The Data Firehose and AI in Government: Why Data Management is a Key to Value and Ethics

Teresa M. Harrison
University at Albany
Albany, NY
tharrison@albany.edu

Luis F. Luna-Reyes
University at Albany
Albany, NY
lluna-reyes@albany.edu

Theresa A. Pardo
CTG UAlbany
Albany, NY
tpardo@ctg.albany.edu

Nic DePaula
University at Albany
Albany, NY
ndepaula@albany.edu

Mahdi M. Najafabadi
University at Albany
Albany, NY
mnajafabadi@albany.edu

Jillian M. Palmer
CTG UAlbany
Albany, NY
jmpalmer@ctg.albany.edu

ABSTRACT

Technical and organizational innovations such as Open Data, Internet of Things and Big Data have fueled renewed interest in policy analytics in the public sector. This revamped version of policy analysis continues the long-standing tradition of applying statistical modeling to better understand policy effects and decision making, but also incorporates other computational approaches such as artificial intelligence (AI) and computer simulation. Although much attention has been given to the development of capabilities for data analysis, there is much less attention to understanding the role of data management in a context of AI in government. In this paper, we argue that data management capabilities are foundational to data analysis of any kind, but even more important in the present AI context. This is so because without proper data management, simply acquiring data or systems will not produce desired outcomes. We also argue that realizing the potential of AI for social good relies on investments specifically focused on this social outcome, investments in the processes of building trust in government data, and ensuring the data are ready and suitable for use, for both immediate and future uses.

CCS CONCEPTS

• **Applied computing** → **E-government**; • **Information systems** → *Data analytics*; Data warehouses.

KEYWORDS

Data Management, Artificial Intelligence, DMBOK, Data Analytics, Policy Analysis

ACM Reference Format:
Teresa M. Harrison, Luis F. Luna-Reyes, Theresa A. Pardo, Nic DePaula, Mahdi M. Najafabadi, and Jillian M. Palmer. 2019. The Data Firehose and AI in Government: Why Data Management is a Key to Value and Ethics.

dg.o 2019, June 18–20, 2019, Dubai, United Arab Emirates

ACM ISBN 978-1-4503-7204-6/19/06…$15.00
https://doi.org/10.1145/3325112.3325245

In *dg.o 2019: 20th Annual International Conference on Digital Government Research (dg.o 2019), June 18–20, 2019, Dubai, United Arab Emirates.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3325112.3325245

1 INTRODUCTION

Never has more data been available to digital government researchers and practitioners than now in light of what Open Data, the Internet of Things, and Big Data have brought. Further, as the theme of this year's DG conference suggests, excitement about artificial intelligence (AI) and its associated computing strategies (e.g., machine learning, k-means clustering, natural language processing, neural networks) is growing, as decision makers at nearly all government levels appreciate what AI offers in terms of improving government efficiencies and supporting policy analysis. The vast streams of data now available to government combined with powerful computational strategies can be valuable tools as well as potentially disruptive technologies that can revitalize our long-standing tradition of using statistical modeling to better understand policy effects and guide decision making, but that may also challenge democratic and other societal institutions in significant ways [3]. However, although much attention is now devoted to improving our capabilities in data analytics, much less attention is devoted to improving our capabilities in data management. In this paper, we discuss the role of data governance and how data management capabilities may be conceptualized in the context of advanced analytics and artificial intelligence in government agencies. We also identify challenges for data management for AI in the public sector. First, in this introduction, we discuss the relationship between AI and data, and issues in the adoption of AI in government organizations.

