Treating Data-Driven Management Seriously: A Way Forward for Public Administration

Jesse Lecy, Arizona State University
Minnowbrook Conference, August 2018

Major advances in technology can serve as harbingers of societal change, offering both the promise for societal progress and the portention of upheavals that create winners and losers. For this reason, e-government scholars treat the information economy as both an opportunity to "transform" governance and service delivery (Mergel et al., 2009), but also a shift that brings major challenges to public administration (Desouza et al., 2015), both in how we theorize about the administrative state in the information age, and in how we prepare students to manage and lead modern organizations. The current digital era provides an opportunity for a critical reflection on public management education.

The most prominent feature of the digital era is the volume of information produced and consumed by individuals. According to one estimate (Hilber & Lopez, 2011), the average person reads 174 newspapers worth of text and produces 6 newspapers worth of text a day, more than five times the volume in 1986. The capacity of two-way communication channels has grown at a dizzying 28 percent a year. The typical office worker receives 120 emails, compose 40 emails in response[1], and send 128 text messages each day[2]. It only requires three minutes spent on each email to fill a full 8-hour work day with just email! Americans spend nearly two hours a day on social media[3], and transmit 2,657,700 gigabytes of data each minute over the internet[4].

The complexity of the information economy impacts the public sector in several ways. New technologies like Twitter and Facebook often supplement official communication channels and create public forums that can be hard to manage, resulting in more work but not necessarily efficiency gains. The 1962 Freedom of Information Act was a landmark advance in government transparency, but it created new policies and technologies for agencies to archive and retrieve information. The DATA Act and similar state and local laws have generated over 200,000 open government datasets that agencies must now manage and distribute. Although there is opportunity for these new communication tools and data resources to improve governance and services delivery, public organizations accrue transactional costs and operating in this data-rich environment requires new skills and training.

I posit that the information economy poses two serious challenges to the traditional public management curriculum. First, there is a direct relationship between the pace at which an environment changes and the need to rely on new data for effective managerial action. The information economy increases communication complexity (communication ties increase exponentially, not linearly, with the number of nodes added to a network) and accelerates the pace at which the organizational environment changes and. As a result, the mental frameworks and decision heuristics of managers become outdated faster. Past experience increasingly becomes ineffective as a guide for reliable strategy. Second, the ability to be a data-driven manager requires two distinctive but complimentary skills: ingesting and digesting data. The ingestion process requires managers to collect new data quickly or find, access, and import

[1] https://www.templafy.com/blog/how-many-emails-are-sent-every-day-top-email-statistics-your-business-needs-to-know/
[2] https://www.textrequest.com/blog/how-many-texts-people-send-per-day/
[3] https://www.socialmediatoday.com/marketing/how-much-time-do-people-spend-social-media-infographic
[4] http://www.iflscience.com/technology/how-much-data-does-the-world-generate-every-minute/

archival data. Actionable knowledge requires managers with the skills to analyze, interpret, and report data efficiently. Note that I use the term "manager" instead of policy analyst, because one of the primary tasks or management is the evaluation of performance of programs and staff in order to green-light projects, manage staff, and allocate resources. These tasks all require data of some kind.

From these premises we can derive a set of principles for effective data-driven management, and propositions about the skill requirements of modern management:

ONE: Using new mobile platforms and survey tools data can be collected in hours or days, and at great volume. Many data collection processes can be automated, can be collected passively through sensors, or can be generated as a byproduct of administrative processes. Data-driven managers should be capable of defining useful performance indicators, selecting appropriate survey instruments, and using the most common tools like Google Forms, KoBo Collect, ArcGIS Collector, and sensors to gather information quickly and at low cost.

TWO: Data is not always numbers. It can be text, spatial coordinates, raster data from sensors, or images. And it rarely exists in rows and columns and never arrives clean. Effective data-driven managers understand common data structures and can work with data in a variety of formats. This requires basic data programming skills. Professional programs disadvantage students when they work solely with spreadsheets and non-scripted software like Excel or SPSS.

