An Inclusive Framework for Data Science
Competencies in NASPAA MPA and MPP Programs

Stephen Kleinschmit, Ph.D.
Director, Master of Science in Civic Analytics
College of Urban Planning and Public Affairs
University of Illinois at Chicago

Abstract

The academic landscape of data science has experienced substantial growth within the last decade, primarily through the establishment of graduate programs within computer science departments, newly formed data analytics departments, and business schools. As a response to the growing demand for public servants with advanced data skills, a small but growing number of institutions have responded by establishing public data science programs. Anchored within the realms of public policy, public administration, and urban planning, these programs can provide a comparative basis for establishing a public data science competency framework.

This whitepaper highlights the core insights derived from the process of establishing the Master of Science in Civic Analytics at the University of Illinois at Chicago. It is intended to inform the development of competencies for NASPAA MPA and MPP programs that wish to incorporate data science concentrations within their curriculum. It is also meant to introduce program directors to civic technology-informed perspectives of data science, serving as a contrast to other approaches in the field. My intent in crafting this paper is to broaden the construction of public data science beyond the domain of public policy analysis, smart cities, and federal agencies towards a broadly inclusive framework that can appeal to a wider population NASPAA member institutions. My proposed framework is intended for all programs but is intended to show the value of data science for those that specialize in preparing students for careers in state and local government, and the nonprofit sector.

1. Introduction

The public sector is drowning in data, so much so that is struggling to analyze most of the data it collects (Newcombe, 2017). The ubiquity of data technology, the rise of new regulatory problems, and adoption smart infrastructure are trends that have evidenced a need for public servants trained in big data fundamentals. Graduate programs in public affairs have been one of the primary means for supplying practitioners that can adapt to the changing needs of public service. To date, information technology is not a core skill in many public management or policy curriculums, and the diminished emphasis on IT skills within accreditation standards has led to a widening gap between theory and practice in the field (McQuiston and Manoharan, 2017). Consequently, the field of public affairs is producing a workforce unprepared for the digital world (Ganapati and Reddick, 2016).

In explaining the need growing need for public data scientists, one can look towards the trends in the private sector for clues. The US Bureau of Labor Statistics estimates an overall growth of 30% in Mathematician and Statistician positions between 2018-2028, far exceeding the estimate of 5% for all occupations (BLS, 2018). *The McKinsey Report* estimated a shortage of 150,000-190,000 data scientists in 2018, with 1.5 million workers managers and analysts with big data skills (Ash Center at Harvard University, 2015). IBM predicts that by 2020, nearly 2.7 million positions in data science and analytics will be posted *annually* (BHEF, 2019). At odds with the ascendant nature of private-sector employment is the reality of the public sector, where "deep questions remain about the ability for many areas of government and civil society to identify, cultivate, and retain individuals with the necessary skills for success in a world increasingly driven by information technology" (Freeman, 2013). Much of this skepticism can be traced to the sector's ability to recruit and retain talent, rethink compensation models, and reform its training programs (Ash Center at Harvard University, 2015). The latter, rethinking the nature of its training programs, is the primary purpose of this paper.

In working to estimate the current degree of capacity for training public data scientists, one can look to the establishment of degree programs and concentrations in the field. Of the nearly 250 operational data science degree programs tracked by the *NC State Institute for Analytics* in May 2019[1], only 6 programs were actively training students with primary coursework in government data science (2.4%). Given that nearly 1 out of 6 workers in the United States are employed by the public sector (BLS 2018) and government expenditures amount to 37.9% of the country's GDP[2], it seems evident there that the current level of production of program graduates is insufficient to meet sectoral needs.

2. Literature Review

Data science, with regards to the modern context of this work, began in earnest in 2001 (Cleveland 2001, Press 2013, Donoho 2017, Brady 2019), though the principles that came to define the field are traced back to the early 1960s (Donoho 2017). From the outset, scientists such as John Tukey defined its principles as being complementary to (and not wholly defined by) statistics and took into account such elements as the rapid progression of computer technology, the increasing size of datasets, and the rising importance of quantification across disciplines (Tukey, 1962). Crucial to the establishment of data science are the concepts of interdisciplinarity, the merging of research methods and exploration of data across disciplinary silos, trends that have also emerged within broader academic inquiry (Cleveland 2001).

The 1990s brought dramatic improvements in computing power and data storage capabilities moved enterprise-scale data processing from mainframes to personal computers. Since that time, Internet connectivity and data usage have exploded, and developments in Information Communication Technologies (ICT) have inspired movements in the public sector, including e-government (eGov) and mobile government (mGov). These movements were primarily associated with the implementation of 1st and 2nd generation information technologies, including websites and social media (Mergel, 2010). Information technologies served as outlets for public information, portals for public service delivery, and as mechanisms for citizen feedback. The early 2000s brought the Open Data movement, which helped facilitate the availability of machine-readable datasets for download, a critical tool for academic research and government accountability.

