

Big Data and Evidence-Driven Decision-Making: Analyzing the Practices of Large and Mid-Sized US Cities

Alfred Tat-Kei Ho

School of Public Affairs and Administration, University of Kansas, Lawrence, Kansas, USA

Email: alfredho@ku.edu

Abstract

With the growing ease of collecting, transmitting, storing, processing, and analyzing massive amounts of data, Big Data has caught the attention of local officials in recent years. Based on a multi-layered institutional theories and an extensive analysis of the 30 largest cities and 35 selected mid-sized cities in the U.S, this study examines how U.S. cities are using mobile phone apps, sensors, data analytics, and open data portals to pursue Big Data opportunities, and what institutional factors influence their choices. The results show three distinct clusters of data practices among the selected 65 cities. Socio-demographics, cultural institutions, professional networks, and an internal data-driven culture as indicated by the use of performance budgeting are significantly associated with more extensive Big Data initiatives. The paper concludes by discussing the implications for Big Data practices and the theoretical development of e-government research.

1. Introduction

For the past two decades, e-government development has transformed governmental operations, public service delivery, and many government-citizen interactions [5, 17, 20]. Through digitalization, website development, and online transactions, e-government developments have the potential to not only make government operations more efficient, effective, responsive, and transparent [9, 35], but also more data-driven. For example, many administrative forms and transactions in the public sector are now fully digitized. Web-based services and network processing also generate a large volume of machine data [36]. New technologies, such as mobile phone apps and connected sensor technologies, provide even more sources of data on the output and outcomes of public policies and service delivery [18].

With these developments, access to a large volume of data is a reality for many local governments. From crime records, health records, social service records, library records, utility service records, public service requests, city websites and search histories, and social media content, to data, pictures, sound, and videos captured by cameras and sensors, local officials do not lack data to inform decision-making.

Local governments now have an unprecedented opportunity to pursue Big Data. “Big Data” refers to the use of a massive amount and diverse forms of data to conduct analysis so that the pattern and relationship of the data can be used for classification, clustering, anomaly detection, prediction, and other decision-making needs. The use of Big Data in the public sector is not only characterized by the complexity, volume, and velocity of data [7,23], but also by data democratization and multi-sectoral collaboration, since individual information technology departments are unlikely to have sufficient expertise and administrative capacity to realize the full potential of Big Data.

Based on an extensive analysis of the 30 largest cities and 35 selected mid-sized cities in the U.S., this study examines whether large and mid-sized cities in the U.S. are pursuing Big Data practices. It also analyzes the clustering of city practices from a multi-layered institutional perspective and tests whether the adoption of Big Data practices is influenced by the macro socio-demographic structure and cultural institution of a community, by involvement in professional networks across cities, or by the institutionalization of evidence-based decision-making within a city. The results of a two-step hierarchical cluster analysis and logistic regressions show that institutional factors at all three layers are important. The paper concludes by discussing the implications for future Big Data practices and the roles of national networks. It also discusses the implications for the theoretical development of e-government research.

2. Understanding Big Data from a Multi-Layered Institutional Perspective

Big Data is characterized by the volume, variety, complexity, velocity, and veracity of data [2, 4, 7, 12, 18, 23]. The Big Data phenomenon leads to new possibilities for evidence-based decision-making and strategic performance management, especially at the local level. For example, by aggregating multiple sources and diverse formats of data across policy fields, such as public safety and emergency services, neighborhood services, urban planning, 311 customer service, public health, local education, and local

infrastructure planning and management, local policymakers can better understand the complex dynamics across services and policies and break down departmental silos in strategic planning and program management [18, 19]. Also, by disaggregating the data across space and time and using predictive analytics to examine the data patterns for different neighborhoods, property parcels, or public service users, local managers can customize service planning and pursue proactive actions to serve the needs and priorities of different sub-groups within a community.

Behind these data-driven practices are a few institutional logics that sometimes conflict with the existing practices and norms of local governance, and can potentially create hurdles to the adoption of Big Data practices. The term “institutions” refers not only to the formal and informal processes and rules of organizations, but also the systems of norms and meanings that influence and constrain behaviors in society [24, 27, 38]. Therefore, institutional logics constitute the conceptual framework for analyzing the interrelationships among institutions, individuals, and organizations in social systems [41]. These logics provide the frames of reference that guide sense-making by actors, influence their choice of language and symbols in organizational activities, shape their reasoning in strategic decisions and action plans, and contextualize how power, symbols, rituals, and actions are interpreted.

In pursuing Big Data, managers adopt different practices, such as trying to digitize administrative records and agent activities as much as possible, working closely with departments to develop an inventory of data, developing mobile apps to collect detailed data about clients, and using sensors, social media, and the internet to track the preferences and activities of targeted populations [3,18]. They also invest resources in software, hardware, and staff power so that they can enhance their organizations’ ability to process and analyze data. Finally, and perhaps challenging the conventional perception that statistics and engineers are usually introverts, Big Data managers in the public sector, such as Chief Data Officers, are usually highly connected. They work regularly with peers within their city organization and partner with community stakeholders to develop strategic data plans, data policies, and pilot projects. Many also collaborate with civic hackers, such as programmers in Code for America Brigades in different cities, to pursue analytics initiatives and open data pilot programs [18].

These practices not only underscore the unique characteristics of Big Data that have already been suggested in the literature, such as the growing volume, complexity, variety, velocity, and

customization of data practices, but also suggest a potential paradigmatic shift or a new set of institutional logics in local management and governance. In this study, four institutional logics of Big Data are especially emphasized because they seem to drive the fundamental reasoning, symbolism, and rituals of Big Data practices among cities (see Figure 1).

