The Salience and Urgency of Enterprise Data Management In the Public Sector

Teresa Harrison, Ph.D
Center for Technology in Government
University at Albany/SUNY
tharrison@albany.edu

Theresa A. Pardo, Ph.D
Center for Technology in Government
University at Albany/SUNY
tpardo@ctg.albany.edu

Mila Gasco, Ph.D
Center for Technology in Government
University at Albany/SUNY
mgasco@ctg.albany.edu

Donna S. Canestraro, PMP
Center for Technology in Government
University at Albany/SUNY
dcanestraro@ctg.albany.edu

Abstract

In this emerging topics paper, we argue that enterprise data management is a key enabler for new and innovative uses of data. Given widespread recognition of the public value potential of these new uses of data, enterprise data management capability is increasingly salient and recognized as urgent. We further argue that creating capability for enterprise data management is poorly understood. However, since enterprise data management is a future practitioner imperative, new research from the digital government community addressing the challenges to creating such capability is required. We illustrate the salience and urgency of enterprise data management through three vignettes that highlight the potential of such efforts to reorganize the public sector along new data oriented lines. A focus on the role of governance and the chief data officer as key enablers to creating public value from data highlight the need for research in these areas.

1.0. Introduction

The siren calls of big data, open data, and integrated data, together with data analytic tools such as data mining and machine learning, beckon to today's organizational leaders with the promise of powerful strategic information resources that can be used to generate increasingly sophisticated analyses of organizational operations. Corporations have been quick to recognize the value that big data and data science bring to decision making about all aspects of management. Having harnessed the business advantages of integrating data within enterprise system infrastructure at the turn of the millennium, large corporations now seek to exploit the digitized information generated through clickstreams; social media; cookies, sensors, and meters; geo-locations; and biometrics, to name a few new types of data whose value is being explored by data scientists using predictive analytics in data-driven decision making for business advantage. As Davenport, Barth, and Bean [6] point out, big data differs from traditional data in that its advantages are achieved largely through analyzing the continuous streams of relevant data and their processes in real time (as opposed to static data that is stored), and incorporating the insights of data scientists engaged in on-demand decision analytics into business and operational functions at the core of the enterprise. Further, as Joseph and Johnson [12] suggest, bringing big data to bear on policy problems can expose new information patterns or unsuspected correlations that might point to better understandings or new interpretations and new ways to construct systems of work that lead to "transformational" or t-government, the next stage in the evolution of e-governance.

Focusing on data rather than technology, however, is a relatively new perspective within the broad scope of information technology management in government organizations. This perspective foregrounds data, both its quantity and its quality as a driver of organizational processes, as well as all related operations including data stewardship, data governance, data standards, data quality management, data architecture, and security [26]. Following the example of business, the federal government, as well as certain state governments have appointed leaders in data management to the role of "chief data officer" (CDO) to help guide this complex multi-dimensional undertaking. This new position differs from, but does not replace, traditional data managers; instead, the CDO is given the responsibility to lead *organization-wide* data strategies and "….put data on the organization's business agenda and in the minds of other executives and officers. Under the leadership of a CDO, business strategies reflect and exploit data, particularly big data, instead of treating data merely as a by-product of running the business [16]. The public sector faces considerable challenges to achieving enterprise-level benefits from these new sources of data. We know, for example, that although much government data is large if not big and increasingly open, this data is often housed in agency-controlled data siloes, with infrastructures served by outdated hardware and legacy software. Many government units have little or no tradition of cross-agency collaboration and their employees lack training related to data stewardship. Perhaps of most concern is the absence of overarching state-wide strategies for

URI: http://hdl.handle.net/10125/50170
ISBN: 978-0-9981331-1-9

data sharing and integration that might be used to galvanize efforts to address these daunting issues.