The emergence of data science as an academic discipline was not without controversy. The overlapping realms of long-established disciplines such as statistics and computer science

[1] NCSU Insitute for Advanced Analytics. https://analytics.ncsu.edu/?page_id=4184
[2] OECD GDP by County. https://data.oecd.org/gdp/gross-domestic-product-gdp.htm

lead leaders of several prominent professional organizations to fight over the field's "intellectual home" (Donoho, 2017). Today, though there remains no standard agreement as to the intellectual home (or even definition) of data science, research has sought to bound the parameters of the field by delineating the characteristics that differentiate it from other areas of academic inquiry. I reference Wil van der Aalst (2016), who defines the four defining characteristics of data sciences:

- *Statistics*. At the heart of data science lies the foundation of applied statistics. A data scientist must understand its fundamentals, such as validity, probability theory and inference, causality, data screening, cleaning, coding, and transformation.
- *Data mining*. Mining begets an understanding of "big data" and methodologies to derive knowledge from vast quantities of collected data. Combined with machine learning and artificial intelligence, there is a dramatic increase in the ability for analysts to discover new relationships and hypotheses. Mining also invites an understanding of coding to collect and analyze data not traditionally examined due to technological or formatting constraints.
- *Database management.* An understanding of the structures of data storage, including types of databases, information management principles, and synthesis.
- *Distributed systems*. Coordinating resource sharing across networked computers as a means to solve complex problems.

In addition to these core characteristics, the practice of data science involves the command of constantly evolving "hard skills" in coding, programming, and visualization. To adequately define public data science, the field agnostic principles of data science must be merged with the core principles of public service. Much like the intellectual discourse that worked to differentiate the principles of public administration from business administration and management, public data science must be rooted in the confines of public institutions, such as:

- *State administration*. The ownership and provision of public goods under the authority vested by citizens. Recent movements in network governance and contracting have expanded the locus of this definition past simple government ownership into hybrid relationships with nonprofit and private organizations (Wise, 2010).
- *Public accountability*. The actions of servants are accountable to the public through democratic mechanisms such as political oversight, civic participation, governance processes, administrative rulemaking, and investigative journalism. The notions of government accountability reflect "a tension between the seemingly neutral concept of performance and a broader context of democratic -and necessarily partisan-judgements" (Olsen 2015, Nielsen and Moynihan 2016)
- *Constitutional framework*. Governments are ceded the right to engage in regulatory police powers under the "health, safety, and general welfare" provisions of the Constitution. This system affords citizens formal rights such as due process of law, freedom of the press, unreasonable search and seizure – the interpretations of which can change through the evolution of social attitudes or technologies that enable new forms of expression (Balkin, 2004). This system also includes the principles of federalism, the separation of powers between branches, policymaking, administrative rulemaking, and regulatory enforcement.

Aside from understanding the institutions of government, there are numerous aspects of public service to be considered. Its practice includes “soft skills” such as professional ethics and commitments to social equity that complement technical skills like budget forecasting and program evaluation. There is also a desire for academic programs to incorporate the principles of nonprofit administration and network governance into their curriculums, which posits an important question for this whitepaper – Is public data science solely anchored in the principles of government, or should it be defined with a more sectoral lens? In seeking to establish competencies, I argue the latter, that defining the fundamental nature of public data science is inclusive to perspectives reflecting the involvement of nonprofit actors and network governance.

2.1 Evolution of the Information Environment

Public data science is not a new enterprise, but rather an evolution of information technology. Early practitioners of public data science have included broadly trained data scientists, policy analysts, and information technologists, though relatively few serving in these roles has formal training in the principles delineated in this paper. Consequently, new management approaches, governance structures, and policy frameworks are missing and pose a challenge for governments to operate effectively in the age of big data (Liu and Quan, 2015). Public affairs programs are obliged to evolve to train a new workforce prepared to address a multitude of new problems in the decades to come. A review of civic technology and the smart cities literature reveal several trends:

- *New regulatory needs*. The introduction of data-driven infrastructure has invited a host of societal concerns. From ridesharing to corporate surveillance, government agencies need improved capacity for monitoring and enforcing laws within a digital society (Arner, Barberis, and Buckley, 2017).
- *Complexity*. IT use within government must evolve to derive an understanding of a complex social reality (Helbig, Gil-Garcia and Ferro, 2005). Specialized data science skills remain outside of the realm of traditional academic curriculums and professional development pathways in public service. Solving this problem will require “new ways to form concepts from data, to do descriptive inference, to make causal inferences, and to generate predictions” (Brady, 2019).
- *A new institutional structure*. A smart city is one that integrates technology-driven approaches to into its structure and operations, focusing on improving its economic development, governance, sustainability, infrastructure, and service delivery (Gil-Garcia, Pardo and Nam, 2015). City information offices, data teams, and other specialized units are an indicator of the evolution that is occurring in local contexts and posits a growing need for new types of public employees to staff them.
- *Security and privacy concerns*. Public sector data specialists will need to understand the emerging environment of data security and proactively implement solutions to preserve the integrity of civic data (Keymolen, Prins and Raab, 2012). These specialists also must understand the privacy implications of data in public contexts to ensure citizen privacy and comply with state and federal law. Balancing government transparency with accountability, the collapse of distinctions between public and

private spheres, and increasing monitoring and profiling of citizens are also significant concerns for governments to address (Scassa, 2014).