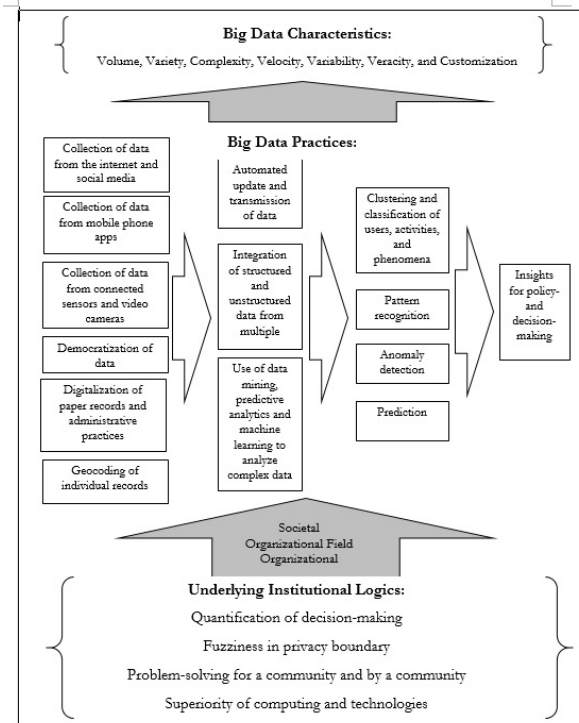


Figure 1. Institutional Logics and Practices of Big Data

Quantification of decision-making: Decision-making components can be quantified and outcomes can be optimized and visualized through statistical tools. This logic differs from the traditional logics of public decision-making, which allow for more ambiguity and focus more on procedural legitimacy, political representation, interest and power alignment, and gamesmanship.

Fuzziness in privacy boundaries: The volume and complexity of Big Data are built on the norm that many individual data can be collected, analyzed, reported, and shared either by governmental or non-governmental actors. In the traditional policymaking setting, these practices would have been unacceptable to many citizens. However, under the Big Data paradigm, individual citizens seem to be more tolerant of these practices, allow for greater fuzziness in privacy boundaries, and are more willing to share personal information, images, and activity records

with service providers in exchange for greater convenience, immediacy, connectedness, and cost-efficiency. These are new norms of data governance and suggest a new set of institutional logics that shift the boundaries of governmental and private activities.

Problem-solving for a community and by a community: There is a sub-culture among civic hackers and computer programmers who are interested in public policy issues that analytics should be used to advance the common good. This normative view of technologies is consistent with the value system of Millennials, who tend to be more civic-minded and want to use their talents and resources to make a difference in society. This sub-culture is also advocated and practiced by many high-tech companies in Silicon Valley, such as Google and Facebook. This normative perspective co-exists with the cognitive framing of public policymaking that community problems should be solved by a community, not just by the government. Stakeholder involvement and cross-sectoral partnership are also well-accepted principles in Big Data governance. These norms and principles of engagement may also emerge out of necessity in the public sector. Because of resource constraints and fiscal stress, many data managers in city governments have to leverage external expertise and volunteers to achieve their strategic data plans. As a result, Big Data practices in the public sector have a strong communal sentiment, and Big Data managers are not typical “bureaucrats”. They often do not believe that public policy problems can be addressed solely by governmental efforts and leadership, nor do they see that the solutions to community problems are “owned” by the government. Instead, they are more willing to work across sectoral and jurisdictional boundaries, and they are less hesitant to share ideas, tools, and solutions with colleagues within or even outside their organizations. These logics of engagement and interaction contradict the traditional logics of local governance, which tends to focus on government-led policy solutions, place-based identity, jurisdictional loyalty, and leadership by elected officials.

Superiority of computing and technologies: Finally, Big Data symbolizes the triumph of computing power and technology in society. It also signals a new possibility that computer-based problem solving may be superior to or more optimal than human decision-making [28]. This view is certainly controversial. However, it also has powerful implications, potentially changing the roles of human and machine in public decision-making and questioning some of the traditional values of local governance, such as the importance and necessity of

deliberative democracy and participatory decision-making by average citizens.

As discussed above, some of these Big Data institutional logics may not be acceptable to some members of society. There are also good reasons to question the appropriateness and validity of these logics. For example, instead of highlighting the efficiency gains of Big Data analytics, some researchers have cautioned about the potential threat of privacy violations or institutionalized discrimination through computer algorithms [22, 37]. Others question the supremacy and insightfulness of computing analytics and argue that Big Data may only generate a lot of “digital exhaust” [4, 25, 26]. There are also others who suggest that the public sector has its own unique organizational context, such as concerns about national security, its mission to protect individual liberty and privacy, and the legacy of departmental structure. As a result, there are inherent challenges in applying the business logics of Big Data to public policymaking and program management [21].

These controversies and debates suggest that while Big Data has become more technically feasible and less cost-prohibitive, there are other institutional factors in society, such as social norms of bureaucratic behavior, political ideologies, and public expectations of the roles and responsibilities of government that may influence the discretion and willingness of governmental agencies in the pursuit of Big Data. These potential conflicts between traditional institutional logics and Big Data logics may happen at three different levels: at the societal level, the professional field level, and at the organizational level within a city government.

For example, at the macro societal level, communities that uphold more traditional values of privacy, individual liberty, and limited government may be less enthusiastic about Big Data practices that involve the use of sensors, mobile phones, and web technologies by governmental agencies to track individuals’ activities and locations, or that use data analytics to classify residents and public service users. Communities that have a higher percentage of less-educated residents, or communities that have a high elderly population, may also show less appreciation and understanding of Big Data.

The struggle between different institutional logics in society can be reinforced or discouraged by organizational fields, such as professional networks. Through isomorphism, professional norms, information sharing, and peer pressure within a network, organizational actors are incentivized to adopt or resist certain practices in society or in their profession [8]. Hence, if the professional network of

city management questions the validity and reliability of quantified decision-making and remains steadfast about the supremacy of traditional governance logics, such as deliberative democracy, city leadership under the influence of the professional field are less likely to see the value of Big Data. On the other hand, if the professional network among city officials encourages the practice of Big Data and tries to empower its members through training, information sharing, capacity building exercises, national recognition, and other symbolism, then city managers under the field influence may feel more encouraged and positive about Big Data practices.

