Big Data in Public Affairs Education

Ines Mergel
Syracuse University

ABSTRACT
Public affairs schools face the challenge of including emergent topics in their curricula to prepare students for the public sector job market. Some such topics reflect advances in the use of information technologies; others reflect updates to industry standards or changing needs of public sector information management professionals. This article focuses on big data that are created through citizens' use of new technologies and the combination of administratively collected data with online data. Big data require changes in government information management skills, including collection, cleaning, and interpreting unstructured and unfiltered data; real-time decision making based on early signals and patterns that emerge; and new organizational roles and tasks, such as open innovation and change management. This article reviews the existing literature, compares big data requirements in neighboring disciplines, and suggests 13 modules for a big data syllabus that extend Mason's PAPA model of ethical considerations for the information age.

KEYWORDS
Information and technology management, big data, public management, MPA curriculum

Traditionally, Master of Public Administration (MPA) programs at public affairs schools teach government technology classes using either a project implementation approach or an information science approach. Many MPA programs regularize project and policy implementation classes by adding skills and knowledge training that is no longer restricted to technology projects. At the same time, public managers are less likely to design databases or set up servers as part of their job requirements. Such so-called information technology (IT) projects are often outsourced to contractors, sometimes with negative consequences for implementation and internal capacity building; the HealthCare.gov launch failure is but one well-known example (Overby, 2013). Parallel to these developments, a new wave of government information is emerging that poses challenges for public managers: so-called big data sets created through online interactions of citizens and government entities. These big data sets are challenging the traditional notion of administratively or scientifically created data sets, such as government open data posted on Data.gov or standardized surveys (e.g., the U.S. census, unemployment reviews).

These developments are problematic for MPA programs. A generalist education in MPA programs will not compete with computer science or engineering programs that are training advanced data scientists. MPA programs should therefore focus on the managerial aspects of innovative data initiatives and redesign their curricula in response to IT innovations in the public sector. Flexible electives can allow MPA programs to quickly offer new training in IT skill building, analytical reviews, and applied cases of emergent government information management.

THE EMERGENCE OF BIG DATA IN THE PUBLIC SECTOR

Big data is an industry term commonly said to derive from a McKinsey Global Institute (2011) report that outlined the emergence of very large data sets that need to be collected and analyzed differently than traditional data, using computational means and algorithms. This report described big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (p. 1). However, the term was first mentioned by two NASA scientists in 1997 (Cox & Elsworth, 1997, cited in Friedman, 2012). Cox and Elsworth described the problem they had with visualizing big data, explaining that big data

> provides an interesting challenge for computer systems: data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of big data. When data sets do not fit in main memory (in core), or when they do not fit even on local disk, the most common solution is to acquire more resources. (Cox & Elsworth, 1997, p. 235)

Gartner's IT glossary provides another definition, focusing on the amount, speed of creation, and unstructured nature of the data itself: "Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (Gartner, n.d.). Management scholars such as Davenport, Barth, and Bean (2012) picked up the topic and defined big data as "the broad range of new and massive data types that have appeared over the last decade or so," predicting a short-term shelf life for the term. Other authors have focused on the distributed creation and use of data sets with the help of new Internet technologies. For example, Cukier (2010) wrote in the *Economist* that "the world contains an unimaginably vast amount of digital information which is getting ever vaster ever more rapidly." Finally, Cukier, and Mayer-Schoenberger (2013) define the use of big data as "the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value" and "things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value."

BIG DATA DEFINITIONS IN NEIGHBORING DISCIPLINES

Beyond the above industry or business definitions, researchers in the social sciences only recently began wrestling with the term *big data*. Scholars in various disciplines have begun calling for additional research to understand how existing algorithms apply to big data sets.

For example, in the field of management studies, George, Haas, and Pentland (2014) focus on the diversity of new data sources and the emergence of new technologies that are either actively or passively creating and submitting data: "Big data is generated from an increasing plurality of sources, including Internet clicks, mobile transactions, user-generated content, and social media as well as purposefully generated content through sensor networks and business transactions such as sales queries and purchase transactions" (p. 321). In political science, Clark and Golder (2014) remark that big data in the form of "technological innovations such as machine learning have allowed researchers to gather either new types of data, such as social media data, or vast quantities of traditional data with less expense" (p. 65). Further, the authors note, technological innovations' "increasing ability to produce, collect, store, and analyze vast amounts of data is going to transform our understanding of the political world" (p. 65).

In the public policy field, Pirog (2014) sees the value of big data mostly in the availability of new data sets, such as the open data posted by the federal government on Data.gov. Value in her mind is created through the combination of traditional open data with geospatial data such as real-time satellite data, GPS locations of cell phones, economic transactions, or Internet search data. Researchers are left with the task of organizing, cleaning, and interpreting the data.

Lazer et al. (2009) focus on real-time availability and see these big data sets as opportunities to create a "second-by-second picture of interactions over extended periods of time, providing information about both the structure and content of relationships" (p. 2).

In summary, the social science articles reviewed here focus on data collection, analytical techniques, and the need for adaptation of existing research methods and theory building based on the potential insights generated by big data. Research communities are still struggling with definitions and agreed-upon methods and outcomes ("Editorial: Community Cleverness Required," 2008). The tasks associated with big data require advanced data analytics and computational power that are usually not located in traditional social science programs but rather in computer science or engineering disciplines.

CHARACTERISTICS OF BIG DATA

Big data sets are created in several forms: messages, updates, and images posted to social networks; readings from sensors worn by human beings, such as fitness sensors measuring distance, calories burned, time/duration, et cetera; GPS signals from cell phones and other wearables; online shopping transactions and preferences; and data collected passively from buildings, public transit vehicles, or video cameras installed on street corners. Such data can be combined with existing administratively collected data; for example, 911 data (emergency calls), 311 data (nonemergency reports), government performance data, or open data in government.