- *New ethical challenges*. The adoption of data technology begets many of the moral dilemmas encountered within public service, intersecting host of new types of ethical problems brought on by the onset of advanced data technologies, such as artificial intelligence. For example, a new reliance on instrumental rationality, combined with the assumptions and social biases programmed into algorithms, can expand economic and social inequality (Kitchin 2016, Kirkpatrick 2016).
- *Open government and public accountability*. The data science paradigm is anchored in the tradition of promoting open access and accountability for government decision-making (McNutt et al., 2016). Dashboards are one form of open data technology that facilitates transparency and improves public decision making (Matheus, Janssen and Maheshwari, 2018). Civic hacking (and "hacktivism") has also emerged as a potent force, with groups such as *Code for America* promoting democratic engagement and enriching civil society (Shrock, 2016).

Since 1986, NASPAA has recognized the emergence of an information society as a driver in improving the digital literacy of public sector employees (Kraemer et al., 1986). Creating an academic framework for public data science invites public affairs programs to consider technological competence as a core practitioner skill, something that has so far alluded the field (Cleary 1990, Brown and Brudney 1998, Ganapati and Reddick 2016, McQuiston and Manoharan 2017). Much of the value of public sector data science is anchored within its interdisciplinarity; by combining knowledge of public problems, institutions, and ethics to expertise in applied statistics and information science, it can emerge as an important mechanism for improving technical competence in the field.

2.2 The Need for New Skills

A new information environment has given rise to the need for public sector employees with advanced data and technology skills not contained within most degree or professional development programs. The problems listed below are a sample of those that typify the challenges of the new operational environment:

- *Inadequate analytical capacity*. Without proper methods underlying analysis, the use of data in government agencies is essentially useless (Archenaa and Anita, 2015). Advanced data science skills are not taught in many of the feeder disciplines that populate public service (Manoharan and McQuiston, 2016). Public affairs schools, political science departments, sociology, public health, and other allied disciplines provide largely introductory sequences in data analysis and manipulation. The programs that do require advanced statistical and data science skills, such as those housed within computer and information science departments, are heavily oriented towards delivering graduates to the private sector (Freeman, 2013).
- *Real-time data.* Public agencies have typically been anchored in a periodic information reporting paradigm, working with data collected at discrete intervals (monthly, quarterly, annually). The rise of sensor-derived data brings the need for

specialists with skills to manage “a vast deluge of real-time, fine-grained, contextual and actionable data, which are routinely generated about cities and their citizens by a range of public and private organizations” (Kitchin 2016, 2).

- *Automation.* Real-time data has led to a demand for artificial intelligence and machine learning techniques to sort through massive amounts of data. Their use can improve the processing of data, allowing better data-informed decision making. Artificial intelligence can also reduce administrative burdens, help resolve resource allocation problems, and take on significantly complex tasks” (Mehr, 2017).
- *New data formats*. The collection of voice, video, and picture data invite new applications and measures of accountability for public agencies. From red light camera photos to police body camera footage, new types of data invite practitioners to possess specialized skills for screening and analyzing non-text data (Kim, Trimi, and Chung 2014).
- *Siloed Data.* Public data is distributed across an expansive institutional landscape. Integrating this data into consolidated datasets allows agencies and researchers the opportunities to ask new questions, derive new insights from data, and develop novel methodologies for solving public problems. Preventing this integration includes “data and technological incompatibility, the lack of institutional incentives to collaborate, and the politics and power struggles around a pervasive silo structure in most governments” (Gil-Garcia, 2013). Consequently, addressing silos entails removing legal and technological barriers that inhibit the use of data for solving social problems (Liebman, 2018).
- *Visualization.* Visualization tools allow citizens to navigate, visualize, and query public data. Traditionally this has been a specialist skill, but within the realm of civic technology there exists a trend in visualization platforms used for accountability tools by nonprofit groups (Bekkers and Moody 2011,)
- *Quantification bias*. There is a danger that data, and not the issues raised by operations or constituencies, will drive public questions and analysis” (Mergel, Rethmeyer and Isett, 2016, Lavertu 2016). A return to the positivist tradition and instrumental rationality begets training for students to understand the ethical implications of their methods, the limitations of quantitative methods, and recognition of the value of mixed-method and qualitative research methodologies.

These problems beget a need for new types of training designed for new regulatory demands. The data science paradigm will not simply be the domain of large agencies and strategic-level policy analysts; its skills will also be critical to improving the delivery of public services in small contexts. Essential to realizing the potential of this paradigm will be a need to restructure elements of the field’s academic and professional development programs.

2.3 Gaps in Training

Public affairs graduate degree programs (MPA, MPP, etc.) have been important mechanisms for educating advanced generations of public service practitioners. Their curriculums have traditionally provided exposure to data analysis and research methods, with 89% of programs requiring a course in this subject area. Budgeting and financial forecasting, policy analysis, program evaluation, survey research, and geographic information systems are all elements of accredited programs, though there are substantial differences in their degree of emphasis. In order identify the gaps in training, I am using the challenges identified in Section III to specify what I believe to be the imperative elements of modifying public affairs curriculums to meet current and future challenges.