Finally, conflicts between traditional local governance logics and Big Data logics can happen within the organization of a city. For example, local governments with less fiscal capacity, less enthusiasm about the power of computing and quantitative analysis, and less willingness to embrace risk and change are also less likely to adopt Big Data as a new way of decision-making and a new paradigm of government-citizen-community relationship. On the other hand, if a city has a long-established tradition of evidence-based policymaking and welcomes the use of data in budgeting, program management, planning, the adoption of Big Data practices may become easier as it is consistent with the existing institutions of decision-making.

3. Research Hypotheses

Using the multi-layered institutional perspective of Big Data discussed above, this study hypothesizes the following:

About the impacts of societal institutions:

- 1a. Communities with a more-educated population are more progressive in adopting Big Data practices.
- 1b. Communities with a higher ratio of Millennials (between 18 and 34) in the population are more progressive in adopting Big Data practices.
- 1c. Communities with a higher ratio of civilian employees working in the information industry are more progressive in adopting Big Data practices.
- 1d. More populous communities are more progressive in adopting Big Data practices.
- 1e. Communities with a more traditionalist culture are less progressive in adopting Big Data practices.

About the impacts of organizational field:

- 2a. Communities that are members of the national network of What Works Cities initiated by

Bloomberg Philanthropies, are more progressive in adopting Big Data practices.

- 2b. Communities that are members of Results for America, are more progressive in adopting Big Data practices.
- 2c. Communities that are members of the White House's police data initiatives are more progressive in adopting Big Data practices.

About the impacts of organizational factors:

- 3a. Communities that have institutionalized data-driven decision-making, such as performance budgeting, are more progressive in adopting Big Data practices.
- 3b. Communities that have more resources to invest in information technologies are more progressive in adopting Big Data practices.
- 3c. Communities that have a team structure in data management are more progressive in adopting Big Data practices.
- 3d. Communities that have a mayoral form of government are more progressive in adopting Big Data practices.

Past studies have shown that the socio-demographic structure of a community and the influence of key external stakeholders, such as local businesses and community leaders, are an important impetus to kick-starting an e-government reform and overcoming internal resistance to change [39, 42, 45]. Therefore, this study hypothesizes that a higher ratio of information industrial workers, a higher ratio of Millennials, and a higher ratio of residents who have a graduate or professional degree are associated with more progressive adoption of Big Data practices. Also, this study hypothesizes that larger cities are more progressive in adopting Big Data practices. Larger cities face a greater need for innovative solutions to address more complex policy problems. As a result, they are usually more willing to adopt innovative reforms or technological change [29, 43, 44].

Another socio-institutional factor of interest is the political culture of different states. Past studies have found three major political cultures among U.S. states – moralistic, individualistic, and traditionalistic [10], and traditionalistic states are less likely to adopt policy innovation and have less policy-relevant competition [11, 15]. Since these findings are still valid in today's institutional environment and cities are nested within these state cultures, cities in traditionalistic states should be less interested in Big Data practices, especially when many Big Data tools and technologies are in the risky early stage of development, and when

Big Data has the potential to conflict with traditional values such as privacy and individual rights.

At the organizational field layer, several factors may help the development of Big Data. For the past few years, many mayors of major cities, such as Michael Bloomberg, the former Mayor of New York City and Martin O’ Malley, the former Mayor of Baltimore, have been very vocal about the need for evidence-based decision-making. New initiatives, such as the stat movement in cities and the “What Works Cities” and “Results for America” initiatives by Bloomberg Philanthropies, have been launched to support the development. The White House has also launched a number of smart city initiatives and the Police Data Initiative to help cities adopt data-driven policymaking and management. If implemented effectively, these national networks should encourage Big Data adoption by cities.

At the organizational level of analysis, organizational capacity issues, such as financial resources, staff capacity, data quality control, system compatibility and newness, and the ability to plan and engage other participants, are also important considerations in e-government practices [6, 13, 33]. This study hypothesizes that communities with a larger IT budget and a team structure in data management are more progressive in adopting Big Data practices. Also, communities that are practicing performance budgeting have already institutionalized the culture of data-driven decision-making. They are also more likely to have a robust data system to track and measure program activities and outcomes. All these may reduce the organizational hurdles to Big Data and so these city governments should be more ready to adopt the practice.

Finally, some past studies have found that partisan rivalry is favorable to e-government development because electoral competition motivates policymakers to cut taxes and reduce the size of the government bureaucracy. This in turn puts pressure on policymakers to look for technology-enabled alternatives to deliver essential public services more cost-effectively [40, 42]. Furthermore, as discussed above, there is a growing network among mayors of major cities to push for evidence-based decision-making. Hence, either because of political pressure to innovate, or because of the impact of isomorphism, the mayoral form of government should be more willing to embrace the idea and practices of Big Data.

4. Data and Methodologies

To test the above hypotheses, the 30 most populous cities and a sample of 35 cities that have a population between 100,000 and 500,000 are selected as the subjects of analysis (see Table 1). While one factor in

the selection of these cities was the need to ensure that their population range was reasonably diverse, some of the cities were selected because of their participation in different national networks, such as What Works Cities and the White House’s police data initiative at the end of 2015. Other cities were selected because even though they are located near those cities that are nationally recognized as leaders in the Big Data movement, these cities did not participate in any national network of data initiatives and did not seem to show much interest in data analytics or open data portal development. Hence, these cities were included to act as the control group in the analysis.