1.1 The relationship between AI and data

Although not all AI systems need data (i.e. recorded information from users and the environment), in current knowledge-based and machine learning systems, data are a necessary component of AI [7][15]. Piletic [29] notes that "AI cannot go anywhere without big data." The data must be "big" because AI requires large datasets in order to learn. The examples of the virtual assistants Cortana and Siri, as Piletic describes, have been successful because of their access to the endless amounts of information provided by users, upon which systems may learn natural language. Other AI technologies have similar data requirements [7]. For AI devices, the more information available to process, the more the system can

learn and increase accuracy [21][23]. However, we should note, data is useless without appropriate analytical strategies. The devices connected to the Internet of Things, for example, produce a never-ending tsunami of data from millions of devices, but all that amounts to little without being harnessed by the software, algorithms, and intelligence that can yield insight.

1.2 The role of AI in governments and industry

Given the appropriate data resources, AI computational strategies can accomplish significant and useful tasks for government. AI can be used to automate repetitive tasks and in so doing free up government labor for other tasks, increase the speed of transactions in the provision of government services, and more accurately assess the outcomes of policy options, to name a few [24]. In the private sector, there is growing recognition about the role of big data and AI in their enterprises. A 2018 survey of Fortune 1000 executives, 77% of which were from financial services companies, reports that 97% of them are investing in building big data and AI initiatives [4][28]. Of the participants, 77% have responded that the growing availability of data "is empowering AI and cognitive initiatives within their organizations" [4]. The financial services industry has been at the forefront of the development of robust data management over a period of decades. However, industries such as life sciences, while possessing vast amounts of scientific and patient data, are new to enterprise data management [4]. Government agencies may be similar in that they possess large amounts of sensitive data; however, this data may not be managed well enough for AI applications [24].

1.3 Issues of AI in governments

Providing a sound foundation for building AI systems is no easy accomplishment. Part of the problem may be that government organizations are not equipped with appropriately curated data resources, computing systems, and human expertise to mount AI initiatives [2]. AI also poses the potential for ethical quandaries, because AI strategies sometimes produce outcomes that are not well understood or integrated into governmental agendas and can thus run afoul of social and political values of government institutions and democracy more broadly. ProPublica's recent critique of the bias produced by AI software used in making prison sentencing recommendations is a telling case in point [1]. DeSouza [10] proposes that "leaders must design and implement governance and policy that promotes a skilled workforce that collaborates with academia and the private sector, risk management frameworks, secure systems, and modern technologies" (p. 5). An essential task for those wishing to exploit the potential of AI is to determine what governance and policy actions are required to establish this foundation of appropriate skills, collaborations, ethical and values considerations.

An anecdote may help illustrate the problem. One of the authors of this paper was asked to keynote at a conference for the launch of a new book on Big Data, privacy and security. The room was full of computer and data scientists who were driving the field of big data analytics. Surprisingly, the topic they asked the keynoter to address was the problem they were beginning to experience in their efforts to get access to and use data to create the kind of value they imagined possible. They were finding that the data they wanted to access and use was not fit for their use. There were a range of data quality issues. Further, in some cases, the data was not available due to privacy concerns –whether valid or otherwise– and often, there was no clear path to a policy framework that could be used to ameliorate those concerns.

These issues, especially concerning data quality, are borne out by a recent study that reported that data quality is still a major issue for data scientists. Sixty percent of the data scientists surveyed spend most of their time cleaning and organizing data instead of mining data for patterns (9% of their time) or refining algorithms (4% of their time). "Messy data is by far the more time-consuming aspect of the typical data scientist's work flow. And nearly 60% said they simply spent too much time doing it" [8]. Of even greater concern is KPMG's [19] recent survey reporting that, in the context of the increasing amount of organizational data that in many cases has not been governed by traditional controls, 49% of CEOs surveyed "are concerned about the integrity of the data upon which they base their decisions" (p. 14).

1.4 Data governance and data management for AI

Understanding quality, security, and privacy issues and their implications for the data held in government systems requires deep understanding about that data, its various uses, and the laws that govern the use and exchange of data assets. These types of issues are increasingly being dealt with through enterprise data governance bodies. Such bodies, often under the direction of Chief Data Officers (CDOs), are working to create the kind of policy coherence necessary to ensure that data is both fit for use as well as attending to security and privacy issues. These require executive level attention and authority. The appointment of CDOs in government agencies, however, is a new and understudied phenomenon.