One agreed-upon way to define big data is to describe the characteristics involved. McAffee and Brynjolfsson (2012), for example, describe three *Vs* of big data: volume, velocity, and variety. Other scholars add the characteristics of veracity, variability, and visualization.

Volume refers to the scale of data that can no longer measured in megabytes but that might still be processable by social scientists. In big data, volume is expressed in terabytes and petabytes, for which large server capacity is necessary (McKinsey Global Institute, 2011).

Velocity refers to the real-time analysis of streaming data: moving from periodic data collection such as U.S. Census surveys to near/real-time collection and analysis of almost-live data. This can include cell phone data from phone logs and social media stream data from Twitter or Facebook feeds or Foursquare check-ins. A specific example is the U.S. Geological Survey's "Did you feel it?" Internet intensity maps that combine scientifically collected earthquake data with expressed impacts posted on Twitter (Atkinson & Wald, 2007).

Variety refers to the different forms of data that are no longer nicely structured in a database. Instead, data formats vary, including photo, audio, Web, social media, video, geospatial, and mobile, all of which are highly unstructured (Neumann, Park, & Panek, 2012).

Veracity refers to highly fuzzy and unstructured data that create uncertainty surrounding the data and their quality. For example, during an incident, emergency managers must now triangulate between multiple incoming 911 phone from cell phones that might be passing a scene, changing their geo-location, and providing varying degrees of quality in their reporting. In addition, observers at the scene may be posting what they see to social media sites like Twitter and Facebook, but it takes a while for emergency responders to report formally trusted information about the incident. These different types of data make it difficult for public managers to organize incoming data and use them as a basis for decision making.

Variability is a characteristic of big data that is similarly challenging to public managers. The meaning of data can change quickly, so that analysis and interpretation needs to be continuously updated to reflect the most recent changes. For example, consider the current searches people are conducting on Google. At one point in time, a search for a terrorist organization might have implicated the person searching as a sympathizer. However, now that terror organizations are in the news 24-7, such searches might be primarily for informational purposes and do not necessarily indicate a desire to affiliate with the terrorist group. As another example, Onnela et al. (2007) inferred friendship networks and affiliations from mobile phone data. The challenge for big data analysts is to review the context and content in order to understand nuances, such as sentiment or sarcasm.

Finally, *visualization* vis-à-vis big data can help in presenting the information understandably and readably to human beings, such that people can identify patterns or landscapes in the data. The pure volume of mentions of a brand, organization, or incident on social media is not necessarily a reliable indicator of a thing's viral nature or impact. Instead, one must measure how the number of mentions leads to value for an organization. The value of big data is one of the most difficult points to measure. In a recent McKinsey and Company report, Kayyali, Knott, and Van Kuiken (2013) estimated that especially in the public health sector, big data will create economic value by combining existing data sets in real time, interpreting the data faster, intervening in real time, and creating innovative IT health market segments.

BIG DATA IN PUBLIC AFFAIRS EDUCATION

Technological advances in Internet technologies and a shift from off-line to online interactions among citizens, as well as between citizens and government, have led to the creation of new types of data sets. In the public sector, there is a noticeable push from the White House Office of Science and Technology Policy and the National Science Foundation to encourage agencies to review their need to use big data and for researchers to design projects and deliver innovative insights concerning big data (Mervis, 2012). Currently, data types collected by government agencies include mostly administratively data, scientifically measured and collected data, and most recently—but to a far lesser extent—social media or Internet data. In combination, these different data types fall under the umbrella of big data. New to government information or government data are those data sets created through the active use of Internet technologies, such as clickstream data from the Web, social media content (tweets, blogs, Facebook posts, etc.), and video data from retail and other settings and from video entertainment (Davenport, Barth, & Bean, 2012). For intelligence purposes, researchers and government organizations have been analyzing terrorist networks using a combination of online accounts such as e-mail exchanges, cell phone data for calls and text messages, and bank transactions to derive affiliation, plans, or leanings (such as recent investigations into recruits for Syrian terror organizations based on social media interactions, online searches, etc.) (Asal & Rethemeyer, 2015). However, as Desouza (2014) shows, government agencies see the need to invest in big data skills but have not made progress in institutionalizing capacity building or incorporating such skills into existing standard operating procedures.

In a recent article, Mergel, Rethemeyer, and Isett (2015) outlined the three main skills that public administration students need to acquire to productively work with big data:

1. managing and processing large accumulations of unstructured and semistructured data;
2. analyzing that data into meaningful insights for public operations; and
3. interpreting that data in ways that support evidence-based decision making.

A review of handbooks and course listings from the *U.S. News and World Report* (2012) top 10 ranked public affairs schools that offer a concentration in information and technology management reveals that MPA programs have not yet incorporated big data topics in their curricula. Instead, as shown in Table 1, all 10 ranked programs rely on traditional e-government and information management classes and even then omit most recent topics, such as social media and new technologies.

The MPA programs listed in Table 1 usually offer both public policy and information management classes but rely on neighboring disciplines to provide advanced informatics classes, such as system design, telecommunication, GIS analysis, and even government information management. These neighboring disciplines include political science, geography, and computer science departments located in

TABLE 1.
Overview of Information and Technology Management Concentrations in the Top 10 Public Affairs Schools' MPA Programs