Table 1. List of Selected Cities

Cityname	Population in 2014	City population ranking in the US, 2014	Form of Government	Cityname	Population in 2014	City population ranking in the US, 2014	Form of Government
New York, NY	8,491,079	1	Mayor-Council	Fresno, CA	515,986	34	Mayor-Council
Los Angeles, CA	3,928,864	2	Mayor-Council	Sacramento, CA	485,199	35	Council-Manager
Chicago, IL	2,722,389	3	Mayor-Council	Kansas City, MO	470,800	37	Council-Manager
Houston, TX	2,239,558	4	Mayor-Council	Mesa, AZ	464,704	38	Mayor-Council
Philadelphia, PA	1,560,297	5	Mayor-Council	Atlanta, GA	456,002	39	Mayor-Council
Phoenix, AZ	1,537,058	6	Council-Manager	Colorado Springs, CO	445,830	42	Mayor-Council
San Antonio, TX	1,436,697	7	Council-Manager	Miami, FL	430,332	44	Mayor-Council
San Diego, CA	1,381,069	8	Mayor-Council	Oakland, CA	413,775	45	Mayor-Council
Dallas, TX	1,261,047	9	Council-Manager	Tulsa, OK	399,682	47	Mayor-Council
San Jose, CA	1,015,785	10	Council-Manager	New Orleans, LA	384,320	50	Mayor-Council
Austin, TX	912,791	11	Council-Manager	Lexington, KY	310,797	61	Urban County
Jacksonville, FL	853,382	12	Mayor-Council	Cincinnati, OH	298,165	65	Mayor-Council
San Francisco, CA	852,469	13	Mayor-Council	Saint Paul, MN	297,640	66	Mayor-Council
Indianapolis, IN	848,788	14	Mayor-Council	Jersey City, NJ	262,146	74	Mayor-Council
Columbus, OH	835,957	15	Mayor-Council	Fort Wayne, IN	258,522	77	Mayor-Council
Fort Worth, TX	812,238	16	Council-Manager	Durham, NC	251,893	81	Council-Manager
Charlotte, NC	809,958	17	Council-Manager	Baton Rouge, LA	228,895	96	Mayor-Council
Detroit, MI	680,250	18	Mayor-Council	Rochester, NY	209,983	103	Mayor-Council
El Paso, TX	679,036	19	Council-Manager	Salt Lake City, UT	190,884	124	Mayor-Council
Seattle, WA	668,342	20	Mayor-Council	Worcester, MA	183,016	131	Council-Manager
Denver, CO	663,862	21	Mayor-Council	Providence, RI	179,154	134	Mayor-Council
Washington DC	658,893	22	Mayor-Council	Fort Lauderdale, FL	176,013	135	Council-Manager
Memphis, TN	656,861	23	Mayor-Council	Chattanooga, TN	173,778	141	Mayor-Council
Boston, MA	655,884	24	Mayor-Council	Jackson, MS	171,155	143	Mayor-Council
Nashville, TN	644,014	25	Mayor-Council	Carly, NC	155,217	160	Council-Manager
Baltimore, MD	622,783	26	Mayor-Council	Sunnyvale, CA	149,980	168	Council-Manager
Oklahoma City	620,602	27	Council-Manager	Hampton, VA	136,879	189	Council-Manager
Portland, OR	619,360	28	Commission	Columbia, SC	132,067	195	Council-Manager
Las Vegas, NV	613,599	29	Council-Manager	Topka, KS	127,215	213	Council-Manager
Louisville, KY	612,780	30	Mayor-Council	Hartford, CT	124,705	218	Mayor-Council
Albuquerque, NM	557,169	32	Mayor-Council	Independence, MO	117,494	231	Council-Manager
Tucson, AZ	527,972	33	Council-Manager	Rochester, MN	111,402	249	Mayor-Council
				Richmond, CA	108,565	265	Council-Manager

In the winter of 2015 and spring 2016, extensive online research of these cities was done to understand whether these cities had any special sensor and analytics initiatives, whether their budgets contained any performance measures, whether they had any open data portals, and whether they offered any mobile phone apps on their websites. Executive orders related to open data and data analytics initiatives, policy documents related to data privacy concerns, and any inter-governmental agreements on data collaboration were also collected and analyzed.

To verify whether the online research results were valid and to collect more information about the internal practices of data management and evidence-driven decision-making, a copy of our preliminary research findings was sent to the responsible staff in the information technology departments or to the chief data officers, if available, of all 65 cities so that they could provide feedback and data updates from April to June, 2016. More than twenty phone and in-person interviews with officials responsible for their cities’

data initiatives were also conducted in fall 2015 and spring 2016.

To test whether certain data-driven practices cluster together among the 65 selected cities, a two-step clustering analysis is used. In the first step, various practices are transformed into 24 binary variables (yes/no) about specific data practices (see Table 2 later). This step avoids different scaling of variables, which may cause biases in the clustering results. Then a cluster analysis is conducted using a squared Euclidean distance matrix created from the binary variables using simple matching coefficients.

The results of cluster analysis produce different groupings of cities based on their evidence-based decision-making practices. Logistic regressions are then used to analyze whether the hypothesized institutional, demographic, and organizational factors have any significant impact on the extensiveness of city data initiatives.

5. Findings

Among the 65 cities included in the study, 60 offer mobile phone apps to engage the public. Among these 60 cities, 24 offer one smartphone app, another 23 cities offer two to five apps, and some of the larger cities offer more than five apps.

Connected sensors and video cameras offer another new source of Big Data for local governments. While the use of sensors is not a recent phenomenon, the advancement of wireless technologies and sensor devices opens up new possibilities for data collection. Today, sensors and video cameras can collect all kinds of data, including temperature, air quality, pedestrian and vehicle movement, audio, and video images, and can transmit the data instantly by wireless or broadband technologies for real-time analytics purposes.