THREE: The value of data is greatly enhanced when it is combined with other data sources, thus linking data is a valuable skill. Students should be exposed to useful and ubiquitous sources like Census data, and should learn how to make datasets commensurate by aggregating data to equivalent units of analysis and practice merging. Many cases will require the use of crosswalks, disambiguation, or approximate matching to generate shared keys.

Data skills are necessary but not sufficient skills to achieve performance gains. If data is not turned into actionable information and integrated into a managerial or decision-making process it will not benefit an organization. There are several challenges to overcome to make this happen, though. Generating insights from data is hard, and organizations and employees by nature will resist change. Systems and incentives are needed to make sense of the data and facilitate organizational learning.

ONE: Simply reporting information to employees, even if summarized in nice reports or dashboards, will fail to create performance improvement because of information blindness, the inability to make sense of information that one does not help generate (Duhigg, 2016).

TWO: Information blindness is countered through managerial experiments or data experiments. Managerial experiments are conducted when teams design and implement interventions in the management of programs and collect data to determine their impact. Data experiments are conducted by identifying natural or quasi-experiments in datasets that can offer the types of evidence needed to guide managerial practice. They typically require strong counterfactual reasoning skills.

THREE: Enterprise software is the enemy. It robs an organization of deepening internal capacity by hiring outside experts to build the system, has long and rigid development cycles that prevent feedback necessary for learning, and it hard-codes a model of performance into survey tools, databases, and analytical models embedded in the system design. As a result, enterprise systems prevent organizational learning, and they create dependencies on expensive outside contractors. Organizations that embrace

open-source tools and agile development processes are more likely to experience organizational learning, empower employees, and catalyze performance gains.

These principles pose two serious challenges to the classic PA curriculum that arise in a data-rich world:

ONE: In a data-rich world, experiments are more powerful than theory. Theory helps us understand the interrelated parts of a complex system, but it typically requires the use of experts that have spent years honing abstract models of the system. Experts are detrimental to the development of a performance culture in organizations because they centralize power by encoding knowledge in abstract and technical terms that prevent dialog. In doing so, they minimize opportunities for learning and the agency of employees in charge of implementing programs. Alternatively, when employees participate in managerial experiments they think critically about the design and implementation of programs, develop insight into what creates impact, and deepen their commitment to a performance culture.

TWO: In a data-rich world, counterfactual reasoning is more powerful than causal modeling[5]. Counterfactual reasoning with data requires the identification of groups that represent states of the world with and without some treatment (a policy or program)[6]. While it is relatively easy to teach counterfactual reasoning with data, as it is primarily about logic, it is comparatively much harder to teach causal analysis using statistical models[7]. Our MPA curriculums have been influenced largely by the disciplines of program evaluation and applied econometrics, which are powerful tools for conducting rigorous estimates of program impact. The biggest need in most organizations, however, involves descriptive analysis to assess the current state of populations or programs, or descriptive statistics using data from managerial experiments.

A Way Forward:

The MPA curriculum was developed in an era where data was scarce, policy cycles were slower, and formal program evaluation was more common. As a result, MPA programs teach courses that emphasize the evaluation of programs using formal research design and econometrics, not managerial experiments using agile design cycles. As a result it is common for MPA programs to over-emphasize the importance of formal statistical models and under-emphasize the usefulness of general data skills. Perhaps more important, students that are technically inclined have historically been tracked into MPP programs or policy analysis concentrations. We sometimes make assumptions that managers will consume and not produce evidence, and thus they do not need analytical skills beyond budgeting. If we are to embrace a data-driven management paradigm, however, the management and analytics divide becomes false dichotomy that must be bridged. For manager to be truly data-driven, they need to develop the capacity to manage, access, manipulate, and ingest large volumes of data, and the ability to use the data to generate meaningful insight and actionable information.

