-being of customers and citizens."

Steve Lucas
Global Executive Vice President and General Manager
Database and Technology
SAP
Source: GCN (http://s.tt/1d5NB)

According to the McKinsey Global Institute, government holds some of the greatest potential for productivity gains through proper leveraging of big data.[15] Anticipate that government will indeed develop and mature its ability to exploit big data as part of a growing and maturing government management discipline – *i.e., government enterprise architecture*.



Potential Productivity by Sector
McKinsey Global Institute[16]

> "In the same way that past Federal investments in information-technology R&D led to dramatic advances in supercomputing and the creation of the Internet, the initiative we are launching today promises to transform our ability to use Big Data for scientific discovery, environmental and biomedical research, education, and national security,"
>
> *Dr. John P. Holdren,*
> *Assistant to the President and Director of*
> *the White House Office of Science and Technology Policy*[17]

Conclusion

Big data carries many, big implications – for better and for worse. State government should be preparing now for the potential of big data and ensure current investment in technology allows for future leverage of big data capabilities. NASCIO will explore a number of topics in the future related to big data and will stay connected to developments at the federal and national level as the research agenda moves forward. Topics that can be anticipated in the future include enterprise architecture management, privacy, data lineage and quality, cost / benefit analysis, the necessary training agenda and applications in specific government lines of business.

Calls to Action

- Anticipate more activity including case studies, management approaches, software and services around big data. Big data is coming and it is of growing importance but, actual purchases should be restricted to those supporting a strong cost benefit analyses. Any early investments in big data should be limited to specific opportunities instead of blanket capabilities – or full suite enterprise solutions.
- Evaluate investment in big data in line with other needed technology investments such as legacy modernization. And, coming from the broader perspective in enterprise investment and procurement, evaluate and plan IT investment with consideration of future maturation in data management capabilities that include: big data; location aware data; sensor data; data generated along value chains and supply chains involving inter-enterprise relationships such as public-private partnerships.
- What state CIOs should avoid is the proliferation of big data initiatives within agencies that create "big data silos." The whole point of big data is to pull together a variety of data from multiple sources and, investigate or discover new data relationships.
- Best to start with highly focused big data initiatives with clear objectives that can be achieved relatively quickly, particularly if there is no formal data management or data governance in place for managing data currently within a broader enterprise architecture discipline or function. Big data initiatives need to be planned so they fit into the state's enterprise architecture. Such initiatives emphasize the need for enterprise architecture.
- Evaluate investment in analytics with an anticipation of more types of relevant data, and growing diversity of data. Anticipate maturation of capabilities in business intelligence and business analytics when evaluating investment in these capabilities. Be informed on vendor partner long term strategy for big data management and analysis.
- Continue to explore and evaluate the technology that is maturing for collecting, managing and analyzing diverse data sources within the context of *issue management*. *(See Appendix C)*
- Use the current and anticipated demand for employing big data as an incentive and motivator to establish and/or grow internal data management discipline within a formal enterprise architecture program. Investment in big data must be founded in investment in data management.
- Monitor what data is being automatically collected by government and the commercial sectors. Be aware there may be security and privacy issues with automatic collection of data on citizens.
- Stay tuned to NTIA code of conduct regarding the capture of data – www.ntia.doc.gov.
- Stay tuned to NIST and TechAmerica research agendas.

Contributors

***Peter Aiken**, President, DAMA International*

***Glen Bellomy**, Datacenter Architect, Symantec*

***Brian H. Cameron**, Ph.D., Executive Director, Center for Enterprise Architecture, The Pennsylvania State University*

***Patricia Cummens**, Government Strategist, ESRI*

***Matthew D'Alessandro**, Business Development Manager, Motorola; President, The IJIS Institute*

***Jack Doane**, Director, Department of Finance, Information Services Division, State of Alabama; Co-Chair of the NASCIO Enterprise Architecture & Governance Committee*

