University of the Philippines-Diliman

National College of Public Administration and Governance

Center for Policy and Executive Development

BUILDING CAPACITY FOR BIG DATA APPLICATIONS IN GOVERNMENT

Lecturer:

DR. XUN WU

Professor, Division of Public Policy,

The Hong Kong University of Science and Technology

14 May 2019

NCPAG Case Room, RP De Guzman St.,

University of the Philippines-, Diliman, Quezon City

Program

WELCOME REMARKS
Dr. Maria Fe Villamejor-Mendoza
Professor and Dean, University of the Philippines-
National College of Public Administration and Governance (UP-NCPAG)

INTRODUCTION OF THE GUEST LECTURER
Dr. Ebinezer R. Florano
Associate Professor and Director, Center for Policy and Executive Development, UP-NCPAG

LECTURE
Dr. Xun Wu

SYNTHESIS
Dr. Noriel Christopher Tiglao
Associate Professor, UP-NCPAG

LECTURER'S PROFILE
Dr. Xun Wu

Dr. Xun Wu earned his PhD from the University of North Carolina, Capel Hill. Professor Xun Wu is a policy scientist with a strong interest in the linkage between policy analysis and public management.

Trained in engineering, economics, public administration, and policy analysis, his research seeks to make contribution to the design of effective public policies in dealing emerging policy challenges across Asian countries. His research interests include policy innovations, water resource management, health policy reform, and anti-corruption.

He has consulted for the World Bank, the Asian Development Bank, UNEP, International Vaccine Institute, and numerous government agencies on a variety of topics, such as infrastructure planning and development, environmental and social impact assessment, design and implementation of randomized control trials (RCTs), and Public-Private Partnership.

Prior positions in the academe: faculty and director of the Institute of Water Policy, Lee Kuan Yew School of Public Policy, National University of Singapore (2001-2015) and lecturer, Renmin University of China (1988– 1992).

Source: https://iems.ust.hk/people/faculty-associates/xun-wu

MESSAGE FROM THE UP-NCPAG DEAN
Dr. Maria Fe Villamejor-Mendoza

"Professor Xun Wu shall offer how government, the academe, the public sector and everyone of us, we, make sense of big data."

"In the tradition of the NCPAG guest lecture series, NCPAG took the opportunity of Dr. Wu's current engagements in the Philippines to request a lecture to our students and publics. The lecture has generated much interest outside the college."

"We have experts in data science analysis present —Dr. Erwin Alampay and Dr. Noriel Christopher Tiglao. Let us see the mixture of expertise on this topic today."

BIG DATA DEFINED

Big data refers to the dynamic, large and disparate volumes of data being created by people, tools and machines; it requires new innovative and scalable technology to collect, host and analytically process the vast amount of data gathered.

Big data includes information garnered from social media, data from internet-enabled devices (including smartphones and tablets), machine data, video and voice recordings, and the continued preservation and logging of structured and unstructured data.

4Vs of Big Data

- Volume: the amount of data being created is vast compared to traditional data sources
- Variety: data comes from different sources and is being created by machines as well as people
- Velocity: data is being generated extremely fast – a process that never stops, even while we sleep
- Veracity: big data is sourced from many different places, as a result you need to test the veracity/quality of the data.

"Public sector is still falling far behind in Big Data Applications."

While there is growing interest in big data applications in public sector, public sector is still falling behind the private and science sectors in using big data.

Public agencies are in the early days of their big data efforts.

List of Types of Potential Big Data Projects in the Public Sector

The use and combination of multiple large datasets, from both external and internal sources.

The use and combination of structured and unstructured data in analysis activities.

Real-time or near-real-time streams of incoming data which are structurally handled and analyzed.

The use of advanced analytics and algorithms, distributed computing and/or advanced technology to handle very large and complex computing tasks.

Innovative use of existing datasets and/or data sources for new and radically different applications.

Potential for Big Data Applications in the Public Sector

- Strengthening Government Service Delivery: Big data analytics can be used by governments to improve existing services and to draw on novel datasets to drive entirely new public services
- Smarter Policymaking: Policymakers are using satellite imagery, cell phone data – and real-time – policy insights
- Deepening Citizen Engagement: By applying machine learning to online and social media, governments can be more responsive to citizen sentiment, ushering in a new dimension of civic engagement.