For example, in 2012, New York City used monitoring sensors to detect elevated flow levels in sewer pipes and alert city staff to perform inspections and preventive maintenance [31]. In 2013, the city experimented with remote sensor technology to monitor the frequency and volume of sewer overflows to guide future infrastructure investment [32]. In 2014, the City Government of Chicago, the University of Chicago, the School of the Art Institute of Chicago, and Argonne National Laboratory launched a new project using a network of 40 sensor nodes installed on lampposts to collect data on weather, air quality, light intensity, and the number of Wi-Fi and Bluetooth devices within a 100-foot range. The data will then be integrated and analyzed to understand pedestrian movements and public health concerns [1]. Kansas City, Missouri also has a pilot program using connected sensors in one of its downtown areas to

measure vacancies in parking lots, the impact of snowfall on the ground, and the volume of pedestrian traffic so that city officials can understand the service needs of different areas instantly.

Out of the selected 65 cities, 52 (80 percent) have some form of connected sensor initiatives (see Figure 2). More than half of them (29 cities) are related to weather and environmental monitoring. Twenty-seven cities (42 percent) also use connected sensors to improve their transit and transportation system. Initiatives related to public safety and justice or smart parking management are also popular. The use of sensors for smart government buildings is relatively less. Only seven cities have reported such initiatives.

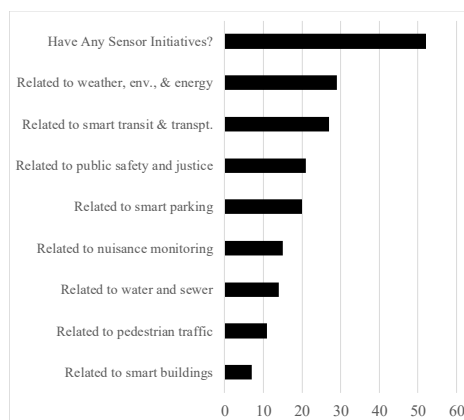


Figure 2. Number of Cities with Sensor Initiatives

Analytics initiatives are another important component that supports the decision-making subsystem of evidence-based policymaking. Based on web and policy document research, 49 cities (75 percent) have reported some form of data analytics initiatives (see Figure 3). The majority of cities rely on their information technology departments to take the lead in these initiatives. However, a few cities also have other departments to co-lead these initiatives with information technology departments or even to take the leading coordinator role. For example, six cities have involved the City Manager's Office or the Mayor's Office, and six cities let their performance management units coordinate various departments to pursue data analytics initiatives. Fifteen cities also have designated Chief Data Officers. To foster coordination among departments and provide general policy direction for the city-wide data initiatives, 30 cities have established a multi-departmental committee or team structure. Also, 28 cities (43 percent) have used a partnership with Code for America to launch pilot analytics programs.

It should be noted that the scope, approach, and the degree of sophistication of data analytics vary significantly among cities. Some have more elaborate

initiatives, involving multiple departments, programs, and external stakeholders, while some may only have one or two initiatives and only one department, such as the police, may be actively involved in analytics programs.

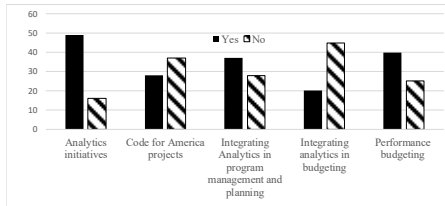


Figure 3. Number of Cities with Analytics Initiatives

Among different analytics initiatives, 37 cities have initiatives related to specific programs, such as public safety, or nuisance control. Only 20 cities have initiatives focusing on budgeting, such as revenue forecasting and program spending analysis. These numbers are fewer than the number of cities that practice performance budgeting (n = 40). This shows that while the majority of the 65 cities are using data for performance measurement purposes, not all have moved to the next step and utilized more sophisticated analytics to understand and analyze performance data to inform budgetary decision-making.

Among the 65 cities studied, 50 cities (77 percent) have an open data portal (see Figure 4). This allows them to use data to engage the public and encourage crowdsourcing ideas and citizen-initiated service delivery [16]. It should be noted that most of these open data initiatives are still city-focused.

Only 20 cities collaborate with their counties and post county data on their portals, and another 20 cities collaborate with other local governments, primarily school districts. Collaboration with federal agencies is rare. Only three cities have such an initiative. These findings show that even though open data initiatives have the potential to break the departmental and city silos and allow more inter-governmental collaboration, data sharing and interoperability is still a challenge [30, 34].

Because privacy is an important public concern in the development of Big Data [14], city policymakers and managers need to pay attention to this governance issue. However, among the 65 cities examined, only 34 cities have posted any privacy policies on their open-data portals. Even if there is a policy, some cities do not have a well-developed one and many simply refer to generic policy language provided by the private contractors of their open data portals. Only 27 of the surveyed cities (42 percent) had their own privacy policies and posted them online. This suggests that many important data governance issues, including privacy, are still under-explored by city management

and have not yet received sufficient formal attention from local policymakers.

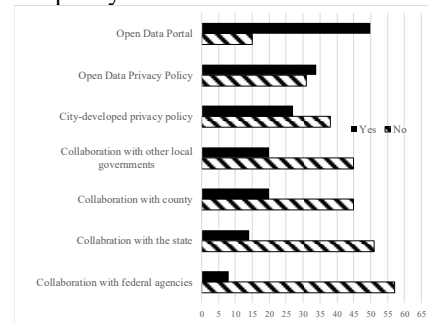


Figure 4. Open Data Practices

The results of the cluster analysis are shown in Figure 5. Three clusters have emerged out of 24 data practices. Table 2 shows the characteristics of these three clusters. The first cluster, including Jacksonville, Houston, New York, and 15 other cities, have relatively high use of all Big Data practices, such as mobile phone apps, sensors, analytics programs, and open data practices. The second cluster, including Louisville, Jackson, Boston, San Diego, Seattle, and ten other cities, have selective sensor initiatives, strong open data presence, and selected analytics programs. The third cluster of cities (n=32) tends to have less depth and breadth in Big Data practices. Still, the majority of them have an open data portal, 34 percent have analytics initiatives in program management, many of which are led by the police or some service departments.