As organizations from across industrial sectors mobilize to exploit AI, they are rediscovering the importance of data management as a core competency [4]. Government agencies may be well advised to similarly develop a set of data management practices as a prerequisite to the establishment of robust AI initiatives. Below we review several critical dimensions of data management [16] and illustrate the ways that they contribute to the creation of AI capabilities. We conclude by drawing attention to five important challenges, somewhat unique to the government context, that must be addressed before governments can feasibly pursue AI for social good. We suggest that individuals occupying the relatively new position of the Chief Data Officer in government are often in the best position to guide their organizations in the development of mature data management and help government ensure the use of data that is fit and appropriate for the purposes to which it is applied.

2 DATA MANAGEMENT PRACTICES IN GOVERNMENT

While new and emerging technologies have changed the context within which data management is performed, the key concepts in data management have not changed radically over the years. One of the most important frameworks for data management practices has been created by the Data Management Association (DAMA) and is expressed in the Data Management Body of Knowledge

(DMBOK) [16]. The DMBOK includes 9 dimensions (see Figure 1): data governance, data architecture & design, database management, data access management, data quality management, master data management, data warehouse & business intelligence management, records management, and metadata management. The framework provides a conceptual foundation to understand data management practices and to organize a knowledge base of current practices and expertise in the use of existing and new data resources by government agencies.

Figure 1: Data Management Framework (DMBOK).

2.1 Data Governance as an Overarching Principle

The structures and practices required for exercising authority and control over the management of data assets are collectively referred to as data governance. Data governance structures, such as formal policy-making bodies, and regularized, systematic practices, enable the organization to make decisions about data and to create policies that stipulate how people and processes are expected to behave in relation to data. The overarching purpose of enterprise data governance is to ensure that data used throughout the organization is managed properly, according to policies and best practices [20] with implications for several important outcomes. For example, establishing data governance enables an organization to create the technology tools and policy to minimize security issues and ensure privacy related to the creation, access, and use of data [4].

Through data governance, the agency creates the overarching policies and structures that enable data users to be confident in the origins and characteristics of their data and trust that it will be appropriate for the uses to which it is applied. To derive the greatest value, current policy analytics tools make use of datasets integrated from dynamic, multiple, and disparate systems, whose provenance may be ambiguous, assuming of course that the existence of such data is even known. Governance tools such as a data catalogue reduce time spent finding data and increases the time devoted to model creation [6]. However, understanding the data's origin, format and lineage is vital for determining the uses to which the data can be applied [5]. Enterprise data governance is an enabler for AI and our ability to realize its potential. Through effective governance, an agency can ensure that a set of subsidiaries but essential data management practices, discussed below, are performed routinely with high standards of effectiveness.

2.2 The Basics of Managing and Integrating Data

Data Architecture & Design involves identifying the data needs of the enterprise, and designing and maintaining the master blueprints to meet those needs. Using master blueprints can guide data integration, control data assets, and align data investments with business strategy. This also includes Data Modeling, defined as the process of discovering, analyzing, and scoping data requirements, and then representing and communicating these data requirements in a precise form called the data model. This process is iterative and may include a conceptual, logical, and physical model. Moreover, data architectures connect the organizational strategy with business processes, databases and information systems [30]. **Enterprise architecture** enables AI and analytics given that it facilitates the connectivity between operational and analytic platforms using consistent naming standards and data definitions across systems within the organization. By connecting data to business processes and systems, data architecture provides the information needed to plan and implement introduction of new analytic platforms and data repositories for AI.