Institutional Challenges for Implementing Big Data Projects

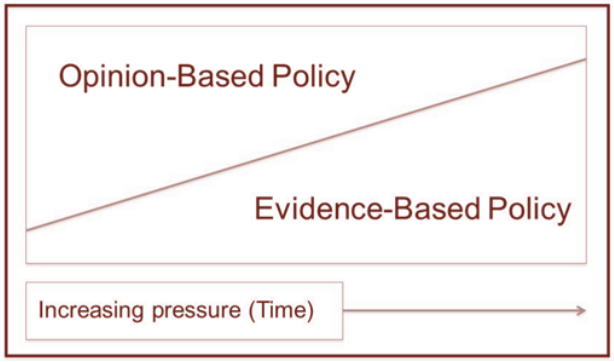
- **Financial investment:** Need for high initial investments to implement data analysis technology and acquisition of appropriate equipment, capable of processing the huge data flow
- **Data ownership:** The challenge here is to determine who is the legal "guardian" of the civilians' data. The government must take responsibility for the use and protection of that data.
- **Data security:** The purpose of data protection is to ensure the privacy and security of information. In the public sector this is a particularly difficult challenge because government agencies must implement policy changes that address real -time threats.
- **Civil liberty:** Profiling is one of the potential applications of the big data that needs careful evaluation.
- **Equality:** The issue of equality concerns the treatment, in the public sector, of individuals and groups who do not fully participate in the information society, because they do not have the resources, the means, time or knowledge to do so.

"Why is it so difficult to mobilize cross-agency collaboration?"

- *Different organizational interests*
- *Different organizational cultures*
- *Lack of consensus on what is to be done*
- *"Transaction costs" – communication, travel*
- *Competing priorities*
- *Centralization within organizations, slow decision-making processes, multiple veto points*

"The Culture of Evidence-based Policy is Essential for Big Data Projects."

Dynamics of Evidence-based Policy



Source: Adapted from Gray (1997)

Shortcomings in Data Governance

- Much of the data collected and stored in an agency's transaction processing systems lacks adequate integrity.
- Data governance is difficult when it is localized to individual systems and departments within an agency.
- When one considers aggregating and connecting data across agencies, there is limited guidance in terms of policy and legal frameworks.
- When it comes to data governance, agencies are already resource-constrained and thus do not have the "bandwidth" to invest in building better governance processes .

Assessing Big Data Readiness

- Organizational Alignment: How the big data uses are aligned with the organization's current structure, its main activities and its strategy
- Organizational Maturity: The Maturity of e-government initiatives within the organization
- Organizational Capabilities: Whether the organization possesses the capacities to use big data, to create value from it, and to ensure that no negative consequences arose from big data uses.

"Big Data is a Leadership Problem—All Leaders Need a Working Knowledge of Data Analytics."

- *Many managers do not know enough about data science to be able to effectively make use of the data scientists in their organization*
- *Managers need not become experts in analytics technology, predictive analytics, or experimentation themselves. But they do need to be able to ask relevant questions and have a sense*
- *Managers need to know whether data are providing actionable insight, or simply providing the illusion of actionable insight.*

Chief Data Officer (CDO)

- The Chief Data Officer is a leader who creates and executes data and analytics strategies to drive value
- The role is responsible for defining, developing and implementing the strategy and methods by which the organization acquires, manages, analyzes and governs data
- It also carries the strategic responsibility to drive the identification of new opportunities through more effective and creative use of data

QUESTION 1

Hon. Celeste Balatbat (Chief of Misssion II, DFA Foreign Service Institute)

If we want to capacitate our people on big data, where do we start? Data are unreliable. How do we start to sift on how to sort unreliable and reliable data? What if the opinion that is coming out is skewed? Only reflective?

ANSWER: One way to address the need to capacitate is to start with a training of key people who will be handling big data. It is also better, in using and sifting through data, to look at it from the perspective of individual projects to make the data more specific.

Data also have its shortcomings. What can be done is triangulation of data of different types and sources, so that we do not just rely on one piece or source of data and so we can minimize the bias from interpreting data. The use of a combination of multiple large data sets can help fill in the shortcomings of using data. As Dr. Xun Wu said, "As long as you do not rely exclusively on one specific type of data, you should be fine."

There is, however, more checking that needs to be done when looking at data, especially given the nuances that might not be obvious when looking at it. We must also look at how representative data are. We must not take for granted that big data have nuances and details, too, and we must ask questions such as who is involved, where the data came from, etc.