Beyond these impediments, it is also worth remembering that the goals of businesses and governments are quite different, since businesses are focused on earning profits by developing and maintaining a competitive edge; government, on the other hand, is focused on promoting economic growth, maintaining the peace and providing for security, creating the foundation for sustainable development, and respecting ethical and legal relationships between the state and citizen [13] [30]. Thus, the solutions and strategies that work for business may not always be directly applicable to the public sector; government decision makers will always need to be sensitive to the economic, legal, and social constraints on enterprise-level innovation.

However, it is also clear that an enterprise level data management approach that enables integrating data from multiple government data sources, open data from any source, and relevant big data generated through digitized transactions has much to offer to the public sector. These benefits begin with increasing efficiency and effectiveness, thus saving time and money for taxpayers and making it possible to improve services to them through personalization or tailoring [30]. Enterprise-level data management also bears the promise of improving the ability to engage in data-driven decision-making for policy issues.

In this emerging topics paper, we argue that enterprise data management is a key enabler for new and innovative uses of data and due to the recognition of the public value potential of these new uses of data enterprise data management capability is increasingly salient and recognized as urgent. We further argue in this conceptual paper that creating capability for enterprise data management is poorly understood. Enterprise data management is a future practitioner imperative and new research from the digital government community around the challenges to creating such capability is required. By "enterprise data management" we refer to a model of information system management in organizations that focuses first and primarily on data, seeking to "manage heterogeneous data sources, validate the quality of data, devise a common data model by integrating information, build analytical and presentation layers, and manage end-to-end metadata in the analytical and presentation layers" [27]. The objective of such management systems is to create an "integrated enterprise-wide data environment" that "ensures consistency of information with a 'single version of truth' "and thus encompasses data stewardship, data governance, data standards, data quality management, data architecture, and security [27].

We begin by reconsidering the term "enterprise" as it has been traditionally used in referring to the development of enterprise architectures and infrastructures, reminding ourselves of the organizational and business problems that seem inevitably implicated when carrying out enterprise level action. We then consider what "enterprise" might mean in the context of data management in the public sector, where governments may derive benefits from sharing their data across both agency boundaries and administrative levels. In the context of government data sharing, enterprises are nested within each other, thus further complicating already complex data environments. We present three illustrations of initiatives that either require enterprise data management or are uniquely focused on the creation of such capability, to illustrate some of these challenges involved in making optimal use of data, whether it is administrative, statistical, open, closed, integrated, or big data. Finally, we consider the relatively new role of the Chief Data Officer as the individual chiefly responsible for forging productive collaborations among governmental units, cultivating a culture of data stewardship, and creating data governance models. We raise a number of questions about the future of this position in light of its myriad challenges as well as from the perspective of e-government practice and research.

2. "Enterprise" Level Structures and their Challenges

Organizations have long recognized the crippling effects of creating and maintaining disparate information systems within their boundaries, but it was not until the 80's and 90's that large corporations began the move en masse to "enterprise" systems as a way of bridging gaps between data systems. The enterprise system was viewed as a way of replacing previously disparate and often idiosyncratic information systems in an effort to share information across multitudes of business functions, such as sales, manufacturing, suppliers, and accounting and across hierarchical levels [19].

2.1. Enterprise Structures for Business Organizations

Corporate experience with enterprise systems led to the recognition that adopting a new information system invited changes to organization processes and functions. This is because of the inevitable need to

coordinate decisions about information with the broadest possible view of what business does and how it achieves competitive advantage. As Davenport [5] suggested, an enterprise system can be "viewed as an opportunity to take a fresh look at the company's strategy and organization." In making decisions about type, definition, form, structure, quantity, quality, and flow of information as well as where and how to standardize, changes inevitably need to be made in how organizational units conduct work, from job design, work sequencing, and training to the possibility of organizational restructuring [17]. Given such major effects, information technologists and business managers at the highest level were advised to work hand-in-hand in order to mediate between "the imperatives the technology and the imperatives of the business. If the development of an enterprise system is not carefully controlled by management, management may soon find itself under the control of the system" [5].