Rank	School and program name	Concentration name	Courses
1	Carnegie Mellon University, Heinz College	Public Policy and Management, no information management track. Instead, offers a Master in Information Systems.	Information Systems for Managers Introduction to Geographic Information Systems Multi-Media
2	Syracuse University, Maxwell School, Department of Public Administration and International Affairs	Technology and Information Management study program	Digital Innovation Management Digital Government Social Media Science, Technology, and Public Policy
3	University at Albany–SUNY, Rockefeller College of Public Affairs and Policy	Information Technology and Management concentration	Data, Models, and Decisions Topics in Information Resource Management Technology in Financial Market Regulation Cybersecurity: Long-Term Planning and Risk Management Foundations of Government Information Strategy and Management Information Technology, Globalization, and Global Governance Management Information Systems Building a Case for IT Investments in the Public Sector
4	Rutgers University, Newark	Core Curriculum Cluster One - Foundation	Technology and Public Administration
5	Georgia Institute of Technology	Graduate Certificate in Science, Technology and Society	Science, Technology, and the Economy Science, Technology, and Security Science and Technology Beyond Borders
6	Indiana University, Bloomington	Information Systems concentration	Vector-Based Geographic Information Systems Public Management Information Systems Database Management Systems
6	University of Nebraska, Omaha	Public Administration/ Management Information Systems	Management of Software Development Advanced Systems Analysis and Design Data Communications Managing the Distributed Computing Environment Database Management Managing the IS Function
8	Harvard University, Harvard Kennedy School	No specific technology or information management concentration	Technology and Policy Privacy, Technology, and National Security Technology, Security, and Conflict in the Cyber Age Human Rights Advocacy Using Video, Social Media, and Participatory Media
8	University of Texas at Austin, Lyndon B. Johnson School of Public Affairs	Technology, Innovation, and Information Policy	Information Policy Intro to Geographic Info Systems Technological Innovation in Defense

Notes. Based on *U.S. News and World Report* (2012) ranking of public affairs programs. Rankings 6 and 8 reflect ties among four different schools.

iSchools and engineering or business schools. All schools in Table 1 offer advanced data analysis skills, such as the Georgia Institute of Technology's science and technology classes and Indiana University's Public Management Information Systems class.

This reliance on outside disciplines is not surprising given traditional MPA programs' focus on core public administration classes and the relatively recent emergence of topics such as big data, the Internet of things, and the smart city, which are mostly driven by industry. An interesting new program at Northeastern University has begun to address the challenges of big data: the Network Science Institute (networkscienceinstitute.org) includes needed theoretical and methodological network analysis methods as well as computational social science skills.

Given the apparent lack of coverage in the MPA curriculum and the concurrent increased need for big data skills in government (Helms, 2015; Mervis, 2012), I propose that five dimensions of the topic should be covered in a management-oriented big data class to prepare future public managers for some of the pressing problem areas in the public sector. These dimensions start from Mason's (1986, p. 5) so-called PAPA model, which distinguishes four ethical considerations of information access:

1. **Privacy:** What information about one's self or one's associations must a person reveal to others, under what conditions and with what safeguards? What things can people keep to themselves and not be forced to reveal?
2. **Accuracy:** Who is responsible for the authenticity, fidelity and accuracy of information? Similarly, who is to be held accountable for errors in information and how is the injured party to be made whole?
3. **Property:** Who owns the information? What are the just and fair prices of exchanges? Who owns the channels, especially the airways, through which information is transmitted? How should access to this scarce resource be regulated?
4. **Accessibility:** What information does a person or an organization have a right or a privilege to obtain, under what conditions and with what safeguards?

Mason's PAPA model considers issues that mostly have to do with the information attributes themselves. While this remains a concern, today's big data world calls for training public managers in additional dimensions that focus on managerial challenges. Thus I propose expanding Mason's model to these five critical dimensions that public managers must consider when dealing with big data: (1) the ethical dimension, (2) the technological dimension, (3) the process dimension, (4) the organizational and institutional change dimension, and (5) the analytical dimension.

1. Ethical Dimension

Big data proponents in industry, consulting companies, and business schools use dramatic terms such as *big data revolution* (Iansiti & Lakhani, 2014; McAffee & Brynjolfsson, 2012), predicting big data's inevitable and mostly beneficial impact on society, the economy, or public health (Kayyali, Knott, & Van Kuiken, 2013; McKinsey Global Institute, 2011). This bias toward the advantageous impact of big data usually omits a critical evaluation of the ethical dimensions of data collection, analysis, and reuse, especially that by the government done without explicit citizen consent (boyd & Crawford, 2012). The NSA's data collection method ("Sniff it all, collect it all") is but one example of mass data collection that interferes with existing privacy laws or the U.S. Constitution. Algorithms that predict interactions or movements can be useful in identifying terrorists but can also mistakenly target peaceful demonstrators or other innocent people (Eagle, Pentland, & Lazer, 2009; Onnela et al., 2007). After the Boston Marathon bombing, people using the social networking site Reddit identified the wrong person as the bomber, when in fact he had committed suicide and had thus dropped off the site for that reason, not because he was involved in the bombing (Wade, 2014).

Other ethical considerations involve the use of public health data collected via fitness apps or sensors such as Fitbit and other wearables, or that derived from the genetic sequencing of individual DNA (Murdoch & Detsky, 2013). These health-related data can—in combination with online purchasing behavior, medical transactional data, social networking interactions, and social security numbers—provide insights into potential health risks for segments of the population or even individuals. Health insurance companies could potentially draw conclusions from such data that could lead to significant discrimination against individuals in terms of insurancecoverage or cost.

At the same time, the individual health data collected and published on social networking sites can also create value in the form of government public health interventions when such individual data is combined with geotagged air pollution data, for instance, or mobility data. If citizens can see what value their data provides, they may be more likely to volunteer personal information, which could help government evaluate the impact of regulations or assess the need for interventions while tracking infectious diseases such as Ebola or flu (Cohen et al., 2012; Morse, 2012; Murdoch & Detsky, 2013).