- *Advanced statistical training*. Many, if not most, public affairs programs provide a 1-2 course methods sequence that exposes students to descriptive and inferential statistics, often culminating with multivariate regression. Data science includes Bayesian statistics, survival analysis, longitudinal analysis, network analysis, simulations, principal component analysis, multidimensional scaling, and other advanced analytical and visualization functions (Blei and Smyth, 2017).
- *Spatial methods*. Spatial methods courses are not a standard part of most public affairs curriculums. Though some programs do include GIS courses and concentrations the technology has not been pedagogically integrated at the program level (Ferrandino, 2014. Spatial methods such as inverse distance weighting, k-means clustering, interpolation, georeferencing, spatial regression, and proximity modeling are all common methods employed within the data science realm, and generally outside of standard desktop program environments.
- *Data science skills*. Common "hard skills" in data science include an understanding of programming languages R or Python, SQL, relational and nonrelational databases, web scraping, web development, visualization platforms, dashboards, Java, Hadoop, Amazon AWS. Many of these skills are traditionally housed in computer or information sciences programs and can only be taken outside of a public affairs program context. With most graduates of these programs going into the private sector, these subjects are anchored in the needs of business and software development communities (Freeman, 2013).
- *Information Life Cycle Management*. There are a range of government challenges to address regarding Big Data, including "access and dissemination; digital asset management, archiving and preservation; privacy; and security (Bertot, Gorham and Jaeger, 2014). A data science program in the public sector must ensure training in safeguards that used during all stages of the life of data: Its creation, dissemination, storage, and destruction. MPA and MPP programs have not typically included coursework in these topics.
- *Ethics*. Ethics courses in public affairs programs are not a common curriculum requirement. Those that do exist primarily focus on managerial ethics, including principal-agent relationships, codes of ethics, and the nature of controls on behavior (Cooper, 2014). The ethics of data technology look at instrumental rationality, the introduction of bias, informed consent, identity protection, the revenge effect of technology, the weaponization of data, data scrubbing and re-identification, and other principles that lie outside the realm of traditional ethics curriculums in the field (Fairfield & Shtein 2014; Roman 2015, Kitchin 2016).
- *Social Equity*. Information scientists need to consider the policy outcomes related to their work, including whether disproportionate impacts will be placed on certain populations, and if there is bias in favor or against certain groups (Jaeger et al. 2015, 184). Inequality

may manifest itself within the cultural assumptions programmed into algorithms and machine learning models (Begbie, 2019). Addressing this will include new forms of transparency, such as making decision-making algorithms open for public review.

3. A Brief Assessment of Public Data Science Degree Programs

The emergence of academic programs in data science is a recent trend. The first graduate program in data analytics began at North Carolina State University in 2007. Since that time, a diverse set of academic programs exploded into existence, with nearly 250 programs in general analytics, business analytics, and informatics established by May 2019[3] (with 230 programs alone since 2013!). This growth has resulted in computer and information science-related degrees becoming one of the fastest-growing disciplines in higher education (NCES, 2018). The market for graduate programs in data science and analytics is expanding rapidly. Of the 250 programs tracked by the NC State University Institute for Advanced Analytics (as of 5/2019), few have a primary emphasis in the public

NYU's Center for Urban Science and Progress began offering the M.S. in Applied Urban Science and Informatics in Fall 2013, graduating their first class in 2014 (NYU CUSP, 2014). The University of Chicago's Harris School of Public Policy established the Master of Computational Analysis and Public Policy program in September 2014, becoming the first STEM program in Public Policy. Its curriculum combines coursework in computer science, statistics and public policy (UC CAPP, 2019). Johns Hopkin's MS in Government Analytics program also began operations in 2014, becoming the first government analytics program. Concentrations within public affairs degrees emerged shortly thereafter, with Carnegie Mellon (2016) integrating subject matter into their MPPM degrees. The first dedicated public sector data science degrees at public universities are slated to come online in Fall 2020, with Rutgers offering the Master of Science in Public Informatics, and the University of Illinois at Chicago offering its Master of Science in Civic Analytics.

This section offers a typology of public data science programs that were derived from a review conducted as part of the background research for UIC's Master of Science in Civic Analytics. The programs overlap in the substantial domains of statistics and principles of data science but are grounded in different intellectual traditions. As noted in the description, the programs are also housed in a variety of academic disciplines and structures, each influencing their curriculum design. All of the programs are either a standalone degree or concentration within ann MPA/MPP/MPPM curriculum unless otherwise noted.

- *Computational Public Policy & Information Science* (University of Chicago, Carnegie Mellon) – The programs emphasize a more academic approach to public data science, with curriculums that heavily emphasize policy analysis and economics. Formal coursework in computer science, large scale computing and artificial intelligence are also distinguishing elements. Though housed in policy schools, the specialized nature of these program's coursework would be difficult to model in replicate by many NASPAA member programs.