Recent moves by businesses to invest in enterprise data management strategies, stimulated largely by hopes that external unstructured data might be integrated with more finely granular data within the organization to provide business value, has pushed questions related to data to the forefront of organizational decision making. But while organizations are focused increasingly on data management, they have not been necessarily successful in realizing the value of this new focus (see e.g. [2]). Indeed, questions remain about how the growing amount of data can be used profitably, a focus of great relevance as well to organizations in the public sector. In suggesting that alignment between data and organizational goals is key to the ability to derive value from big data, Bean [1] advises organizations "to take a step back and think about their key business drivers: What are they? Which ones would benefit from more sophisticated data-driven decision-making, in particular, the ability to iterate through data more rapidly and integrate new sources of data? And then think about how to put in place the processes and make sure they have the organizational alignment and skills to make that happen. Organizations that have a very clear view of what they're trying to achieve and how they're going to achieve it are going to have the greatest probabilities of success [1].

2.2. Enterprise Planning for Government Organizations

Government agencies in the 21st century can hardly be blamed for succumbing to the hopes and aspirations that have pushed business organizations to move to enterprise level information infrastructures. Indeed, in 2012 under the Obama Administration, the federal government launched an effort to better understand, manage, and act upon the enormous data resources generated through research and government activity, in part to use data analysis to advance scientific knowledge, but also to make progress on achieving national goals in health, energy, defense, and education [18]. But while the data deluge we are experiencing is relatively new, pleas for data sharing and collaboration among government agencies, for interoperability, and for shared data governance have been, for some time, well represented in the literature of e-government. For example, Pardo, et al., [2006] observed that government managers need information external to their particular organization for problem solving and attempt to integrate data imported across agency boundaries. Their case study suggested that the success of such initiatives depends on data sharing, interoperability, a change in agency culture, and the support of legislators and policy makers to remove legal constraints on such action and provide a foundation for enabling adaptable governance structures. Other research has explored the role of trust as a factor that, beyond brute technological capabilities, is needed to cultivate information sharing by clarifying roles and responsibilities, developing collaborative decision making, and communicating respect for the autonomy of participating organizations. And Dawes [7] has argued that a future "infrastructure for digital government requires an extended view of enterprise that goes far beyond a single organization to encompass all the parts of a government as an interconnected whole operating in a complex social and economic environment" (p. 258). The vision of the future generated by European Commission participants in her study is one that includes the theme of public-private-civic sector relationships that are focused on "sharing responsibilities and exchanging information among networks of diverse organizations in ways that generate public value and satisfy public requirements for fairness, accountability, and competence" (Dawes, [7]). Recent work by Susha et al., [29] looks closely at the types of networks identified by Dawes, calling them "data collaboratives". Data collaboratives, as defined by Susha et al., [28] are "cross-sector (and public-private) collaboration initiatives aimed at data collection, sharing, or processing for the purpose of addressing a societal challenge" [28]. Their work focuses on the coordination problems and the creation of mechanisms to match demand for data with supply, finding that data collaboratives exhibit a "bazaar" form of coordination; "In data collaboratives, the matching

is often defined by what kind of data is on offer and the incentives and control are low" [29].

2.3 Clockwork Government vs. Information Ecosystems

Thus, recognition of both the need for enterprise data management and some of the factors inhibiting its realization are hardly new; but it is clear that creating an enterprise data management strategy that encompasses both internal and external partners is considerably more complicated than might have been supposed. Our failure to recognize this may due to a simplistic conceptualization of the way government works. Eggers and O'Leary [9] observe that "Some systems operate like clockwork. Government does not. Government is closer to a mud-wrestling match than a precision time-keeping mechanism". The assumption of "clockwork government" gives rise to the expectation that government employees carry out their duties systematically, predictably, and in a way that meshes perfectly with the mission of the agency and with the interests of other governments units with which there might be actual or potential collaboration.