2. Technological Dimension

The technological dimension of big data involves emergent topics that are challenging the investment decisions of chief technology officers in government. Examples include the Internet of things, the smart city, and cloud computing, all of which must be explored for their usefulness and applicability in the public sector. The Internet of things refers to the online transmission of information through sensors from buildings, cars, human beings, or animals to machines that store the information in the cloud. Even though human beings are actively involved in wearing data-collection devices, the Internet of things mostly involves machines communicating with other machines. For example, homes can automatically transmit information to utility companies via temperature sensors or smart meters; smart cement can transmit information about contamination levels in oil wells or cracks in bridges; and cars equipped with sensors can transmit the weather and road conditions (Burrus, 2014). Such data, either transmitted actively by citizens or collected passively via interactions with public infrastructure, can help public managers make investment decisions, deploy resources in real time, conduct urban analysis and modeling of public mobility, and decide whether or not to intervene manually.

The use of new technologies, and especially the combination of data collected from different types of sources (social data, the Internet of things), contributes to the notion of the smart city (Bingham-Hall & Law, 2015; O'Grady & O'Hare, 2012). Online interactions with citizens and their environment can improve the quality and performance of a city's services, reduce energy costs, and improve distribution of resources.

3. Process Dimension

The process dimension of big data involves incorporating citizen feedback. New technologies, such as crowdsourcing platforms, as well as political changes and presidential mandates to engage directly with citizens have changed how citizens interact online with government organizations (Miller & Kalil, 2014). Already tested and proven processes such as participatory budgeting have created cultural and procedural acceptance of incorporating large-scale feedback from citizens in government decision making (Novy & Leubolt, 2005). Platforms such as the open innovation tool Challenge.gov have also shown that there is, on the one hand, public enthusiasm for such tools but, on the other, slack capacity vis-à-vis contributing to government innovation.

Brabham (2008) defined crowdsourcing as "a new web-based business model that harnesses the creative solutions of a distributed network of individuals through what amounts to an open call for proposals" (p. 76). Citizens are helping to write policy (Prpić; Taeihagh, & Melton, 2015) and helping government agencies solve problems that agency employees could not.

Especially organizations such as NASA have gained experience using citizen insights to further research projects (Raddick et al., 2010).

NASA's online citizen science platform (science.nasa.gov/citizen-scientists) engages citizens in microtasks such as the identification of small objects on photos or videos. These are usually tasks that machines or algorithms cannot do, instead needing human logic for identification and interpretation.

4. Organizational and Institutional Change Dimension

The organizational and institutional change dimension involves government building its capacity to accommodate advances in big data. Current practices in public administration reflect the use of traditional administratively collected data to monitor performance of government operations, for example, through models like CitiStat. Some officials misinterpret big data as only that data supplied by citizens to government; for example, the description given by the governor of Maryland, Martin O'Malley, of using 911 and 311 data (O'Malley, 2014). Both of these data sets are created by using citizen input via traditional databases and administrative processes; the resulting information is then used to derive real-time insights into emergency and nonemergency situations.

However, what is needed beyond this is increased organizational capacity. As Court (2015) states, we need the "creation of new senior-management capacity to really focus on data and address cultural and skill-building challenges needed for the front line to embrace the change [i.e., organizational change]." Public managers need to understand how to include big data insights into decision making and resource deployment, similar to stock market analysis. This type of organizational capacity building needs to go beyond the IT department and include change managers as well as data scientists in the redesign of processes and systems, so that such systems reflect the changing needs of stakeholders.

Organizational and cultural change challenges will continue to occur as the nature of decision making changes with advances in big data. Data are no longer merely for archiving. Instead of only protecting and safeguarding data, as chief data officers had to do in the past, policy makers and public managers must use data to make better-informed, real-time decisions (Brownson, Fielding, & Maylahn, 2009). The role of the chief data officer will therefore likely change to being a chief change manager or chief innovation officer who must think creatively about how to make data available instead how to archive data.

5. Analytical Dimension

The analytical dimension of big data involves the ability to create predictive insights before an event occurs or to interpret inflowing data in real time. The hope is that the data amount and data types might lead to more accurate predictions, better decisions, and potentially immediate interventions.

Many government agencies are already using dashboards to review, for example, 311 nonemergency reports provided by citizens. However, the insights derived and how public managers act on the data vary. For example, though the Federal IT Dashboard (itdashboard.gov) brings some transparency to government IT investment, actual changes in behavior and buying decisions are not yet observable.

McAffee and Brynjolfsson (2012) give the example of airlines' determining estimated arrival times as useful analysis of big data; airlines look at publicly available data, such as weather forecasts, along with flight schedules, radar station input, subjective pilot estimates, and other internal information to improve deployment of resources, facilitate smoother operations, and reduce flight delays. Lazer, Kennedy, King, and Vespignani (2014), however, warn that human interpretation must be involved, citing the example of Google Flu Trends, which amassed searches of key flu terms by Google users and published the results; computer algorithms can fail in the absence of human interpretation of data, leading to misinterpretation.

The analytical dimension for public managers, therefore, does not necessarily mean a focus on algorithms, which data scientists apply. Instead, this dimension must focus on human interpretation and decisions based on new information provided through big data.

A SYLLABUS FOR TEACHING BIG DATA IN THE PUBLIC SECTOR

To expand public management skills to include the above five dimensions vis-à-vis big data, I propose the following modules for an MPA-level course titled Big Data in the Public Sector.

FIGURE 1.
Proposed Course Syllabus: Big Data in the Public Sector

Proposed Course Syllabus: Big Data in the Public Sector

Module 1: Big Data: Introduction, Concepts, Definitions
The public affairs context

Readings

Ahonen, P. (2015, July). Institutionalizing big data methods in social and political research. *Big Data and Society*. doi:10.1177/2053951715591224.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., et al. (2009). Computational social science. *Science*, *323*(5915), 721–723.

Pentland, A. S., & Berinato, S. (2014, November). With big data comes big responsibility. *Harvard Business Review*. Retrieved from https://hbr.org/2014/11/with-big-data-comes-big-responsibility.