Database Management includes the actions a business take to manipulate and control data to meet necessary conditions throughout the entire data lifecycle. These actions are based on a broad and unified internal perspective about the importance of data assets and what is required to manage and improve them. Database management encompasses, on the one hand, activities associated with the analysis of all new applications to ensure compliance with enterprise data architecture. On the other hand, database management also involves activities related to managing data through their lifecycle [14]. Unfortunately, traditional database management and data lifecycle management practices do not always include data analytics processes within their scope, focusing on the management and archiving of operational data, and leaving out the life of data for the purposes of strategic and policy analysis as well as decision making. In this sense, there is still much to learn about better integrating analytics into data lifecycle management [12].

Master Data Management is related to managing shared data to meet organizational goals, reducing risks associated with data redundancy, ensuring higher quality, and reducing the costs of data integration. Master data defines the essential business entities related to an organization's mission and activities [26]. These essential entities include business partners, clients or employees. In a sense, master data management "enables consistent, shared, and contextual use across systems, of the most accurate, timely and relevant version of truth about business entities" [26]. The lack of appropriate master data management practices increases the likelihood of problems related to data redundancy, duplication and inconsistencies when integrating data for AI and analytics purposes [27]. Master data management has been identified as a key success

factor in the implementation of data repositories with datasets that can be integrated for AI such as big data lakes [27].

Records Management or Content Management encompasses managing unstructured data including planning, implementation, and control activities for its lifecycle. This entails controlling the capture, storage, access, and use of data and information not traditionally stored in the relational databases, focusing on maintaining the integrity of and enabling access to documents and other unstructured or semi-structured information. The value of this unstructured data for AI and analytics has been explored in some policy domains such as health, using medical records to explore alternate policies to improve health and increase efficiencies in the health system [31]. Although the impact of such efforts is still under scrutiny [31], unstructured data has been successfully used in the private sector [25]. Successes in the private sector suggest that unstructured data may be a significant source of value for using AI in other policy and management domains.

Finally, **Data Warehouse & Business Intelligence Management** is the data management dimension that involves planning, implementation, and control processes to provide decision support data and support knowledge for workers engaged in reporting, querying, and analysis. In addition to data warehouses that are commonly used in government organizations for reporting, visualization and statistical analysis of data, there is an emerging concept of a repository to support AI activities, the Big Data Lake. "The Big Data Lake has been conceptualized as a single data repository for an enterprise, typically designed to work in conjunction with Hadoop. Multiple sources such as data from an enterprise resource planning (ERP) system, sensor data from the Internet of Things and data generated from Twitter or other sources would be located in the same repository" [27]. Just like data warehouses, analysts will use data in the lake to find solutions to emerging problems. Data Lakes are not substitutes for Data Warehouses, but constitute a new repository more appropriate for AI-driven analysis. Data lakes hold structured and unstructured data in a "raw" format. Data structure requirements are defined ad-hoc at the moment of the analysis [6].

2.3 Ensuring Quality and Fairness of the AI Process

Data Quality Management focuses on what is done to plan, implement, and control activities that assure that data is fit for analysis and will meet the needs to which it is applied. The data must be appropriately formatted, error free, and organized in ways that are suitable for integration with other datasets [29]. Data quality management consists of choosing quality dimensions that are appropriate for assessing the suitability for use of a dataset based on what is critical for business operation, reporting, or other relevant purpose.

DAMA UK [32] recommends six primary dimensions for assessing quality: accuracy (does the data reflect reality?), uniqueness (is there one authoritative view of the data?), completeness (does the data contain a value for all members in the set?), consistency (can we match records within and across datasets?), timeliness (does the data represent reality at a specified time?) , and validity (does the data conform to the syntax of its definition?). Each dimension is defined through reference to some standard, which is often empirical reality. The dimension is then assessed by determining what to measure. For example, having ascertained how many individuals are empirically members of a group, completeness of data about that group is defined as "The proportion of stored data against the potential of '100% complete'" (p. 8). What is measured might be, for example, the number of blank values for a given data element. For each data dimension one must determine what range or values represent "good" or "bad" quality data, again based on the requirements of the task, e.g., is 95% completeness sufficient for the intended purposes?