[3] Graduate Degree Programs in Analytics and Data Science, https://analytics.ncsu.edu/?page_id=4184

- *Government Data Science* (Georgetown, Johns Hopkins) – Johns Hopkins Master of Science in Government Analytics and Georgetown's Master of Government Data Science populate these programs. Their curriculums are anchored in public policy and management, very much centered on the needs of federal sector workers and contractors for which these institutions have been traditional feeders. The Johns Hopkins program distinguishes itself in that it is the only public data science program offered online, and one of the few that offers concentrations.
- *Urban Analytics and Informatics* (New York University, University of Pennsylvania, Rutgers). This group of institutions operationalizes public data science with a heavy emphasis on urban planning and affairs. The programs in these categories provide an interdisciplinary core that focuses on urban planning, urban policy, and smart cities. Core classes include fundamentals of urban planning, urban economics, and development. These programs are housed in urban science center (NYU), a school of design (UPenn), and a school of planning and public policy (Rutgers), which lie outside the traditional institutional structure of the policy schools and public affairs colleges in which NASPAA operates.
- *Civic Analytics* (University of Illinois at Chicago) – The University of Illinois at Chicago's Master of Science in Civic Analytics program combines the principles of two important movements, civic technology and data analytics. The program's applied emphasis is rooted heavily in the public administration tradition and derived in part from an expansion of the MPA program's former Information Technology and Performance Measurement track. It is the only program that offers a multi-course geospatial methods and visualization sequence as part of its core curriculum, offered as an elective in most of the compared curriculums. It is also the only program currently housed in a public administration department.
- *Concentrations within a Data Science Programs* (American, UT-Dallas). Options in this classification take an inverted approach to the others – exposure to public affairs as a concentration, with the core curriculum in a computer or data science program. Students that seek to study public data science at American University do so primarily as a student within a data science department, rather than a public affairs school. The core curriculum differs significantly from the subject matter that of MPA/MPP/MPPM programs, and exposure to public affairs coursework is largely contained within concentrations. Students can choose from concentrations in business analytics, finance or investigative journalism, in addition to the Applied Public Affairs concentration. UT-Dallas offers a Master of Science in Social Data Analytics and Research, which allows students to pursue elective coursework in public and nonprofit management, and public policy.

The above programs are the early adopters whose programs that will undoubtedly influence the academic landscape of public data science. The typology is intended to evidence the breadth of implementation in the field - whether housed in a public policy school, urban affairs program, or public administration department, each manifestation imparts a different "flavor" to their curriculum. This typology excludes programs who host specializations are anchored in government information management or technology; while related to the area of examination, the coursework is not grounded in the principles of data science and largely reviews managerial topics.

In reviewing the curriculums and course descriptions for the aforementioned institutions, and through my comparative review field literature, it became evident that commonalities existed in academic program design. Broadly, these areas can be merged into subject area domains, classified below.

- *Public Sector Information Technology* – This domain consists of academic preparation in topics relating to the use information technology in government and nonprofits, civil society, ethics, civic activism, regulatory compliance, data security, privacy, and information policy.
- *Data Science* – This domain covers coursework that enables practitioners to collect, generate, analyze, merge, store, and structure data. It is anchored in "hard skills" such as programming languages, database management, infrastructure, and machine learning theory.
- *Statistical and Spatial Methods* – Coursework in this domain in seated in advanced statistical methods, research design, spatial analysis, survey methods, data screening, cleaning, and transformation. While there may be some content overlap, these skills are generally more advanced than those encountered in the research methods core of MPA/MPP programs.

Aggregating coursework into these domains helps to narrow the focus of possible competencies. In the next section, I synthesize the insight from this program review, the literature review, and stakeholder involvement into a proposed framework for NASPAA data science concentration competencies.

4. Competencies and Curriculum Components

The following recommendations are derived from a synthesis of the included literature review, the review of existing public data science programs, faculty experience, input from advisory board members in civic technology organizations, the included case studies in Section 5, and the core competencies derived from the creation of the Master of Science in Civic Analytics program at the University of Illinois at Chicago. My approach is in part grounded through the work of other researchers, whose research has posited competency models for information technology and e-Government (McQuiston and Manoharan, 2016), global cultural competency (Appe, Rubai and Stamp, 2016), health administration (Rissi, 2014), and diversity (Johnson and Rivera, 2007)

Each domain linked to a set of broad competencies, which are further linked with curriculum components that may be used to provide evidence of compliance with a future NASPAA data science accreditation or certification program. They are constructed to account for individual program needs but are focused enough to offer detailed guidance on curriculum.

