Are We Ready for the Era of Big Data?
Issues and Challenges in Big Data Management in the Public Sector

The 2014 APPAM Spring Conference, Washington DC
April 12, 2014

Minyoung Ku, Doctoral Candidate in Public Administration,
mku@albany.edu

Mino Ku, Doctoral Candidate in Computer Science
Konkuk University, South Korea

Presentation Overview

- Big Data and Big Data Management

- Four Dimensions of Big Data Management

- Issues and Challenges in Big Data Management

- Implications for MPA/MPP Programs

- In 2012, the Obama Administration announced it would invest more than $200 million in big data R&D initiatives

- Federal agencies are managing over 150 big data projects (Biometrics Research Group, 2013), and their spending on big data will reach $7.2 billion by 2017

- What is big data?

  ➢ **Volume**: The era of zettabytes already began in 2010 (Grantz & Reinsel, 2010). The U.S. Library of Congress collected 235 terabytes of data (equal to 18 billion pages of digital documents) in April 2011

  ➢ **Variety**: 90% of existing digital data in the the world is unstructured data (e.g., documents, pictures, videos, and voice files) (DCI, 2011)

  ➢ **Velocity**: 90% of existing data in the world has been created in the last two years alone (IBM)

- Big data management is, in a broad sense, the whole effort of an organization to generate, collect, store, organize, process, retrieve, analyze, and share large-scale data in various forms, which are generated by various sources in real time or near real time, in order to create value.

- Four Dimensions of Big Data Management

  ➢ Technical Dimension

  ➢ Data Dimension

  ➢ Managerial Dimension

  ➢ Legal/Political Dimension

- Big Data Collection
  - ➤ Highly-scalable computer processing systems & large bandwidth networks

- Big Data Cleaning
  - ➤ Standards for levels of data cleaning techniques and cleaning-dedicated computer systems

- Big Data Storage
  - ➤ Distributed computing systems with high processing capacity
  - ➤ Accurate data usage prediction and sophisticated metadata are necessary

- Big Data Analysis
  - ➤ Tools to manage a processing job (job distribution, job processing, job monitoring, and collecting sub-jobs) e.g., Hadoop

Data Challenges

- Data Quality
  - ➢ Data accuracy, completeness, reliability, consistency, accessibility, interpretability, and timeliness

- Data Standardization
  - ➢ Reduce data isolation and enhance data integration
  - ➢ Metadata standards / 'community-level' standardization (Lynch, 2008)

- Data Classification
  - ➢ Classify data by diverse standards
  - ➢ A comprehensive data classification standard – the guidelines in technical, managerial, and legal domains

UNIVERSITY AT ALBANY
State University of New York

- Goals and Visions
  - ➢ Clear and realistic goals
  - ➢ Alignment between organizational and project goals and between goals of different organizations

- Executive Leadership and Support
  - ➢ Far-sighted leadership and strong, mid- or long-term support
  - ➢ 26% and 36% of IT experts at the federal and state levels, respectively, have been frustrated with frequent leadership changes in big data management in their organizations (the TechAmerica Foundation, 2013)

- Resource Management
  - ➢ Shortage of IT resources and data scientists in the public sector
  - ➢ Outsourced resources? or In-house resources?

- Organizational Culture
  - ➢ Decision making by HiPPOs, relying on intuition and personal experience (McAfee & Brynjolfsson, 2012)
  - ➢ Data-driven culture; appreciating the value of data and data analytics
  - ➢ Culture in favor of information/data sharing

- Turf War brewing over…

  Information/Data Privacy
  - ➢ Personally Identifiable Information (PII) matters
  - ➢ The foundations of privacy protection, by the Fair Information Practice Principles, includes notice/awareness and choice/consent
  - ➢ The "Do Not Track" debate

  Data Breach
  - ➢ Data Breach Notification Law and Regulation

Rockefeller College
of Public Affairs and Policy

NASPPA Five Core Competencies

#1 The ability to lead and manage in public governance

#2 The ability to participate in and contribute to the policy process

#3 The ability to analyze, synthesize, think critically, solve problems and make decisions

#4 The ability to articulate and apply a public service perspective

#5 The ability to communicate and interact productively with a diverse and changing workforce and citizenry

UNIVERSITY AT ALBANY
State University of New York

Thank you.

Questions and Comments

Minyoung Ku: mku@albany.edu