Module 2: Big Data's Volume, Velocity, Variety, and Veracity
Characteristics of big data and data sources

Readings

Court, D. (2015, January). Getting big impact from big data. *McKinsey Quarterly*. Retrieved from http://www.mckinsey.com/insights/business_technology/getting_big_impact_from_big_data.

Cukier, K. N. (2010, February 25). Data, data everywhere. *Economist*. Retrieved from http://www.economist.com/node/15557443.

Davenport, T. H., Barth, P., & Bean, R. (2012). How "big data" is different. *MIT Sloan Management Review*, *54*(1), 43–46. Retrieved from http://sloanreview.mit.edu/article/how-big-data-is-different.

Gray, E., Jennings, W., Farrall, S., & Hay, C. (2015, June). Small big data: Using multiple data-sets to explore unfolding social and economic change. *Big Data and Society*. doi:10.1177/2053951715589418.

Case

Su, N., & Nagaash, Pr. (2014). *Volkswagen Group: Driving big business with big data.* London, ON: Ivey.

FIGURE 1.
Proposed Course Syllabus: Big Data in the Public Sector (continued)

Module 3: Big Data Ethics
Ethical considerations in the interpretation of personally identifiable information (PII), discrimination, tailoring services and rates, underage users' and minors' using social media and Internet

Readings

Basic HHS policy for protection of human research subjects, 45 C.F.R. 46 (2009).

boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication and Society*, *15*(5), 662–679.

Duhigg, C. (2012, February 16). How companies learn your secrets. *New York Times Magazine*.

Gorman, S., Perez, E., & Hook, J. (2013, June 7). U.S. collects vast data trove. *Wall Street Journal*. Retrieved from http://www.wsj.com/articles/SB100014241278873242991045 78529112289298922.

Hargittai, E., & Hsieh, Y. P. (2013). Digital inequality. In W. H. Dutton (Ed.), *Oxford Handbook of Internet Studies* (pp. 129–150). Oxford, United Kingdom: Oxford University Press.

McCallister, E., Grance, T., & Scarfone, K. (2010). *Guide to protecting the confidentiality of personally identifiable information (PII): Recommendations of the National Institute of Standards and Technology* (NIST Special Publication 800-122). Gaithersburg, MD: NIST. Retrieved from http://csrc.nist.gov/publications/nistpubs/800-122/sp800-122.pdf.

Schneier, B. (2013, April 30). Do you want the government buying your data from corporations. *Atlantic*. Retrieved from http://www.theatlantic.com/technology/archive/2013/04/do-you-want-the-government-buying-your-data-from-corporations/275431.

Zwitter, A. (2014, November). Big data ethics. *Big Data and Society*. doi:10.1177/2053951714559253.

Cases

Acquisri, A., & Gross, R. (2009). Predicting Social Security Number from public data. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(27), 10975–10980.

Dannen, C. (2009, July 1). On Facebook? New algorithm can guess your SSN. *Fast Company*. Retrieved from http://www.fastcompany.com/1305136/facebook-new-algorithm-can-guess-your-ssn.

Hogan, M. (2015, July). Data flows and water woes: The Utah Data Center. *Big Data and Society*. doi:10.1177/2053951715592429.

NSA Domestic Surveillance Directorate. (n.d.). Your data: If you have nothing to hide, you have nothing to fear. Retrieved from https://nsa.gov1.info/data.

FIGURE 1.
Proposed Course Syllabus: Big Data in the Public Sector (continued)

Module 4: Value Proposition of Big Data
Supporting national priorities and economic development opportunities

Readings

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011, Winter). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 21–32.

McKinsey Global Institute. (2011). *Big data: The next frontier of innovation, competition, and productivity*. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Shmueli, E., Mazeh, I., Radaelli, L., Pentland, A. S., & Altshuler, Y. (2015). Ride sharing: A network perspective. In N. Agarwal, K. Xu, & N. Osgood (Eds.), *Social computing, behavioral-cultural modeling, and prediction* (pp. 434–439). Berlin, Germany: Springer International.

Case

Parker, S., & Ramasastry, C. S. (2014). *Luminar: Leveraging big data using corporate entrepreneurship* (W14175-PDF-ENG). Brighton, MA: Harvard Business Publishing.

Module 5: The Technological Dimension
Cloud computing, sensors, machine-to-machine communication

Readings

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., et al. (2010). A view of cloud computing. *Communications of the ACM*, *53*(4), 50–58.

Iansiti, M., & Lakhani, K. (2014, November). Digital ubiquity: How connections, sensors, and data are revolutionizing business. *Harvard Business Review*, *92*(11), 91–99.

Kaufman, L. M. (2009, July/August). Data security in the world of cloud computing. *Security and Privacy, IEEE*, *7*(4), 61, 64.

Onnela, J. P., Waber, B. N., Pentland, A., Schnorf, S., & Lazer, D. (2014). Using sociometers to quantify social interaction patterns. *Scientific Reports*, *4*(5604). doi:10.1038/srep05604.

Paquette, S., Jaeger, P. T., & Wilson, S. C. (2010). Identifying the security risks associated with governmental use of cloud computing. *Government Information Quarterly*, *27*(3), 245–253.

Case

Kanter, R. M. (2014). *The Weather Company*. Harvard Business School Case 314-083. Retrieved from http://www.hbs.edu/faculty/Pages/item.aspx?num=46052.

FIGURE 1.
Proposed Course Syllabus: Big Data in the Public Sector (continued)

Module 6: Predictive Analytics
Potential and current use of big data insights

Readings

Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., & Pentland. A. (2014). Once upon a crime: Towards crime prediction from demographics and mobile data. In *Proceedings of the 16th International Conference on Multimodal Interaction, Istanbul, Turkey, November 12–16, 2014* (pp. 427–434). New York, NY: Association for Computing Machinery.

Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *Economic Record*, *88*(s1), 2–9.

City of Reno dashboard: http://dashboard.reno.gov.

City of Springfield urban planning dashboard: http://gallery.idashboards.com/preview/?guestuser=webgov&dashID=226.

Davenport, T. H. (2013, December). Analytics 3.0. *Harvard Business Review*. Retrieved from https://hbr.org/2013/12/analytics-30.

Ganapati, S. (2011). Using dashboards in government. Retrieved from IBM Center for the Business of Government website: http://www.businessofgovernment.org/report/use-dashboards-government.

Gov.UK performance data dashboard: http://www.gov.uk/performance.

Case

Khachatryan, D. (2014). *Jennie Maze Limited: Enhancing call center performance using predictive analytics* (Babson College Case BAB263-HCB-ENG). Brighton, MA: Harvard Business Publishing.

Module 7: The Process Dimension
Crowdsourcing citizen insights to create innovation for government

Readings

Brabham, D. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, *14*(1), 75–90.

Howe, J. (2006, June 1). The rise of crowdsourcing. *Wired*. Retrieved from http://www.wired.com/2006/06/crowds.

Cases

Boudreau, K. J., Gaule, P., Lakhani, K. R., Riedl, C., & Woolley, A. W. (2014). *From crowds to collaborators: Initiating effort and catalyzing interactions among online creative workers* (Harvard Business School working paper). Retrieved from http://nrs.harvard.edu/urn-3:HUL.InstRepos:12111352.

NASA's citizen scientists website: http://science.nasa.gov/citizen-scientists.

FIGURE 1.
Proposed Course Syllabus: Big Data in the Public Sector (continued)

Module 7: The Process Dimension (continued)

Tushman, M. L., Lifshitz-Assaf, H., & Herman, K. *Houston, we have a problem: NASA and open innovation (A)* (Harvard Business School Case 414-044). Retrieved from http://www.hbs.edu/faculty/Pages/item.aspx?num=47334.

Module 8: Organizational Alignment and New Organizational Structures

- ***Expanding the C-suite in government: chief data scientists, chief innovation managers, et cetera***
- ***Embedding big data in the organizational culture***

Readings:

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*(2), 321–326.

Stone, A. (2015, March 4). Chief data officers: Shaping one of the newest positions in government. *Governing Technology*. http://www.govtech.com/data/Chief-Data-Officers-Shaping-One-of-the-Newest-Positions-in-Government.html.

Module 9: Smart Cities and the Internet of Things

Smart city concepts, functionalities, sensors, data collection

Readings

Bingham-Hall, J., & Law, S. (2015, August). Connected or informed? Local Twitter networking in a London neighbourhood. *Big Data and Society*. doi:10.1177/2053951715597457.

Gil-Garcia. R., Helbig, N., & Ojo, A. (2014, June). Being smart: Emerging technologies and innovation in the public sector. *Government Information Quarterly*, *31*(s1),. I1–I8.

Lee, I., & Lee, K. (2015). The Internet of things (IoT): Applications, investments, and challenges for enterprises. *Business Horizons*, *58*(4), 431–440.

O’Grady, M., & O’Hare, G. (2012). How smart is your city? *Science*, *335*(6076), 1581–1582.

Thrift, N. (2014, June). The “sentient” city and what it may portend. *Big Data and Society*. doi:10.1177/2053951714532241.

Wood, C. (2014, November 3). Smart cities: Understanding the untapped value of sensor data. *Government Technology*. Retrieved from http://www.govtech.com/data/Smart-Cities-Understanding-the-Untapped-Value-of-Sensor-Data.html.

Case

Ovchinnikov, A. (2013). *Global warming revisited (B)* (Darden School of Business Case UV6690-PDF-ENG). Brighton, MA: Harvard Business Publishing.

FIGURE 1.
Proposed Course Syllabus: Big Data in the Public Sector (continued)

Module 10: Big Data in Government
Open data, social media data, administratively collected data, scientific data

Readings

Baack, S. (2015, July). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data and Society*. doi:10.1177/2053951715594634.

Helms, J. (2015, February 25). Five examples of how federal agencies use big data. IBM Center for the Business of Government blog. Retrieved from http://www.businessofgovernment.org/blog/business-government/five-examples-how-federal-agencies-use-big-data.

Lyon, D. (2014, July). Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data and Society*. doi:10.1177/2053951714541861.

Mervis, J. (2012). Agencies rally to tackle big data. *Science*, *336*(6077), 22.

Pirog, M. A. (2014). Data will drive innovation in public policy and management research. *Journal of Policy Analysis and Management*, *33*(2), 537–543.

Case

Atkinson, G. M., & Wald, D. J. (2007). "Did you feel it?" Intensity data: A surprisingly good measure of earthquake ground motion. *Seismological Research Letters*, *78*(3), 362–368.

Module 11: Big Data in Politics
Political engagement, social movements, uncovering online networks, elections

Readings:

Asal, N. N., & Rethemeyer, R. K. (2015, February 26). Building terrorism from social ties: The dark side of social capital. *Civil Wars*, 1–23. doi:10.1080/13698249.2014.981942.

Clark, W. R., & Golder, M. (2014). Big data, causal inference, and formal theory: Contradictory trends in political science? *PS: Political Science and Politics*, *48*(1), 65–70.

Dumas, C. L., LaManna, D., Harrison, T. M., Ravi, S., Kotfila, C., Gervais, N., et al. (2015, August). Examining political mobilization of online communities through e-petitioning behavior in We the People. *Big Data and Society*. doi:10.1177/2053951715598170.

Eagle, N., Pentland, A., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(36), 15274–15278.

Sudhahar, S., Veltri, G. A., & Cristianini, N. (2015, March). Automated analysis of the US presidential elections using big data and network analysis. *Big Data and Society*. doi:10.1177/2053951715572916.