Domains	*Competencies*	*Curriculum Components*
Public Sector Information Technology	1. To Understand the institutional landscape of civic technology and public data science; 2. Explain how information technology can be used to promote government accountability; to 3. Understand the ethical dilemmas posited by information technology in the public sector; and 4. Be knowledgeable of principles of data security and privacy.	Ethics, e-Government, data-driven advocacy, hacktivism, data-driven governance, open data, regulatory compliance with state and federal law, privacy, civic technology, smart cities, technology policy, social equity, data-informed approaches to solving public problems, data security, data misuse and weaponization, FOIA, transparency, records maintenance.
Data Science	1. To understand the principles of big data and scalable computing; to 2. Be proficient in methods for collecting, integrating, analyzing and displaying data generated by public and private organizations; to 3. Be knowledgeable of the fundamentals of commonly used programming and querying languages; and to 4. Utilize systems for collecting, analyzing, and displaying public data.	Data mining, artificial intelligence, machine learning, programming or coding, R or Python, relational and nonrelational databases, SQL, cloud computing, distributed systems, API, web scraping, non-text and unstructured data, siloed data, simulations, development environments, Github, Jupyter, visualization techniques and platforms, neural networks, Internet of Things (IoT), web development, application development, dashboards, Tableau
Statistical and Spatial Methods	1. To understand and apply appropriate statistical methodologies to public problems; to 2. Explain the principles of data classification, cleaning and transformation, and the effect on the output of statistical operations; to 3. Apply appropriate spatial analysis methodologies to public problems; and 4. Understand fundamentals of mapping and visualization of spatial data.	Advanced statistical functions, model specification, probability theory, validity, data cleaning, screening, transformation, linear and nonlinear functions, classification, mapping, spatial analysis, GIS, survey research and design, network analysis, large-N analysis, algorithm development, operations research, econometrics, policy analysis, program evaluation

.

In implementing concentrations within MPA and MPP programs, a NASPAA data science certification program would require at least one dedicated course in each of the three domains for a typical 3-4 course concentration, and to allow programs to map competency attainment through coursework across the core curriculum or electives. Conversely, the insights from the table can be expanded inform the development of coursework models for public data science degree programs, or compressed to into skill-specific training programs or modules. My proposed competency framework is broad enough to allow for experimentation across a diverse set of institutional settings, and but specific enough to pin down the locus of public sector data science to offer prescriptive guidance. I invite future research to examine this framework for expansion or improvement.

5. Case Studies

The following case studies are intended to illustrate aspects the practice of public data science throughout Chicago. True to the city's reputation as an "analytics-driven city" (City of Chicago, 2019) there are diverse manifestations of data science across municipal agencies and nonprofits. The cases below provide insight into possibilities and challenges of digital government, and evident the sort of real-world problems and innovations that occur in local government and nonprofits. These cases are intended to expose readers to operational data science and provide context to the nature of competencies that may include programs that focus on state and local government, or nonprofit affairs. Each example also includes a shortlist of competencies and curriculum components that are not typically covered within the coursework of MPA and MPP programs and demonstrate linkages to the curriculum components mentioned within the previous section.

- *City of Chicago* – Health inspector deployment (applied analytics, machine learning) - In 2014, Chicago's Chief Data Officer, Tom Schenk, began a project working with the City Health Department to reduce the city's incidence foodborne illness. Employing 70 inspectors to oversee 15,000 operating restaurants, a new system was needed to identify dining establishments with a higher level of risk. Developing a custom tool named FINDER, a machine-learning model to detect the real-time incidence of foodborne illness. The inspectors were assigned to more frequent inspections of restaurants identified by the Finder system. Without hiring any additional inspectors, the city reported foodborne illness rate declined by 30% in just three years (Sadilek et.al 2018). *Gaps in current training demonstrated: Advanced statistical methods, data science skills, machine learning, and artificial intelligence principles*

- *ProPublica Illinois* – Parking tickets and bankruptcy study (government accountability and journalism) – Conducting one of the most extensive instances of data-driven investigative journalism, ProPublica Illinois worked to link the city's parking ticket data to bankruptcy data, as a means to understand the disparate effects of the Chicago's parking enforcement program. Contrary to the Northside's relatively quick payment of violations, a substantial number of residents in the city's lower-income South and Western neighborhoods experienced high rates of negative personal outcomes, including vehicle repossession, bankruptcy, and eviction. The story played an integral part in causing the city to modify their ticket issuance policies, as well as the structure of their ticket repayment plans (ProPublica IL, 2018). *Gaps in current training demonstrated: Open data principles, government accountability through data, integrating siloed data across agencies, innovation*

- *Chicago Police Department* – ShotSpotter (smart infrastructure/privacy) – ShotSpotter is a technology that assists police in locating firearm discharges in cities, helping law enforcement to respond to violent crime. The technology operates through triangulation, in which the fixed distance between sensors is measured to the source of the shot. Based on technology first used by the U.S. military in Southwest Asia campaigns, it is able to differentiate the distinct pattern of firearm noise from similar noises, such as the backfiring sound of vehicle exhaust. Once alerted to the presence of gunfire, the system directs law enforcement to a system of cameras to help capture evidence of the crime. Englewood, a Southside Chicago neighborhood long known for gang-related violence, experienced a 43% drop in gun-related violent crime in 2016-2017, a majority of which is attributed to the operation of the Shotspotter system. Though the system has demonstrated the potential of technologically enabled law enforcement applications, critics raise concerns about the accuracy of the data and citizens right to privacy (Southside Weekly, 2017). *Gaps in current training demonstrated: Real-time data processing, privacy, sensor networks, Internet of Things (IoT)*