***Lauren Farese**, Senior Director, Oracle, USA*

***Tim Finnegan**, Sales Director, State and Local Government, SAS Institute*

***Jeremy Forman**, Enterprise Architecture Program Director, Oracle, USA*

***Liz Fossett**, Research Assistant, NASCIO*

***Karen Gee**, HP Applications Practice Leader, State/Local Government and Education Division*

***Sam L. Hearn, Jr.**, Graphic Designer, AMR Management Services*

***Kennan Hogg**, Regional Sales Manager, Software AG*

***Ashwini Jarral**, Assistant Director, Technology Services, The IJIS Institute*

***Sue Ann Lipinski**, Acting Director, Business Solutions, State Office of Technology, State of West Virginia*

***Janet Mostow**, Enterprise Architect, Oracle Public Sector Technology*

***Paulina Orlikowski**, USPS State & Local Division, Portfolio Manager, HP*

***Bob Otterberg**, HP CTO, State/Local Government and Education Division*

***Dan Paolini**, Deputy CTO for Data Management Services, Office of Information Technology, State of New Jersey*

***Carolyn Parnell**, Chief Information Officer, State of Minnesota; Co-Chair of the NASCIO Enterprise Architecture & Governance Committee*

***Bethann Pepoli**, Chief Technology Officer , EMC*

***Van Ristau**, Chief Technology Officer, DLT Solutions Inc.*

***Doug Robinson**, Executive Director, NASCIO*

***Sanjay Shirude**, VP Research and Education, DAMA International*

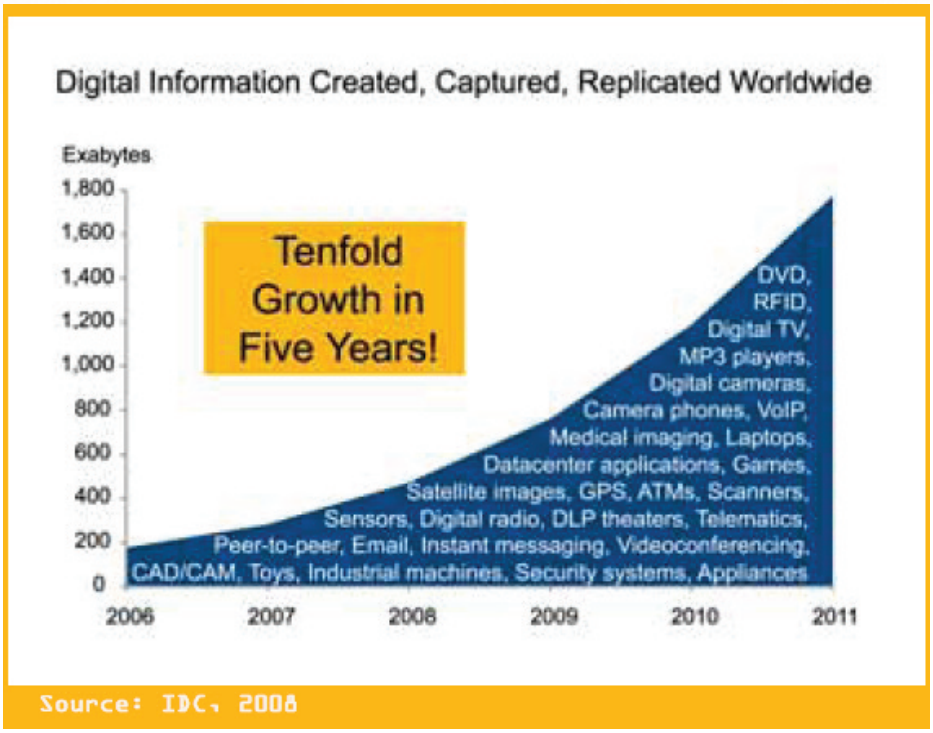
***Shawn Vaughn**, Membership and Communications Coordinator, NASCIO*

***Tom Walters**, Enterprise Business Process Architect, Commonwealth Office of Technology, Office of Enterprise Technology, Division of IT Governance*