The assumption of "clockwork government" may fail in all government pursuits, but failure is particularly impactful when it comes to base data, the building block of all data-driven operations. The "data life cycle" consists of tasks related to the planning, collection, creation, and curation of data that, for better or worse, will determine the quality of the data. Unfortunately, most government agencies are not oriented or sensitized to such tasks. The people who execute these tasks do not have jobs as "data stewards." They are most likely doing something else that happens to require data collection and handling but that is not the principal responsibility of the position.

The "clockwork government" metaphor suggests that organizational members make their decisions to optimize utility or advantage. However, organizational theorists have known for decades that the rationality of organizational members is bounded by the constraints of the situation and employees' cognitive capacities, and the time available for decision making [26]. Instead, decision makers are "satisficers" tending to seek satisfactory, rather than optimal, solutions in light of current exigencies, rather than looking beyond to other interdependencies. Instead of the mechanistic overtones of "clockwork government" a more apt metaphor might be that of the information ecosystem, which focuses on the inter-relationships between organizations who might share information. An ecosystem is open, by definition, simply because it consists of input and output relationships. Open government ecosystems have been conceptualized as "departments, bureaus, and offices interacting in multiple ways with each other; some of these offices are interacting with counterparts on state and local levels. These layers comprise multiple contexts with quasi-independent decision makers, customized technologies, legacy systems, strained budgets, amounting to complexity at every level. All this may work fairly well within given organizational units, but the more organizational units that are interconnected, the harder it becomes to predict and manage as issues related to coordinating technological and organizational infrastructure are presented. The complexities and limitations of this practice context make data sharing and the process of enabling data access difficult [11].

The ecosystem metaphor conveys the complexity and interdependencies that enterprise data management inevitably needs to address. In so doing, it calls attention to issues that lie latent within the problem of enterprise-level action. First, within the context of public sector organizations, it is not quite clear where the "enterprise" begins and ends. Given multiple layers of government, and their legal, political, policy, and data dependencies up and down the hierarchical organization of layers, it becomes apparent that defining the "enterprise" may be one of the most important tasks facing the designers of enterprise data management systems for public organizations, especially those that cross local, state, and federal levels. Indeed, while diverse information systems may reside in bureaus within a state agency, the "enterprise" at issue may encompass the offices of one agency together with those of other agencies in state government, and these, of course, may all be interdependent with agency offices in the federal government and/or an international organization, such as the Worth Health Organization. In other words, government "enterprise" data management systems may well consist of multiple networked organizational enterprises.

Second, we have argued that government ecosystems occur and evolve naturally, but they might also be strategically managed for the purpose of achieving some value or policy vision [11]. We have further advocated the use of "strategic ecosystem thinking" that focuses on identifying the organizational components of the ecosystem, understanding how transactions take place between components, specifying the conditions and resources needed so that these transactions create value, and defining metrics that make it possible to assess the health of ecosystem operation. These suggestions may provide some helpful advice to guide the initial development of

enterprise data management capability. What is not clear is which individuals, in which organizations, occupying which roles, are those who are or should be charged with organizing, resourcing and leading the necessary strategic ecosystem thinking toward the development and execution of an enterprise data strategy.

3. Enterprise Data Management Projects in the Public Sector: Three Vignettes

We illustrate the salience and urgency of enterprise data management through three vignettes that highlight the potential of such efforts to reorganize the public sector along new data oriented lines. The vignettes are provided to illustrate the saliency and urgency of creating enterprise data management capability as an enabler to solve critical public problems and some of the very real challenges of designing and executing enterprise level data management in the public sector.

3.1. Local Enterprise Data Management: Smart Cities and Urban Blight

Urban blight, the deterioration of living environments of cities with abandoned and ruined structures [25], is a persistent and costly problem for cities and local governments across the United States. On average, a single property slated for demolition can incur expenses for a local government in excess of $65,000 [4]. With over 18 million such properties in various conditions of distress across the U.S., the overall potential negative economic impact of this inventory could, over the next five to seven years, reach $1.17 trillion. The yearly costs of a blighted property include direct costs from government including code enforcement, police and fire dispatches, legal, administrative, engineering and property maintenance. In addition, there are indirect costs including uncollected taxes and utilities, decreased tax revenue due to devalued adjacent properties, and the lost opportunity costs of reduced economic development and investment interest activities [25].