Unfortunately, these "primary" qualities cannot be taken for granted, thus data scientists, will need to determine if the data can be trusted; that is if the quality of their data is appropriate for a specific analytic purpose and, if not, what actions can be taken to correct it. Within the context of AI, an equally important issue however is the nature of the data itself from the perspective of its collection and its stewardship. Since that data may be used for such tasks as predictive modeling and case-related decision making, the datasets must be free of bias and stand up to tests for systematic discrimination. For example, tools used in predictive policing have received substantial criticism because of the datasets upon which they are based [11]. Critics charge that such tools are racially biased because they rely on data produced through historical policing practices that have been concentrated in areas that over-represent poor and minority populations producing biased predictions [17].

There is widespread concern that data taken from their original production venues and used elsewhere for modeling and algorithm-driven decision making may produce outcomes that can distort [9] and discriminate [1]. Data scientists must be knowledgeable about the contexts from which datasets have been created or extracted and make judgements about the appropriateness of using such data for their purposes. Since datasets are likely to be integrated across organizational boundaries, government agencies must maintain thorough records that annotate their sources and preserve information about the contexts of origin.

To these ends, **Metadata Management** comprises the planning, implementation, and control activities to enable access to high quality, integrated metadata. Commonly defined as "data about data," metadata includes information about technical and business processes, data rules and constraints, and logical and physical data structures. It describes the data itself (e.g., databases, data elements, data models), the concepts the data represents (e.g., business processes, application systems, software code, technology infrastructure), and the relationships between the data and concepts represented. As Vemuganti [33], puts it "The importance of metadata cannot be overstated. Metadata drives the accuracy of reports, validates data transformations, and ensures accuracy of calculations and enforces consistent definition of business terms across multiple business users" (p. 5-6). This is especially critical for volatile Big Data, changing over time and requiring short-lived but rapid analytical treatment. Industry professionals recommend that metadata be integrated into agency repositories, such as a data catalog equipped with metadata functionality. Metadata management would provide that any metadata existing in, for example, business applications, data warehouses, and data quality systems be linked, enabling "changes to be detected and policies applied immediately, without

manual steps. This ensures reliable data training is fed into the AI model" creating efficiencies [6], reducing erroneous results [33], and minimizing inappropriate uses of data.

Finally, **Data Access Management**, also known as data security management, refers to the principles by which security policies and procedures are defined, planned, developed and executed. These policies and principles are the basis for determining who has access to data, what authentication procedures are required, and how data and information assets are audited; they are foundations for protecting individual privacy. Regardless of advanced data analytics, government agencies are required by statute and regulation to safeguard access to personally identifiable data, although, as the 2014 and 2015 data breaches at the Office of Personnel Management illustrate, the need for data protection is constant, and government does not always measure up to these challenges [22]. However, these issues become increasingly relevant because the greatest value obtained in using AI often comes from integrating volumes of data from multiple sources that generate trails of potentially identifiable data (PII). Reverse engineering strategies can undermine efforts to strip data of PII in social media channels [18]. When joined and analyzed together, integrated datasets provide the potential for, as O'Leary [27] puts it, "...the ability to assemble multiple views of the customer [that] may provide inappropriate insights" (p. 72).

Data breaches that violate privacy are always troubling to those whose personal information is exposed. However, as Mehr [24] has pointed out, the problem is compounded when individuals have not opted in or been given an opportunity to opt out of such initiatives. This contributes to the erosion of trust in government by its citizens. She recommends "treading carefully" (p.11) on issues of privacy and making AI a citizen-centric process in which citizens are educated about AI and invited to "co-create" ethics and privacy rules for the use of their personal data (p. 12).