QUESTION 2

Prof. Karl Jondoc (Asst. Professor, UP School of Economics)

Policy research is a chain (data generation -> analysis -> use). Each part of the chain has gaps. What specific interventions should happen at each part of the chain?

ANSWER: There is really a need for more people who can do data analytics in projects that involve big data. There will need to be a division of labor per department in handling data for aspects of this chain. There will also be a need to get people from these departments to work with big data together.

"As long as you do not rely exclusively on one specific type of data, you should be fine."

Open Forum

QUESTION 3

Dr. Ebinezer R. Florano (Director, CPED)

What are the techniques that can be used to handle big data more meaningfully?

ANSWER: Some of the techniques we need in studying and making sense of big data have actually been provided for us by prior trainings and methods we have learned before.

QUESTION 4

Marlon Marquina (Director IV, Information Technology Audit Office, Commission on Audit)

We've tried to push this big data 3 years ago. There is a lack of trust in the government based on experience. What are tips on how to overcome this challenge?

ANSWER: There is a very big tendency for those who hold data not to share what they have. The push for big data will only be successful if national government opens data. If there is no strong mandate from the government regarding big data, sharing will not happen voluntarily.

> *"The push for big data will only be successful if national government opens data."*

Synthesis **Dr. Noriel Christopher Tiglao**

In response to the lecture, Dr. Tiglao cites the following realizations on the science and application of big data :

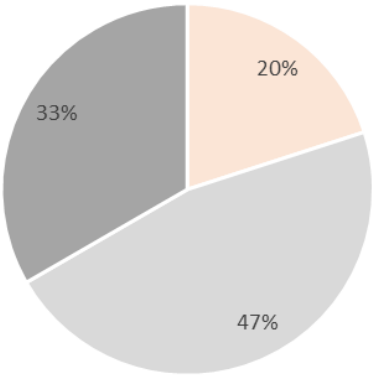
- Culture on the use of existing data
- Who drives the Big Data
- Need to make organizations information centric
- Need to use structured and unstructured data
- Murphy's Law – positive side, don't overestimate
- Big Data requires inter-agency coordination
- Opinion-based vs Evidence-based policy
- Lack of appropriate job description
- Leadership and working knowledge of managers
- It should be demand-driven

Feedback

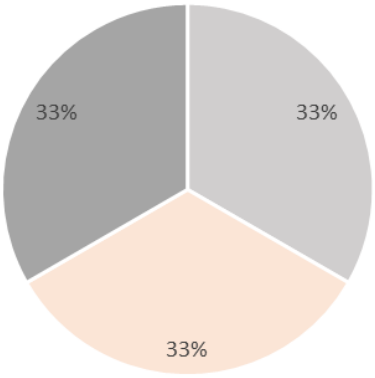
The respondent-attendees rated the lecture on promptness of technical advise, quality (satisfaction with advise received), timeliness, and relevance.

Good | Better | Best

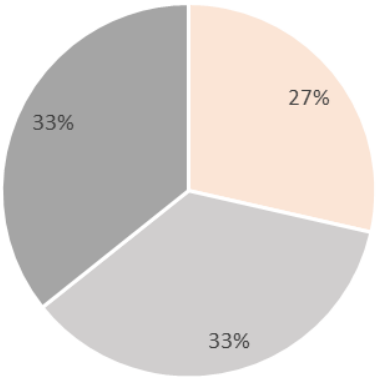
Promptness of technical advise



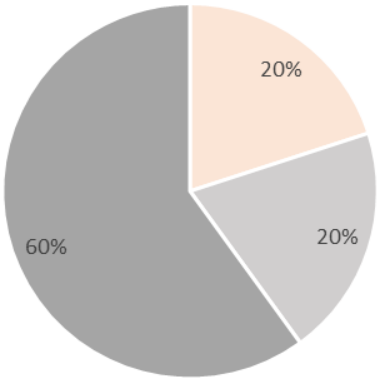
Quality



Timeliness



Relevance



BUILDING CAPACITY FOR BIG DATA APPLICATIONS IN GOVERNMENT

organized by

Center for Policy and Executive Development

University of the Philippines 1908
CPED
NCPAG 1952

text
Ric Arvin Agapay | Adrienne Onday | Jillian Jocelyn Somera

photos
Crinezza Veil Mendoza