Addressing urban blight relies on the ability of governments to think and work regionally due to the behaviors of some property owners, in particular corporate owners, who might own multiple properties in contiguous cities and move from one city to another as violations are identified and fines and other legal actions are undertaken by a single city. These owners, known as “bad actors” make a regional approach to urban blight necessary. Reversing the cycle toward urban blight relies on the ability of city governments to work together as an enterprise to create enterprise data management capability including governance and technical data sharing capability. The few studies that have addressed data driven decision making in the context of urban blight have generally concluded that integrating and sharing information and resources across city boundaries can give governments critical leverage to be more proactive [5]. However, how data is integrated and shared across boundaries and what the role of different stakeholders is in the process of capturing, managing, using and sharing code enforcement information is an underexplored topic.

Due to the lack of relevant knowledge, the cost of urban blight to their communities and the increasing recognition of the critical role of regional coordination around property and code data in addressing urban blight, four New York State Cities came together with funding from the New York State Department of State’s Local Government Efficiency (LGe) Program and in partnership with [to be identified], to develop a shared regional information resource using code enforcement and property information to help inform the programs and policies aimed at combatting urban blight. The cities have spent several years developing a workable collaboration based on shared interest, trust and a clear set of roles and responsibilities that is now enabling the design and development of the necessary enterprise level resources including data governance. The cities regularly refine their understanding of the data problem and the changing context of use to ensure that they have the capability necessary to successfully use and reuse data gathered from many sources, in particular within the context of code enforcement. Code enforcement is the critical local government business process focused on identifying early warning signs and preventing properties from entering the cycle of decline from neglected to vacant to blighted.

The cities have new capability to address urban blight by collaboratively developing the rules they will all live by in terms of collecting, managing and using data, including sharing; they work together to identify and share information about the programs of work they carry out to address urban blight, determining the data most critical to share across governments and then leveraging existing and new technologies to create the most appropriate resource to share information. They are using the grant to support the costly activities of designing a governance structure, identifying core data elements, data standards, and setting expectations regarding process and data quality. Setting forth a foundation to collect, manage and use data and establishing a network to share current practices will allow all cities to provide services and promote health,

safety, and economic development throughout the capital region.

3.2. State Agency Enterprise Data Management: Data and Child Welfare Services

In contrast to the collaborative effort initiated by these cities, the creation of certain data infrastructures in New York state government has taken place under the aegis of the federal government, who collects information from states related to legislative mandates, as illustrated by the following example. In 1993, the federal government passed a law that encouraged states to create a Statewide Automated Child Welfare Information System (SACWIS) that was intended to track children receiving various child welfare services within each individual state. SACWIS offered financial rewards and operational support for the creation of an enterprise-level centrally located system used to manage, track, and report on children in the state's child welfare system. The concept of SACWIS in New York State (NYS) was to build a single data repository to connect the agency that provided child welfare administrative oversight with offices in the 57 counties and New York City as well as not-for-profit agencies that actually provided these services to clients. While states had some latitude in system design, the federal government stipulated 90 system requirements and required the development of a single comprehensive and centralized statewide information system that would function as the location of the "state case record."

As it turned out, New York elected to borrow and build on a SACWIS originally created for Texas. By 2001, six years and $216 million later, only three of the anticipated five system modules were operational, with numerous complaints from users that the modules were neither functional nor reliable. Caseworkers, who had not been invited to participate in the design or testing, reported that the new system, dubbed "Connections," actually required more paperwork to do their jobs, resulting in less time spent with the children they served (cite Too little Too late). As a result, many not-for-profits created local systems to track information about clients. Ultimately, only 36 states (and no tribal organizations) elected to create a SACWIS.