3 CURRENT CHALLENGES OF DATA MANAGEMENT FOR AI IN THE PUBLIC ENTERPRISE

Government organizations face important challenges associated with data management its practices. Exploitation of data for AI and improved policy and decision making will depend upon crafting successful strategies to overcome such challenges.

Complexity of the Government Enterprise. National, state/provincial and local governments are complex organizations composed by a set of agencies that typically have disparate missions, goals and values. In addition, government policies and management practices need to respond to a set of stakeholders that also represent a complex set of values and interests that are frequently in conflict. Historically, each agency has focused on the development of their own data management practices and information technology capabilities, creating an environment where promoting enterprise-wide data governance is a challenging task. Although recent trends show that governments are working on more centralized structures for managing their information resources, integrating IT and data management across boundaries is very much a work in progress.

Legacy systems. Similar to other larger organizations, governments still rely heavily on legacy systems where implementing current practices of data management is more complicated because of the nature of the databases and the technical tools available to map data and document metadata. Data management of legacy systems depends on experts in the database and the programming language used for its implementation (frequently COBOL). Moreover, there is an additional pressure created by the retirement of these experts, whose departure will mean a loss of knowledge about the meaning, context and nature of the data stored in such legacy systems.

Incentive structures for information management. Also similar to other larger organizations, IT departments in government have incentive systems that push for fast responses in the development of systems and solutions to operational problems. However, there is a lack of incentive to promote the development of governance and data management practices. The lack of proper definitions of the meaning, origin and context of the data creates important voids in the potential application of the data for AI or other analytic tasks. Without the proper incentive system the whole AI enterprise is at risk.

Digitization of unstructured data. As we described in the previous section, AI applied to unstructured data constitutes an important area of opportunity, allowing policy makers to learn from electronic records better ways of improving the quality of life of the citizens, increased efficiencies as well as the impact of policy and government programs. Governments produce a large amount of unstructured data, and digitizing these data represents a challenge by itself, considering that it is not about scanning, but uncovering the meaning and context of the data in the documents –and in many cases– anonymizing the record before analysis to protect the privacy rights of the individuals that own the records.

The continuing challenge of data quality. Traditionally, data quality has been understood as a multidimensional concept involving intrinsic and contextual characteristics of data [34]. Good data management practices have a potentially high impact on the improvement of the intrinsic dimension of data quality, involving issues of accuracy, completeness, and consistency, among others, as previously discussed. However, the contextual quality of the data will continue to be a challenge. That is to say, using data for a purpose different than the purpose for which it was created also poses a challenge for the interpretability, fitness for use or relevancy of the data for the new application [9].

3.1 The CDO as a necessary role

Data management of the type needed to pursue AI initiatives in the public sector is predicated upon taking an enterprise view of government's data resources. This raises the question of what role in government is appropriately positioned to manage this substantial undertaking? Elsewhere we have argued that individuals playing the role of the Chief Data Officer (CDO) are well positioned to exercise leadership in the process of creating data-centric agencies within the public sector [13]. The Chief Data Officer is the individual who should possess a deep understanding of an agency's data and of the agency's business and, based on this, can cultivate and incentivize a culture of data stewardship within the agency.

Governance is an overarching principle of data management which lays out the policies and plans that allocate decision rights and responsibilities to organizational actors involved in managing

data assets. The CDO is the individual who should be charged with leading such governance efforts, forging the collaborations and structures that are required, and ensuring they are sustainable. Finally, if the burgeoning troves of data within the public sector, combined with AI and its analytic tools, are to produce outcomes with public value, the CDO, along with other government leaders, must also exercise creativity required to close the gap between data readiness and data use. What policy options are best explored with AI? What services are best redesigned with AI? What repetitive agency tasks are most effectively automated? Realizing the potential of AI for social good relies on investments in explorations focused on how to use it for social good, but also in the ongoing process of building trust in government data and ensuring it is ready and suitable for use, both in those explorations and beyond.