- *Chihacknight* – Chihacknight is Chicago's weekly civic technology and hacking forum. Held each Tuesday night in the city's iconic Merchandise Mart, its format consists of two primary elements. The event begins with a presentation by an invited speaker(s), who represent a diverse array of organizations in government, nonprofits, industry, advocacy, and academia. The second half of the evening invites participants to participate in breakout groups, each offering a different avenue of technology-enabled citizen activism. This includes groups using API-derived transit data to optimize and improve routing in the city, developing an application to automatically warn residents of elevated levels of airborne petroleum petcoke from refineries, and seminars on the nuances of requesting data and documents through Freedom of Information Act (FOIA) requests. (Chihacknight, 2019) *Gaps in current training demonstrated: Civic technology, technology-based activism, accountability, digital governance processes, open data.*

The case studies presented present examples of applied public sector data science, and the types of institutions academic programs may interact with. Though I have used only examples from Chicago, the domain of public data science is far from being constrained to major U.S. cities; civic technology groups and data-driven governments exist all over the United States. There are relatively few contexts where the delivery of public services could not be improved or optimized through the application of data science principles, though managers must remain cognizant of the limitations of data-driven decision making.

6. Summary and Conclusion

The whitepaper proposes a framework for data science competencies within NASPAA MPA and MPP programs, helping to define the locus of its academic landscape. It is also aimed at helping to educate deans and directors about the challenges that information technology will present to graduates of their programs. While news stories may provide the impression that public data science is the purview of federal agencies and information offices in major U.S. cities, there is a real and immediate need for operational data scientists at all levels of government and the nonprofit sector – within urban, suburban and rural contexts. Many professional academic fields produce graduates destined for the private sector with hard skills substantially more advanced to than those taught our field's professional programs; consequently, public data science degrees and concentrations are a sorely needed complement to the field's existing academic programs and provide professional training options.

The call for improved technical capacity in the public workforce has largely been ignored by the field of public affairs, even in the face of an established body of research offering prescriptive guidance in addressing. Manoharan and McQuiston (2016) note that NASPAA once had IT competencies, but "removed the IT requirement from its standards for institutional accreditation because of the subject area's diffuse focus" (184). My hope is that my contribution to the NASPAA data science initiative will provide a foundation for reestablishing information technology as a core element of its wider accreditation standards, in addition to its important purpose of helping to define the bounds of this newly emerging academic discipline.

Bibliography

Appe, S., Rubaii, N., & Stamp, K. (2016). Advancing global cultural competencies: International service learning within NASPAA member programs. *Journal of Public Affairs Education*. https://doi.org/10.1080/15236803.2016.12002229

Archenaa, J., & Anita, E. A. M. (2015). A survey of big data analytics in healthcare and government. *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2015.04.021

Balkin, J. M. (2004). Digital speech and democratic culture: A theory of freedom of expression for the information society. *New York University Law Review*.

Begbie, N. (2019). How to mitigate negative algorithmic biases in machine learning. *ICTworks*. Available at: https://www.ictworks.org/algorithmic-bias-machine-learning/

Bekkers, V., & Moody, R. (2011). Visual events and electronic government: What do pictures mean in digital government for citizen relations? *Government Information Quarterly*. https://doi.org/10.1016/j.giq.2010.10.006

Bertot, J. C., Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity*. https://doi.org/10.3233/IP-140328

Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1702076114

Brady, H. E. (2019). The challenge of big data and data science. *Annual Review of Political Science*. https://doi.org/10.1146/annurev-polisci-090216-023229

Brown, M. M., & Brudney, J. L. (1998). Public sector information technology initiatives: Implications for programs of public administration. *Administration & Society*, *30*(4), 421–442. https://doi.org/10.1177/0095399798304005

Chichacknight (2019). Available at http://chihacknight.org

Cleary, R. (1990). What do public administration masters programs look like? Do they do what is needed? *Public Administration Review, 50*(6), 663-673. doi:10.2307/976977

Cleveland, W. S. (2001). Data Science: An Action Plan for Expanding the Technical Areas of. *International Statistical Review*. https://doi.org/10.1002/sam

Donoho, D. (2017). 50 Years of data science. *Journal of Computational and Graphical Statistics*. https://doi.org/10.1080/10618600.2017.1384734

Eads, D. (2018). The ticket trap: Front to back. ProPublica Illinois, March 4th. Available online at: https://www.propublica.org/nerds/the-ticket-trap-news-app-front-to-back-david-eads-propublica-illinois

Fairfield, J., & Shtein, H. (2014). Big data, big problems: Emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics: Exploring Questions of Media Morality*. https://doi.org/10.1080/08900523.2014.863126

Freeman Consulting (2013). A Future of failure: The flow of technology talent into government and society. Ford Foundation & John D. and Catherine T. McArthur Foundation

Ganapati, S., & Reddick, C. G. (2016). An ostrich burying its head in the sand? The 2009 NASPAA standards and scope of information technology and e-government curricula. *Journal of Public Affairs Education*. https://doi.org/10.1080/15236803.2016.12002245

George, G., Osinga, E., Lavie, D., & Scott, B. (2016). Big data and data science methods for management research. *Academy of Management Journal*. https://doi.org/10.5465/amj.2016.4005

Gil-Garcia, J. R., Pardo, T. A., & Nam, T. (2015). What makes a city smart? Identifying core components and proposing an integrative and comprehensive conceptualization. *Information Polity*. https://doi.org/10.3233/IP-150354

Helbig, N., Gil-Garcia, J. R., Ferro, E., Pardo, T. A., & Gil-Garcia, J. R. (2005). Understanding the Complexity of E-Government: Multi-method Approaches to Social Phenomena. *66th ASPA National Conference*.