---

[5] For inspiration check out the syllabus for the course "Calling Bullshit: Data Reasoning in a Digital World"

[6] Steven Levitt's famous study on abortion and crime is a good example of applied counterfactual reasoning. The study estimates a complex econometrics model, but the main argument is made through counterfactual reasoning and simple descriptive statistics. Levitt argues that since it took up to two years for states to implement the ruling the drop in crime should occur in each state relative to when abortions became available. We only expect crime to drop in the age cohort born after Roe v. Wade, but not in the cohort born before. And if the theory is true, we would expect to see an increase in crime in countries where abortion was legal became illegal, like Romania. All three of these assertions can be tested using comparisons of means, not complicated models. They show how managers can test their theories by reasoning about what they expect to observe in the data if their theory holds instead of relying on complex inferential models.

[7] Statistics is the art of applying complex mathematical models to simple (rectangular) datasets. Data science is the art of applying simple mathematical models to complex datasets.

Major advances in open-source software offer clear opportunities to update the MPA curricula in ways that address the challenges outlined above.

ONE: Elegant specialized programming languages like R and Python were designed specifically for data analysis and have replaced cumbersome mainframe programming languages of the past. As a result, introductory data programming can be taught in one semester, foundations of data science in two or three.

TWO: A data programming language is your passport into the data science ecosystem, a global community of experts that have turned the internet into a giant collective brain. Access to this global knowledge repository requires a basic understanding of data programming and enough vocabulary to search for answers on Stack Overflow. The ecosystem is continuously developing new tools and resources, making membership more valuable and the opportunity costs of not participating higher. Programs like Excel, SPSS, and Stata do not provide a pathway to citizenship in this new land.

THREE: The expertise needed to develop powerful custom applications and the time it takes to create these tools have decreased exponentially. To create a web app, for example, it used to require seven programming languages (Linux, Apache, SQL, PHP, HTML, CSS, and JavaScript). You can now develop an equivalent web app with basic R skills and a couple of packages. This means tools like automated reports or data dashboards can be developed in a couple of days, not a couple of months, all with free software, making data-driven approaches to management possible in ways that they have never been before.

FOUR: The fields of design thinking and agile management provide solid theoretical foundations for data-driven management, and many practical tools for teaching a version of the scientific process for management.

These opportunities do not come without costs. If the information burden of modern organizations is as high as some research suggests, and increasing the capacity for organizations to utilize data is the most effective way to improve performance, then we need to think seriously about how to adapt the MPA management curricula to prepare effective leaders.

Treating data-driven management seriously requires us to identify challenges that are unique to modern organizations and decide whether new tools might better prepare students to manage the challenges. Are the most valuable skills the ones that we currently emphasize, or has the MPA curriculum become burdened with path dependency? Do powerful new data science tools, the ability to collect and analyze data quickly and cheaply, offer an opportunity to re-think some of the fundamental skills that managers need to be effective?

As I have argued above, modernizing the curricula is not as simple as adding data science electives. The data-driven approach to evidence and causality through counterfactual reasoning requires a slightly different mindset than a traditional program evaluation approaches that rely heavily on inferential statistics, research design, and well-structured data. MPA curricula are already operating at the saturation point, so new course material would likely supplant current content. Embracing such an approach would force current instructors to kill many of their darlings. Thus, change can be hard. But it may be eminent, and our field will be better off if we take the path of purpose and intention instead of the path of least resistance.

References:

Duhigg, C. (2016). *Smarter faster better: The secrets of being productive*. Random House. *CH8: Absorbing Information.*

Desouza, K. C., Swindell, D., Smith, K. L., Sutherland, A., Fedorschak, K., & Coronel, C. (2015). Local government 2035: Strategic trends and implications of new technologies. *Issues in Technology Innovation*, (27).

Mergel, I., Schweik, C. M., & Fountain, J. E. (2009). The transformational effect of Web 2.0 technologies on government.

Harrison, T. M., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N., ... & Pardo, T. (2012). Open government and e-government: Democratic challenges from a public value perspective. *Information Polity*, *17*(2), 83-97.

Mossberger, K., Wu, Y., & Crawford, J. (2013). Connecting citizens and local governments? Social media and interactivity in major US cities. *Government Information Quarterly*, *30*(4), 351-358.

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *science*, 1200970.

Hilbert, M. (2012). How much information is there in the "information society"?. *Significance*, *9*(4), 8-12.