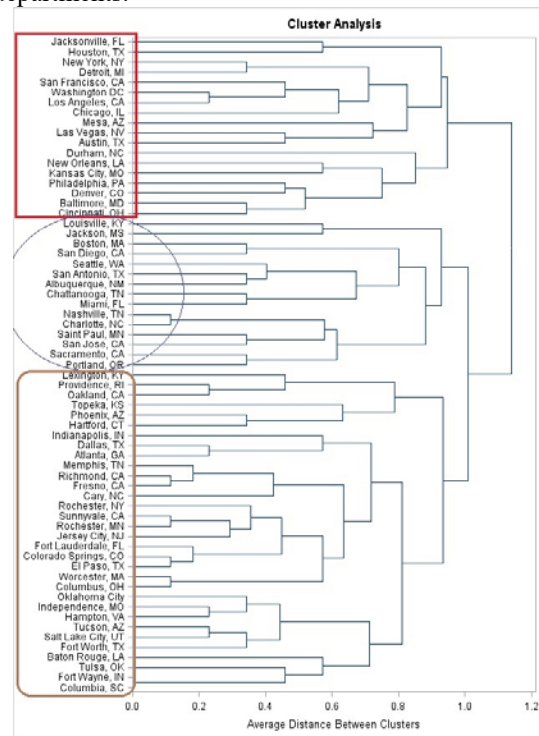


Figure 5. Cluster Analysis of City Data Initiatives

Table 2. Characteristics of the Three Clusters

Practice	Cluster 1		Cluster 2		Cluster 3	
	N	Mean	N	Mean	N	Mean
City app: 0	18	0%	15	13%	32	9%
Cityapp: 1	18	28%	15	0%	32	59%
Cityapp: 2-3	18	22%	15	13%	32	31%
Cityapp: 4-5	18	6%	15	40%	32	0%
City app: 6-10	18	11%	15	33%	32	0%
City app: > 10	18	33%	15	0%	32	0%
city developing own apps	18	78%	15	33%	32	22%
Have a central app directory	18	61%	15	67%	32	22%
Sensor initiatives: 0	18	6%	15	7%	32	34%
Sensor initiative: 1	18	6%	15	7%	32	25%
Sensor initiatives: 2-3	18	61%	15	20%	32	28%
Sensor initiatives: 4-5	18	6%	15	67%	32	6%
Sensor initiatives: >5	18	22%	15	0%	32	6%
Code for America	18	72%	15	67%	32	16%
Analytics focus: program management and planning	18	100%	15	53%	32	34%
Analytics focus: budgeting	18	56%	15	33%	32	16%
Having a chief data officer	18	72%	15	13%	32	0%
Have an executive order and policies on analytics programs	18	39%	15	7%	32	0%
Open data portal	18	94%	15	93%	32	59%
Executive orders on open data	18	83%	15	73%	32	25%
Performance dashboard	18	100%	15	93%	32	56%
Open data privacy policy	18	67%	15	40%	32	28%
Open data collaboration with the state	18	50%	15	7%	32	13%
Open data collaboration with the federal government	18	22%	15	7%	32	9%

A logistic regression is used to analyze the impacts of institutional and organizational factors behind these patterns of clustering. The dependent variable is whether a city belongs to the high data initiative clusters (clusters 1 or 2) or the relatively lower data initiative cluster (cluster 3), with high usage coded as 1 and low usage coded as 0.

The results confirm the hypotheses based on multi-layered institutional theories (see Table 3). At the macro-societal level, communities with more educated residents and more populous cities are more likely to be progressive users of Big Data. Cities in traditional state cultures, however, are significantly less likely to do so, after controlling for other factors.

As hypothesized, the professional network, “What Works Cities” provides significant support for cities’ adoption of Big Data practices. The astronomically high odd ratio of the logistic regression results shows that once a city has joined the network, it is almost certain that the city will adopt some practices of Big Data, either because of the support they receive or because of peer pressure and isomorphism.

Within a city, the institutionalization of data-driven decision-making through performance budgeting is also a positive and significant factor. Having a team structure also helps. This is consistent

with the institutional logics of Big Data that analytics are used for and by communities.

Table 3. Logistic Regression Results

	Coefficient	Standard error		Odd ratio
Societal Institutional Factors:				
Percentage of population with graduate or professional degree	0.32	0.16*		1.37
Percentage of population aged 18-34	-0.21	0.18		
Percentage of civilian employees working in the information industry	-0.56	0.75		
City population (in log)	3.36	1.11***		28.70
Traditionalistic state culture	-3.85	1.77**		0.02
Individualistic state culture	-0.80	1.39		
Organizational Field Factors:				
What Works Cities Initiative 2015	8.65	3.55**		5693.40
Results for America 2015	-0.07	1.53		
White House Police Data Initiative 2015	1.22	1.68		
Organizational Factors:				
Have performance budgeting	4.77	2.19**		118.11
Information technology spending per-capita	-0.01	0.02		
Have a data team structure	2.03	1.18*		7.62
Involvement of city manager's or mayor's office	-0.28	1.71		
Mayor council form of government	3.02	1.60*		20.44
Notes: R-square = 0.56; Max-rescaled R-square = 0.74. The intercept is not reported here Likelihood ratio: 53.16*** Wald statistics: 13.24 *** * significant at the 10-percent level. ** significant at the 5-percent level. *** significant at the 1-percent level.				

Finally, the mayoral form of government is also positively associated with Big Data practices, after controlling for demographic differences among cities, the effect of national networks, and other managerial strategies of data initiatives. In the interview results with data officials of different cities, many pointed out that mayors are important leaders of Big Data initiatives. If they set a clear vision and become the champions of the initiatives, they can rally community support and help cities overcome some of institutional resistance to Big Data, such as departmental silos and refusal to adopt evidence-based logics in decision-making.