Case

Diesner, J., Frantz, T., & Carley, K. M. (2005). Communication networks from the Enron email corpus: "It's always about the people. Enron is no different." *Computational and Mathematical Organization Theory*, *11*(3), 201–228.

FIGURE 1.
Proposed Course Syllabus: Big Data in the Public Sector (continued)

Module 12: Big Data in Public Health

Predicting pandemic outbreaks, tracking infectious diseases, assessing health implications, using sensors, implementing mobile health applications

Readings

Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, *357*, 370–379.

Cohen, B., Ferng, Y. H., Wong-McLoughlin, J., Jia, H., Morse, S. S., & Larson, E. L. (2012). Predictors of flu vaccination among urban Hispanic children and adults. *Journal of Epidemiology and Community Health*, *66*(3), 204–209. doi:10.1136/jech.2009.099879.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., et al. (2009). Computational social science. *Science*, *323*(5915), 721.

Morse, S. S., Mazet, J. A., Woolhouse, M., Parrish, C. R., Carroll, D., Karesh, W. B., et al. (2012). Prediction and prevention of the next pandemic zoonosis. *Lancet*, *380*(9857), 1956–1965. doi:10.1016/S0140-6736(12)61684-5.

Morse, S. S. (2012). Public health surveillance and infectious disease detection. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, *10*(1), 6-16. doi:10.1089/bsp.2011.0088.

Olson, S. H., Reed, P., Cameron, K. N., Ssebide, B. J., Johnson, C. K., Morse, S. S., et al. (2012). Dead or alive: Animal sampling during Ebola hemorrhagic fever outbreaks in humans. *Emerging Health Threats Journal*, *5*. doi:10.3402/ehtj.v5i0.9134.

Module 13: Data Analysis Practicum

- ***Social network analysis for social scientists***
- ***Introduction into NodeXL for social media data collection and analysis***

Readings

Contractor, N., Wasserman, S., & Faust, K. (2006). Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example. Academy of Management Review, 31(3), 681–703.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137–144.

Gray, E., Jennings, W., Farrall, S., & Hay, C. (2015, June). Small big data: Using multiple data-sets to explore unfolding social and economic change. Big Data and Society. doi:10.1177/2053951715589418.

Schroeder, R. (2014, December). Big data and the brave new world of social media research. Big Data and Society. doi:10.1177/2053951714563194.

CONCLUSION

The suggested syllabus for the Big Data in Public Affairs class is derived from the needs of practitioners and the apparent lack of MPA classes that address comparable issues. The development of government information management or information technology management concentrations and course offerings seems to have stalled, and the MPA programs reviewed in this article (see Table 1) rely heavily on courses offered in neighboring disciplines, such as computer science, engineering, or business. This does not seem to be the best scenario for MPA programs, students, or future public managers. For example, tuition dollars from MPA students interested in technology management innovations in the public sector do not accrue to MPA programs. Also, given the nature of other disciplines, I predict that most of these outside classes will not provide contextual information or qualitative insights into the most recent developments observable in the public sector. At the same time, the fast-moving trend toward integrating new technologies in government, not only for external public affairs, but also for internal use of new data sets, requires that future public managers be schooled in big data.

The modules for the suggested Big Data in Public Affairs class provide a systematic review of the five management dimensions vis-à-vis big data that future public managers should be aware of, as well as a research methods practicum in which students collect, analyze, and interpret social media data. The integration of both managerial practices and analytical skills provides MPA students with insights into emergent big data concepts, which can be combined with traditional performance management or CitiStat classes. The modules focused on specific public affairs domains—such as government, public health, or politics—are designed to capture a variety of practical applications that classes in computer science or engineering programs might not address.

The integration of such a class into a generalist MPA curriculum is not meant to compete with existing public management or public policy classes that focus on the analysis and interpretation of administratively collected data. Instead, this new course is designed to augment existing data analysis classes that focus on clean data sets. Big data is messy and far from clean. Similarly, methods of big data analysis are still developing. Nevertheless, it is important to prepare MPA students for the new realities of the public affairs workplace and equip them for careers in organizational roles such as data scientists and chief innovation officers.

REFERENCES

Asal, N. N., & Rethemeyer, R. K. (2015, February 26). Building terrorism from social ties: The dark side of social capital. *Civil Wars*, 1–23. doi:10.1080/13698249.2014.981942.

Atkinson, G. M., & Wald, D. J. (2007). "Did you feel it?" Intensity data: A surprisingly good measure of earthquake ground motion. *Seismological Research Letters*, *78*, 362–368.

Bingham-Hall, J., & Law, S. (2015, August). Connected or informed? Local Twitter networking in a London neighbourhood. *Big Data and Society*. doi:10.1177/2053951715597457.

boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication and Society*, *15*(5), 662–679.

Brabham, D. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, *14*(1), 75–90.

Brownson, R. C., Fielding, J. E., & Maylahn, C. M. (2009). Evidence-based public health: A fundamental concept for public health practice. *Annual Review of Public Health*, *30*, 175–201. doi:10.1146/annurev.publhealth.031308.100134.

Burrus, D. (2014). The Internet of things is far bigger than anyone realizes. *Wired*. Retrieved from http://www.wired.com/insights/2014/11/the-internet-of-things-bigger.

Clark, W. R., & Golder, M. (2014). Big data, causal inference, and formal theory: Contradictory trends in political science? *PS: Political Science and Politics*, *48*(1), 65–70.

Cohen, B., Ferng, Y. H., Wong-McLoughlin, J., Jia, H., Morse, S. S., & Larson, E. L. (2012). Predictors of flu vaccination among urban Hispanic children and adults. *Journal of Epidemiology and Community Health*, *66*(3), 204–209. doi:10.1136/jech.2009.099879.