Jaeger, P.T., Gorham, U., Bertot, J.C. & Taylor, N.G. (2015). Teaching information policy in the digital age: Issues, strategies, and innovation. *Journal of Education for Library and Information Science*, 56(3).

Keymolen, E., Prins, C., & Raab, C. (2012). Trust and ICT: New challenges for public administration. *Innovation and the Public Sector*. https://doi.org/10.3233/978-1-61499-137-3-21

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data applications in the government sector. *Communications of the ACM*. https://doi.org/10.1145/2500873

Kirkpatrick, K., 2016. Battling algorithmic bias: how do we ensure algorithms treat us fairly? Communications of the ACM. 59, 16–17. https://doi.org/10.1145/2983270

Kitchin, R. (2016). The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. https://doi.org/10.1098/rsta.2016.0115

Kraemer, K., Bergin, T., Bretschneider, S., Duncan, G., Foss, T., Gorr, W, Wish, N. (1986). Aurriculum recommendations for public management education in computing. final report of the National Association of Schools of Public Affairs and Administration, Ad Hoc Committee on Computers in Public Management Education. *Public Administration Review, 46*, 595-602. doi:10.2307/975583

Lavertu, Stéphane. 2016. We All Need Help: “Big data” and the mismeasure of public administration. *Public Administration Review*. 76(6): 864– 72.

Liebman, J. B. (2018). Using data to more rapidly address difficult u.s. social problems. *Annals of the American Academy of Political and Social Science*. https://doi.org/10.1177/0002716217745812

Matheus, R., Janssen, M., & Maheshwari, D. (2018). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. Government Information Quarterly. https://doi.org/10.1016/j.giq.2018.01.006

McNutt, J. G., Justice, J. B., Melitski, J. M., Ahn, M. J., Siddiqui, S. R., Carter, D. T., & Kline, A. D. (2016). The diffusion of civic technology and open government in the United States. *Information Polity*. https://doi.org/10.3233/IP-160385

Manoharan, A.P. and McQuiston, J.M. (2016). Technology and pedagogy: Information technology competencies in public administration and public policy programs. Journal of Public Affairs Education, 22(2), 175-186

McQuiston, J. M., & Manoharan, A. P. (2017). Developing e-Government coursework through the NASPAA Competencies Framework. *Teaching Public Administration*. https://doi.org/10.1177/0144739417690582

Mergel, I. (2010): Government 2.0 revisited: Social media strategies in the public sector, in: PA Times, American Society for Public Administration, Vol. 33, No. 3, p. 7 & 10.

Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*. https://doi.org/10.1111/puar.12625

Newcombe, Tod (2017). Drowning in data, cities turn to ‘citizen scientists.’ *Governing*, August 2017. Available at: https://www.governing.com/columns/tech-talk/gov-government-data-cities-citizen-scientists.html

New York University CUSP (2014). NYU’s Center for Urban Science and Progress graduates inaugural class. Available at: https://cusp.nyu.edu/press/nyus-center-urban-science-progress-graduates-inaugural-class/

Nielsen, P. A., & Moynihan, D. P. (2017). Romanticizing bureaucratic leadership? The politics of how elected officials attribute responsibility for performance. *Governance*. https://doi.org/10.1111/gove.12256

Olsen, J. (2015). Democratic order, autonomy, and accountability. *Governance*, 28(4), 425–440.

Press, G. (2013). A very short history of data science. *Forbes*. Available at: https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/

Roman, A. V. (2015). Framing the questions of e-government ethics: An organizational perspective. *American Review of Public Administration*. https://doi.org/10.1177/0275074013485809

Schrock, A. R. (2015). Civic hacking as data activism and advocacy: A history from publicity to open government data. *New Media and Society*. https://doi.org/10.1177/1461444816629469

Sadilek, A., Caty, S., DiPrete, L., Mansour, R., Schenk, T., Bergtholdt, M., Gabrilovich, E. (2018). Machine-learned epidemiology: real-time detection of foodborne illness at scale. *Npj Digital Medicine*. https://doi.org/10.1038/s41746-018-0045-1

Scassa, T. (2014). Privacy and open government. *Future Internet*. 6(2), 397-413. https://doi.org/10.3390/fi6020397

Wasney, M. (2017). The shots heard round the city. *Southside Weekly*, December 19, 2017. Available online: https://southsideweekly.com/shots-heard-round-city-shotspotter-chicago-police/

U.S. Department of Education, National Center for Education Statistics (2018). *Digest of Education Statistics, 2016* (NCES 2017-094)

U.S. Bureau of Labor Statistics (2018). *Occupational Outlook Handbook: Mathematicians and Statisticians*. Available online at: https://www.bls.gov/ooh/math/mathematicians-and-statisticians.htm

van der Aalst, W. (2016). Data science in action. In: *Process Mining*. Springer, Berlin, Heidelberg.