Diana Zavala, HP Information Management & Analytics Industry Executive, U.S. Public Sector

Appendix A. Big data characteristics:

- **Volume** - There is more and more data being generated. It is estimated that the data volume is increasing by 40 percent per year. More data means more data is available for analysis. This includes data generated by sensors, social media, email, photographs, video footage, machine data, spatial data, weblog data.[19] The cost of data storage is decreasing so it is possible and economical to store larger amounts of data for longer periods of time. One issue is the relevance of data. What data should be captured and stored for analysis? The answer is - "Don't know yet!" With the advent of predictive analytics – it is not known what data will prove useful in developing predictive models. So the volume of data to be captured will only increase as analysts explore any and all data to *find* useful data and data relationships. Government is generating vast amounts of data from a broad portfolio of sources such as sensors, mobile applications, health exchanges, tax reporting, life events reporting, network traffic monitoring and regulatory compliance reporting.



Data Growth

- **Variety** – There are a myriad of data types, formats, structures, and models. We have moved from an emphasis on highly structured transactional data that can be stored in rows and columns to data that is unstructured. Unstructured data includes text, video, audio, graphics, maps, social media data, and web images. In fact, most data is of this "unstructured" form. Of course the definition of "structured" data comes from the previous emphasis on organizing data into rows and columns. That term may be redefined in the future because there is structure to other forms of data. The methods

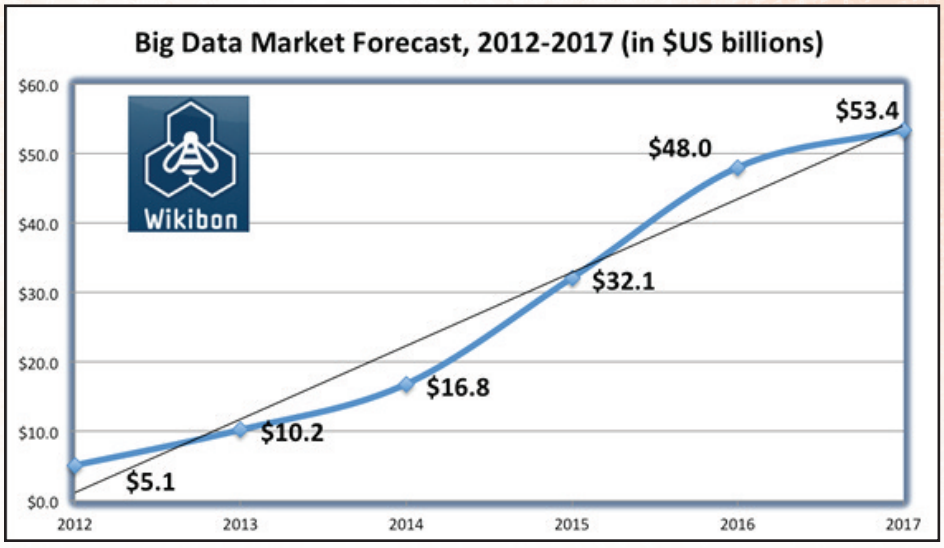
for capturing, organizing and analyzing data is still determined by the "structure" of the data. Video data is analyzed with different methods and procedures than financial transactional data that is primarily numerical. This presents an issue because there may be patterns of interest that need to be uncovered that involve data that is "structured" differently. For example, correlations between behavior captured on video to behavior captured from transactional systems, or machine logs. The variety across government is as vast as the number of agencies, government processes, and government services.

- **Velocity** – This refers to the speed of data creation. With the advent of weblogging, network traffic logging, sensors of various types including RFID, video surveillance, and other data capturing technologies, data is being captured immediately. Data is being analyzed immediately as well. The advances in information technology have made this possible.[21] The velocity is increasing with the advent of more and more mobile apps, sensors, surveillance, network logging, audit and automatic reporting.