Court, D. (2015, January). Getting big impact from big data. *McKinsey Quarterly*. Retrieved from http://www.mckinsey.com/insights/business_technology/getting_big_impact_from_big_data.

Cox, M., & Elsworth, D. (1997, October 18–24). *Application-controlled demand paging for out-of-core visualization*. Paper presented at the 8th IEEE Visualization '97 Conference, Phoenix, AZ. Retrieved from http://www.evl.uic.edu/cavern/rg/20040525_renambot/Viz/parallel_volviz/paging_outofcore_viz97.pdf.

Cukier, K. N. (2010, February 25). Data, data everywhere. *Economist*. Retrieved from http://www.economist.com/node/15557443.

Cukier, K. N., & Mayer-Schoenberger, V. (2013, May/June). The rise of big data: How it's changing the way we think about the world. *Foreign Affairs*. Retrieved from http://www.foreignaffairs.com/articles/2013-04-03/rise-big-data.

Davenport, T. H., Barth, P., & Bean, R. (2012). How "big data" is different. *MIT Sloan Management Review*, *54*(1), 43–46. Retrieved from http://sloanreview.mit.edu/article/how-big-data-is-different.

Desouza, K. (2014). *Realizing the promise of big data*. Washington, DC: IBM Center for the Business of Government. Retrieved from http://www.businessofgovernment.org/sites/default/files/Realizing the Promise of Big Data.pdf.

Eagle, N., Pentland, A., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(36), 15274–15278.

Editorial: Community cleverness required. (2008). *Nature*, *455*(1). Retrieved from http://www.nature.com/nature/journal/v455/n7209/full/455001a.html.

Friedman, U. (2012, October 8). Big data: A short history. *Foreign Policy*. http://foreignpolicy.com/2012/10/08/big-data-a-short-history.

Gartner. (n.d.). Gartner IT glossary: Big data. Retrieved from http://www.gartner.com/it-glossary/big-data.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*(2), 321–326.

Helms, J. (2015). Five examples of how federal agencies use big data. Retrieved from IBM Center for the Business of Government website: http://www.businessofgovernment.org/blog/business-government/five-examples-how-federal-agencies-use-big-data.

Iansiti, M., & Lakhani, K. (2014). Digital ubiquity: How connections, sensors, and data are revolutionizing business. *Harvard Business Review*, *92*(11), 91–99.

Kayyali, R., Knott, D., & Van Kuiken, S. (2013). The big-data revolution in US health care: Accelerating value and innovation. Retrieved from McKinsey and Company website: http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: Traps in big data analysis. *Science*, *343*(6176), 1203–1205.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., et al. (2009). Computational social science. *Science*, *323*(5915), 721–723.

Mason, R. O. (1986, March). Four ethical issues of the information age. *MIS Quarterly*, 5–12.

McAffee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 60–68.

McKinsey Global Institute. (2011). *Big data: The next frontier of innovation, competition, and productivity*. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Mergel, I., Rethemeyer, R. K., & Isett, K. (2015). *Big data in public affairs*. Manuscript submitted for publication.

Mervis, J. (2012). Agencies rally to tackle big data. *Science*, *336*(6077), 22.

Miller, J., & Kalil, T. (2014). Crowdsourcing ideas to accelerate economic growth and prosperity through a strategy for American innovation. Retrieved from White House website: http://www.whitehouse.gov/blog/2014/07/28/crowdsourcing-ideas-accelerate-economic-growth-and-prosperity-through-strategy-ameri.

Morse, S. S. (2012). Public health surveillance and infectious disease detection. *Biosecurity and bioterrorism: biodefense strategy, practice, and science*, *10*(1), 6–16. doi:10.1089/bsp.2011.0088.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Journal of the American Medical Association*, *309*(13), 1351–1352.

Neumann, W. R., Park, Y. J., & Panek, E. (2012). Tracking the flow of information into the home: An empirical assessment of the digital revolution in the U.S. from 1960–2005. *International Journal of Communication*, *6*, 1022–1041.

Novy, A., & Leubolt, B. (2005). Participatory budgeting in Porto Alegre: Social innovation and the dialectical relationship of state and civil society. *Urban Studies*, *42*(22), 2023–2036.

O'Grady, M., & O'Hare, G. (2012). How smart is your city? *Science*, *335*(6076), 1581–1582.

O'Malley, M. (2014). Doing what works: Governing in the age of big data. *Public Administration Review*, *74*(5), 555–556.

Onnela, J., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., et al. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(18), 7332–7336.

Overby, S. (2013). 6 IT outsourcing lessons learned from HealthCare.gov's troubled launch. Retrieved from CIO website: http://www.cio.com/article/2381755/outsourcing/6-it-outsourcing-lessons-learned-from-healthcare-gov-s-troubled-launch.html.

Pirog, M. A. (2014). Data will drive innovation in public policy and management research. *Journal of Policy Analysis and Management*, *33*(2), 537–543.

Prpić, J., Taeihagh, A., & Melton, J. (2015). The fundamentals of policy crowdsourcing. *Policy and Internet*. doi:dx.doi.org/10.1002/poi3.102.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., et al. (2010). Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review*, *9*(1). Retrieved from http://arxiv.org/abs/1303.6886.

U.S. News and World Report. (2012). Best grad schools: Information and technology management. Retrieved from http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-public-affairs-schools/information-technology-management-rankings?int=a95909&int=a06908.

Wade, C. (2014, April 15). The Reddit reckoning. *Slate*. Retrieved from http://www.slate.com/articles/technology/technology/2014/04/reddit_and_the_boston_marathon_bombings_how_the_site_reckoned_with_its_own.html.

ABOUT THE AUTHOR

Ines Mergel is associate professor of public administration and international affairs at Syracuse University's Maxwell School of Citizenship and Public Affairs. Her research focuses on networked governance as well as digital and managerial innovations in the public sector.