4 CONCLUSIONS

AI lives by data but, at the same time, AI may also die by data. Data that is problematic introduces a serious set of risks that can be economically devastating for any company and that may erode trust in the legitimacy of government [2]. Thus, the success or failure of AI and the data analytics that government will increasingly pursue rests on the quality of its fuel. These technologies are vitally dependent on access to vast data stores, but equally as important, such data must be cleaned, integrated, of high quality, structured, and secure. In essence, there needs to be a data management process in place to ensure data are properly used. Unfortunately, it is frequently not possible to assume that data held by government meets these qualifications. However, as we have shown, the DMBOK framework may be used to understand data management practices in government, including in the context of initiatives related to AI. Moreover, as we have argued, the CDO may serve a unique and vital role in undertaking data management practices in government, especially given the varied technical, organizational and ethical issues that Big Data and AI bring to the public sector.

REFERENCES

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[2] Muhammad Anshari and Syamimi Ariff Lim. 2017. E-Government with Big Data Enabled through Smartphone for Public Services: Possibilities and Challenges. *International Journal of Public Administration* 40, 13 (2017), 1143–1158.

[3] D Araya. 2019. Artificial Intelligence And The End Of Government. https://www.forbes.com/sites/danielaraya/2019/01/04/artificial-intelligence-and-the-end-of-government/#2eb7cbda719b

[4] R Bean. 2018. How big data and AI are driving business innovation in 2018. https://sloanreview.mit.edu/article/how-big-data-and-ai-are-driving-business-innovation-in-2018/

[5] P. Brunet. 2018. Data lakes: Just a swamp without data governance and catalog. https://www.infoworld.com/article/3290433/data-lakes-just-a-swamp-without-data-governance-and-catalog.html

[6] P. Brunet. 2018. The Real Competitive Advantage of AI Lies In Data Governance. https://aibusiness.com/data-governance-collibra/

[7] Francesco Corea. 2018. AI Knowledge Map: how to classify AI technologies. https://medium.com/@Francesco_AI/ai-knowledge-map-how-to-classify-ai-technologies-6c073b969020

[8] CrowdFlower. 2016. Data Science Report. https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf

[9] Sharon S. Dawes and Natalie Helbig. 2015. The Value and Limits of Government Information Resources for Policy Informatics. In *Governance in the Information Era*, Erik W. Johnston (Ed.). Routledge, New York, 25–44. https://doi.org/10.4324/9781315736211-11

[10] Kevin C. Desouza. 2018. Delivering Artificial Intelligence in Government: Challenges and Opportunities. http://www.businessofgovernment.org/report/delivering-artificial-intelligence-government-challenges-and-opportunities

[11] Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* St. Martin's Press, New York, NY.

[12] Samuel Greengard. 2015. Getting a Handle on Data. , 1 pages. https://libproxy.albany.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=108661020&site=ehost-live

[13] Teresa Harrison, Theresa A Pardo, Mila Gasco, and Donna S Canestraro. 2018. The Salience and Urgency of Enterprise Data Management In the Public Sector. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. University of Hawai'i at Manoa, Big Island, HI, 2246–2255.

[14] Elizabeth Horwitt. 2008. How data lifecycle management saves millions of dollars. *Enterprise Innovation* 4, 5 (Nov. 2008), 38–39. https://libproxy.albany.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=35691836&site=ehost-live

[15] Fernando Iafrate. 2018. *Artificial Intelligence and Big Data: The Birth of a New Intelligence* (1 edition ed.). Wiley-ISTE, Hoboken, NJ.

[16] DAMA International. 2017. *DAMA-DMBOK: Data Management Body of Knowledge* (ediciÃşn: second ed.). Technics Publications, Basking Ridge, New Jersey.