- **Complexity** – More types of complexity are being encountered partially due to the fact that more kinds of data can be captured. With the growing variety, variableness, sources, and formats, it becomes a challenge to then evaluate data using correlational analysis and other techniques when the data being analyzed is not uniform in format, context, time periods, or frequency of collection. It is challenging to pull data from a myriad of systems used by a wide variety of agencies and lines of business that use different terms and definitions. For example, pulling together school data, court data, child welfare and enforcement, and public safety data in order to ensure children are protected from being placed in unsafe situations.

- **Variability** – This term is different from the term variety. Variability refers to the patterns for data generation. Data generation is not linear. The very pattern of generation of data is itself a type of data. The patterns for data generation can be used to uncover behaviors, relationships and other information that can be explored to find additional correlations. Much of government data have associated cycles where the data is often generated in peaks and valleys. Examples include audit reports, tax returns, accounting cycles, and regulatory compliance.

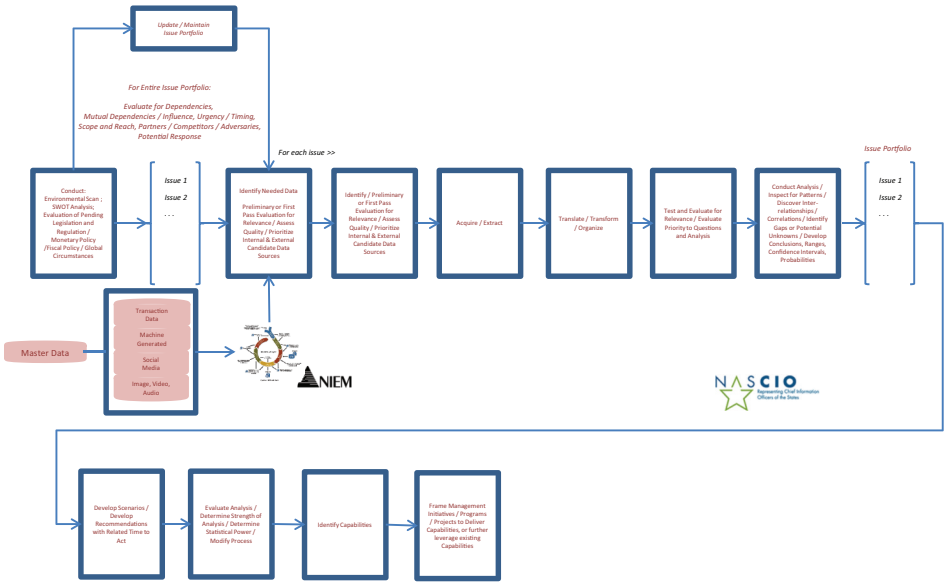
Appendix B. Market Forecast for Big Data



Source: Wikibon 2012

Appendix C. Candidate Rendering of The Issues Portfolio and Big Data Process

The following is a candidate conceptual process model for analytics with a view towards big data and issue portfolio management. This is provided as a starting point for state government to develop a formalized portfolio management for data sources, issues and analytics.



Issue Portfolio Management

DISCLAIMER
NASCIO makes no endorsement, express or implied, of any products, services, or websites contained herein, nor is NASCIO responsible for the content or the activities of any linked websites. Any questions should be directed to the administrators of the specific sites to which this publication provides links. All critical information should be independently verified.

This project was supported by Grant No. 2010-DJ-BX-K046 awarded by the Bureau of Justice Assistance. The Bureau of Justice Assistance is a component of the Department of Justice's Office of Justice Programs, which also includes the Bureau of Justice Statistics, the National Institute of Justice, the Office of Juvenile Justice and Delinquency Prevention, the Office for Victims of Crime, and the SMART Office. Points of view or opinions in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