It should be pointed out that administrative capacity or resource availability, which is measured by the amount of information technology (IT) spending per-capita, is not statistically significant. However, interviews with city officials show that difficulties in hiring competent staff and insufficient training for existing staff are perceived to be the largest barriers to Big Data. Hence, even though IT spending per-capita of a city is not significant here, this does not mean that capacity constraints or inadequate investment in staffing for Big Data initiatives is not a challenge.

6. Discussion and Conclusion

Using data from 65 large and mid-sized cities in the U.S., this study analyzes how city governments are

using Big Data tools to engage the public and improve program management. It also applies a theoretical framework based on theories of multi-layered institutionalism to analyze empirically how different societal, network, and organizational factors influence the practices of cities. The results show that the largest and mid-sized cities in the U.S. have indeed adopted Big Data practices, such as collecting and analyzing various types of sensor data, engagement of the public through open data and crowdsourcing, and integrating data analytics into program management and budgeting. The results show diverse practices among cities, and they can be categorized broadly into three major groups with different emphases on the usage of sensors, analytics, mobile phone apps, and open data platforms. Many societal and organizational factors influence these cities' practices. However, the most impactful factors, as indicated by the odd ratios of the logistic regression results, are the membership in the What Works Cities network and the presence of performance budgeting practices. Both show that if a city can institutionalize a culture of evidence-based decision-making, it is more likely to embrace Big Data.

These results should be encouraging to policymakers and advocates for Big Data and evidence-based policymaking. While it is true that macro-institutional factors, such as the demographic characteristics and historical culture of communities, have some influence on cities' willingness to embrace Big Data, professional networks and managerial practices can make a difference and can over-compensate for an inertia to change. Practices by national organizations and the federal government, such as the What Works Cities program by Bloomberg Philanthropies and the White House's smart city initiatives, are especially some good examples.

This study also highlights the importance of understanding Big Data development from a multi-institutional layers perspective and the need to compare the relative impact of societal, organizational field, and organizational factors. This study is just a preliminary effort to use this type of institutional perspective to analyze e-government practices, such as Big Data, in U.S. cities. Future studies may look into this theoretical framework more carefully and analyze how different layers of institutional factors may interact and impact each other. This study also points out the positive and significant impacts of mayors. Future studies may further examine the role of leadership and understand how mayors and their designates influence the practices of Big Data in different institutional and governance contexts.

Acknowledgment: This research received funding by the IBM Center for the Business of Government, which supports public management research by academic and nonprofit entities. The author thanks Bo McCall for ample research assistance. All errors remain the responsibilities of the author.

References

- [1] Burns, J. (2014). "Chicago will Use Sensors to Gather Data on City and its Residents." <http://www.psfk.com/2014/10/chicago-sensor-monitoring-residents.html> (assessed March 1, 2016).
- [2] Chen, Y.-C., and Hsieh, T.-C. (2014). "Big Data for Digital Government: Opportunities, Challenges, and Strategies." *International Journal of Public Administration in the Digital Age*, 1(1), pp.1-14.
- [3] City of Boston (2015). "Street Bump: Help Improve Your Streets." <http://www.cityofboston.gov/doi/apps/streetbump.asp> (accessed March 1, 2016).
- [4] Clark, W. R., and Golder, M. (2014). "Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?" *PS: Political Science and Politics*, 48(1), 65-70.
- [5] Cordella, A., and Tempini, N. (2015). "E-government and organizational change: Reappraising the Role of ICT and Bureaucracy in Public Service Delivery," *Government Information Quarterly*, 32(3), pp. 279-286.
- [6] Coursey, D., and Norris, D.F. (2008). "Models of E-government: Are They Correct? An Empirical Assessment." *Public Administration Review*, 68(3), pp. 523-535.
- [7] Desouza, K. C. (2014). *Realizing the Promise of Big Data: Implementing Big Data Projects*, Washington, D.C.: The IBM Center for the Business of Government.
- [8] DiMaggio, P. J., and Powell, W.W. (1983). "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review*, 48: 147-160.
- [9] Domínguez, L.R., García-Sánchez, I. M., and Álvarez, I. G. (2011), "Determining Factors of E-Government Development: A Worldwide National Approach." *International Public Management Journal*, 14(2), pp.218-248.
- [10] Elazar, D. J. (1984). *American Federalism: A View from the States*, 3rd ed, New York: Harper and Row.
- [11] Fitzpatrick, J. L. and Hero, R. E. (1988). "Political Culture and Political Characteristics of the American States: A Consideration of Some Old and New Questions." *Western Political Quarterly*, 41(1), pp. 145-153.
- [12] Gartner (2011). *Pattern-Based Strategy: Getting Value from Big Data*. <https://www.gartner.com/doc/1727419/> (Accessed March 1, 2016).
- [13] Gil-Garcia, J. R., and Pardo, T. A. (2005). "E-Government Success Factors: Mapping Practical Tools to Theoretical Foundations." *Government Information Quarterly*, 22(2), pp. 187-216.