[17] Keith Kirkpatrick. 2017. It's Not the Algorithm, It's the Data. *Commun. ACM* 60, 2 (Jan. 2017), 21–23. https://doi.org/10.1145/3022181

[18] Rob Kitchin. 2016. The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374, 2083 (Dec. 2016), 20160115. https://doi.org/10.1098/rsta.2016.0115

[19] KPMG. 2017. Disrupt and grow: U.S. CEO outlook 2017. https://assets.kpmg/content/dam/kpmg/us/pdf/2017/06/us-ceo-outlook-survey-2017.pdf

[20] John Ladley. 2012. *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program.* Morgan Kaufmann Publishers, Waltham, MA. https://www.amazon.com/Data-Governance-Effective-Kaufmann-Intelligence/dp/0124158293

[21] Karl M. Manheim and Lyric Kaplan. 2018. *Artificial Intelligence: Risks to Privacy and Democracy.* SSRN Scholarly Paper ID 3273016. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=3273016

[22] Joseph Marks. 2018. OPM is Still Far Behind on Data Protection Three Years After Devastating Breach. https://www.nextgov.com/cybersecurity/2018/11/opm-still-far-behind-data-protection-three-years-after-devastating-breach/152804/

[23] Bernard Marr. 2017. Why AI Would Be Nothing Without Big Data. https://www.forbes.com/sites/bernardmarr/2017/06/09/why-ai-would-be-nothing-without-big-data/

[24] Hila Mehr. 2017. *Artificial Intelligence for Citizen Services and Government.* Technical Report. Ash Center for Democratic Government and Innovation: Harvard, Cambridge MA. https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

[25] Oliver MÃijller, Stefan Debortoli, Iris Junglas, and Jan vom Brocke. 2016. Using Text Analytics to Derive Customer Service Management Benefits from Unstructured Data. *MIS Quarterly Executive* 15, 4 (Dec. 2016), 243–258. https://libproxy.albany.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=120565630&site=ehost-live

[26] Martin Hubert Ofner, Kevin Straub, Boris Otto, and Hubert Oesterle. 2013. Management of the master data lifecycle: a framework for analysis. *Journal of Enterprise Information Management* 26, 4 (Aug. 2013), 472–491. https://doi.org/10.1108/JEIM-05-2013-0026

[27] Daniel E. O'Leary. 2014. Embedding AI and Crowdsourcing in the Big Data Lake. *IEEE Intelligent Systems, Intelligent Systems, IEEE, IEEE Intell. Syst.* 29, 5 (2014), 70. https://doi.org/10.1109/MIS.2014.82

[28] New Vantage Partners. 2018. Data and innovation: How Big Data and AI are driving business innovation. http://newvantage.com/wp-content/uploads/2018/01/Big-Data-Executive-Survey-2018-Findings-1.pdf

[29] Philip Piletic. 2018. Why Artificial Intelligence Cannot Survive Without Big Data. https://www.smartdatacollective.com/why-ai-cant-survive-big-data/

[30] Neil Ross and Daniel Petley. 2006. Enterprise Architecture: The Value Proposition. *DM Review* 16, 1 (2006), 56–57. http://web.a.ebscohost.com/ehost/detail/detail?vid=9&sid=ec09a9c7-963d-4b22-bfc0-08a922a16a76%40sdc-v-sessmgr02&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=20292617&db=bth

[31] Kyan Safavi, Simon C. Mathews, David W. Bates, E. Ray Dorsey, and Adam B. Cohen. 2019. Top-Funded Digital Health Companies And Their Impact On High-Burden, High-Cost Conditions. *Health Affairs* 38, 1 (Jan. 2019), 115–123.

[32] DAMA UK. 2013. *The six Primary dimensions for Data Quality asseSsment.* Technical Report. DAMA UK. 17 pages. https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf

[33] G. Vemuganti. 2013. *Metadata Management in Big Data.* Technical Report. Infosys Labs Briefings.

[34] Richard Y Wang and Diane M Strong. 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems* 12, 4 (1996), 5–33.