---

[1] "Analytics: The Widening Divide." *MIT Sloan Management Review Research Report*. Fall 2011. Joint study by the MIT Sloan School of Management and the IBM Institute for Business Value. Retrieved on January 11, 2012, from http://sloanreview.mit.edu/feature/analytics-advantage-introduction/
[2] NASCIO Analytics Series, available at www.nascio.org/publications.
[3] Brynjolfsson, E., *Erik Brynjolfsson on Big Data: A revolution in decision-making improves productivity*. Retrieved on June 27, 2012, from http://mitsloanexperts.com/2012/02/15/erik-brynjolfsson-on-big-data-a-revolution-in-decision-making-improves-productivity/.
[4] Nguyen, A., "Big Data to add £216bn to UK economy," *Computerworld UK, April 5, 2012*. Retrieved on June 27, 2012, from www.computerworlduk.com/news/it-business/3349601/big-data-add-216bn-uk-economy/.
[5] Kelly, J., Vellante, D., Floyer, D., "Big Data Market Size and Vendor Revenues." *Wikibon*. July 24, 2012. Retrieved on August 9, 2012, from http://wikibon.org/wiki/v/Big_Data_Market_Size_and_Vendor_Revenues.
[6] *Big Data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute. May 2011. Retrieved on 6/21/2012 from www.mckinsey.com/mgi.
[7] *Rethinking Personal Data: Strenthening Trust, World Economic Forum*. May 2012. Retrieved on June 27, 2012, from www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development.
[8] This concept of avoiding "quests" is derived from Gopal Kapur's "Management's Seven Deadly Sins". See podcasts published at www.center4pm.com/?page_id=163.
[9] See NASCIO's *series on Data Governance; series on Analytics (Do You Think or Do you Know); series on Records Management at* www.nascio.org/publications.
[10] *Integrate For Insight: Combining big data tools with traditional data management offers enterprises the complete view*. Oracle White Paper. Retrieved on 7/17/2012 from Oracles big data website www.oracle.com/index.html.
[11] The term "correlation" in this context is intended to convey the statistical meaning of correlation. Correlation in general statistical terms refers to fact that two variables or measures are related. The coefficient of correlation is an indicator of the strength of a linear relationship between to variables.

[12] DAMA *Data Management Body of Knowledge* retrieved from www.dama.org/i4a/pages/index.cfm?pageid=3364

[13] NASCIO's Top Ten Priority Strategies, Management Processes and Solutions and Top Ten Priority Technologies, Applications and Tools is published annually. The 2012 priorities are available at www.nascio.org/publications.

[14] Yasin, R., "New commission aims to help government tap into big data", May 31, 2012, *Government Computer News*. Retrieved on June 27, 2012, from http://gcn.com/articles/2012/05/31/techamerica-selects-big-data-commission-leaders.aspx.

[15] *Big Data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute. May 2011. P. 100. Retrieved on 6/21/2012 from www.mckinsey.com/mgi.

[16] *Big Data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute. May 2011. P. 100. Retrieved on 6/21/2012 from www.mckinsey.com/mgi.

[17] *OBAMA ADMINISTRATION UNVEILS "BIG DATA" INITIATIVE: ANNOUNCES $200 MILLION IN NEW R&D INVESTMENTS*, White House Press Release, March 29, 2012. Retrieved on 7/19/2012 from www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf.

[18] Lohr, S., "New U.S. Research Will Aim at Flood of Digital Data." *The New York Times*. March 29, 2012. Retrieved on 7/19/2012 from www.nytimes.com/2012/03/29/technology/new-us-research-will-aim-at-flood-of-digital-data.html.

[19] *What is Big Data?* Teradata. Online video of conversation between Teradata Labs President Scott Gnau and Teradata vice president of development, Mayank Bawa. Retrieved on June 21, 2012, from www.teradata.com/videos/what-is-big-data/.

[20] *The Diverse and Exploding Digital Universe*, IDC, March 2008. Retrieved on July 17, 2012 from www.emc.com/utilities/search.esp.

[21] *Big Data*. Online article by SAS retrieved on May 31, 2012 from www.sas.com/big-data/index.html.