- [14] Goldsmith, S. (2015), "Protecting Big Data: Seattle's Digital Privacy Initiative Aims to Keep Innovation on Track with New Data Safeguards." *Government Technology*, September 9. Available at <http://www.govtech.com/data/Protecting-Big-Data.html>.
- [15] Hanson, R. L. (1991). "Political Cultural Variations in State Economic Development Policy." *Publius*, 21(2), pp. 63-81.
- [16] Harrison, T. M., Guerrero, S., Burke, G.B., Cook, M., Cresswell, A., Helbig, N., Hrdinová, J., and Pardo, T. (2012), "Open Government and E-Government: Democratic Challenges from a Public Value Perspective." *Information Polity*, 17(2), pp. 83-97.
- [17] Ho, A. T-K. (2001). "Reinventing Local Governments and the E-Government Initiative." *Public Administration Review*, 62(4), pp. 434-444.
- [18] Ho, A. T-K. (2016). *Ten Actions to Implement Big Data Initiatives: A Study of 65 Cities*. Washington, D.C.: The BM Center for the Business of Government.
- [19] Ho, A. T-K., and Cho, W. (2016). "Government Communication Effectiveness and Satisfaction with Police Performance: A Large-Scale Survey Study." *Public Administration Review*, <http://dx.doi.org/10.1111/puar.12563>.
- [20] Janowski, T. (2015). "Digital Government Evolution: From Transformation to Contextualization." *Government Information Quarterly*, 32(3), pp. 221-236.
- [21] Kim, G-H., Trimi, S., and Chung, J-H. (2014). "Big Data Applications in the Government Sector." *Communications of the ACM*, 57(3), 78-85.
- [22] Kirchner, L. (2015). "When Big Data Becomes Bad Data: Corporations are Increasingly Relying on Algorithms to Make Business Decisions and That Raises New Legal Questions." *ProPublica*, September 2. <https://www.propublica.org/article/when-big-data-becomes-bad-data> (accessed May 25, 2016).
- [23] Laney, D. (2001). "3D Data Management: Controlling Data Volume, Velocity, and Variety." *Meta Group Application Delivery Strategies*, February 6.
- [24] March, J. G., & Olsen, J. P. (1989). *Rediscovering Institutions: The Organizational Basis of Politics*. Simon and Schuster.
- [25] Meier, K. J., and O'Toole, L. J. (2005). "Managerial Networking Issues of Measurement and Research Design." *Administration and Society*, 37(5), 523-541.
- [26] Mergel, I., Rethemeyer, R. K., and Isett, K. (2016). "Big Data in Public Affairs." *Public Administration Review*. <http://dx.doi.org/10.1111/puar.12625>.
- [27] Meyer J. W., & Rowan, B. (1977). "Institutionalized organizations: Formal structure as myth and ceremony." *American Journal of Sociology*, 83: 340-363.
- [28] Monroe, B.L., Pan, J., Roberts, M. E., Sen, M., and Sinclair, B. (2015), "No! Formal theory, Causal Inference, and Big Data are Not Contradictory Trends in Political Science." *PS: Political Science & Politics*, 48(1), pp.71-74.
- [29] Moon, M. J., and Norris, D. F. (2005), "Does Managerial Orientation Matter? The Adoption of Reinventing Government and E-Government at the Municipal Level." *Information System Journal*, 15(1), pp. 43-60.
- [30] Nam, T., and Pardo, T. A. (2011). "Smart city as urban innovation: Focusing on management, policy, and context." *Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance*, pp. 185-194.
- [31] New York City. 2012. "DEP Installs Manhole Monitoring Sensors to Detect Elevated Flow Levels." http://www.nyc.gov/html/dep/html/press_releases/12-63pr.shtml#.VucC5uYzQeE (accessed March 1, 2016).
- [32] New York City. 2013. "Department of Environmental Protection Installs Remote Technology to Help Guide Future Capital Investments and Further Improve Harbor Water Quality." http://www.nyc.gov/html/dep/html/press_releases/13-010pr.shtml#.VucD_uYzQeE. Accessed March 1, 2016.
- [33] Pardo, T. A., and Taewoo N. (2011). "Conceptualizing Smart City with Dimensions of Technology, People, and Institutions." *Proceedings of the 12th Annual International Digital Government Research Conference*.
- [34] Pardo, T. A., Nam, T., and Burke, G. B. (2012). "E-government Interoperability: Interaction of Policy, Management, and Technology Dimensions." *Social Science Computer Review*, 30(1), pp. 7-23.
- [35] Pina, V., Torres, L. and Royo, S. (2010). "Is E-government Promoting Convergence Towards More Accountable Local Governments?" *International Public Management Journal*, 13(4), pp. 350-380.
- [36] Power, D. J. (2014). "Using Big Data for analytics and decision support," *Journal of Decision Systems* 23(2), pp. 222-228.
- [37] Schrage, M. (2014). "Big Data's Dangerous New Era of Discrimination." *Harvard Business Review*, <http://blogs.hbr.org/2014/01/big-datas-dangerous-new-era-of-discrimination>.
- [38] Scott, W. R. (1994). Institutions and Organizations: Toward a Theoretical Synthesis. In Scott, W. R., & Meyer, J. W. (Eds), *Institutional Environments and Organizations: Structural Complexity and Individualism* (pp. 55-80). Thousand Oaks, CA: Sage.
- [39] Siau, K., and Long, Y. (2006), "Using Social Development Lenses to Understand E-government Development." *Journal of Global Information Management*, 14(1), pp. 47-62.
- [40] Siau, K., and Long, Y. (2009), "Factors Impacting E-government Development." *Journal of Computer Information Systems*, 50(1), pp. 98-107.
- [41] Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). *The Institutional Logics Perspective: A New Approach to Culture, Structure, and Process*. Oxford: Oxford University Press.
- [42] Tolbert, C. J., Mossberger, K., and McNeal, R. (2008). "Institutions, Policy and E-Government in the American States." *Public Administration Review*, 68(3), pp. 549-563.
- [43] Weare, C., Musso, J., and Hale, M. (1999), "Electronic Democracy and the Diffusion of Municipal Web Pages in California." *Administration and Society*, 31(1), pp. 3-27.
- [44] West, D. (2001). *State and Federal Government E-Government in the United States 2001*. <http://www.insidepolitics.org/policyreports.html>
- [45] West, D. (2004). "E-government and the Transformation of Service Delivery and Citizen Attitudes." *Public Administration Review*, 64(1), pp. 15-27.