In May 26, 2016, the Department of Health and Human Services (HHS) called on states to create the next generation of child welfare information systems called the Comprehensive Child Welfare Information System (CCWIS), with commensurate financial incentives. The Final Rule was published in the Federal Register on June 2, 2016 [10] and went into effect on August 1, 2016 replacing the existing regulations governing SACWIS [14]. States are currently invited develop a CCWIS, which now stipulates only 14 system requirements and is conceived to run on distributed, reusable, interoperable technologies. CCWIS allows state agencies greater flexibility to develop systems consistent with their practices and business needs by moving away from "data capture" to a "data maintenance" philosophy, as long as a copy of the data is stored and maintained in CCWIS, which must be the source of all data and reports required by federal law. The new system places greater emphasis on multiple and bi-directional data exchanges between child welfare programs in the state departments of health, justice, labor, education, and local agencies as well as other organizations who provide services to children and families.

The federal government is of course interested in streamlining the production, integration, and dissemination of data needed to assess and evaluate their child welfare programs. They have asked states to file their intent to build and deploy a CCWIS by May 2018. Today NYS's child welfare administrative agency is at a decision crossroads: the state can either continue using the existing SACWIS or convert to CCWIS. Continuing with SACWIS bears the risk of financial penalties from the federal government, the loss of federal financial incentives for converting to CCWIS, and perpetuation of the current system's inadequacies. On the other hand, developing a CCWIS requires addressing myriad redundant data and information systems within the current information environment, fixing acknowledged problems with data gaps and data quality, and confronting the uncertainties of designing an entirely new data management system in the absence of models or best practices. The more simple determination of whether or not a CCWIS is feasible will require identifying relevant stakeholders in multiple state agencies, county welfare departments, and local not-for-profit agencies and initiating conversations with them about the prospects for launching enterprise data management.

3.3. State: New York State Data Strategy

State agencies using federal funds must find ways to engage in effective reporting practices. Beyond fulfilling such requirements, New York State government has long been a user of data to support policy and program level decision making. Unfortunately, the ability to take full advantage of the opportunities that new and emerging technologies and analytics tool and techniques provide is inconsistent across the range of NYS government agencies. This inconsistency exists primarily due to capability gaps.

At the enterprise level the state lacks the full set of enterprise data policies, strategies, standards and practices that are recognized as necessary for multi-entity enterprises to fully realize the value of data. For NYS to use data to create new public value, according to the newly hired Chief Data Officer, "there is a need for systematic and standard approaches to capturing, preparing, and managing that data in ways that ensure that data is fit and ready for use."

The CDO is charged with identifying and addressing capability gaps and building the necessary data management infrastructure to ensure that the state has what it needs going forward to realize the public value potential of data and new and emerging technologies including data analytics tools and techniques. In this effort he has identified a set of specific capability gaps as starting points for creating enterprise data management. NYS lacks capability in a number of important areas relevant to enterprise data management. One capability gap is related to the sharing of data across multiple agencies; such sharing is necessary to ensure that data is available to solve today's complex problems. Such sharing is also problematic with many legal and technical hurdles to clear and few prototype successes to point to. A second reason is that the pressure to meet existing program needs make it difficult for agencies to try something new and create pipelines of new knowledge products. A third reason is that government salary structures make it difficult to hire and retain enough in-house data analysts and data stewards, so agencies don't have the capacity to either prepare data for use or work with new linked data. These combined challenges have led to the current situation in NYS; program, administrative and control agencies cannot get the resources necessary to systematically collect, manage and make use of data. Further, to compound the problem, because they don't use data, they don't get new resources. If NYS government is to leverage its data holdings to generate public value, new capability must be created.

The new CDO is seeking to address this capability gap and reverse the negative spiral by developing a NYS Data Strategy including an enterprise data management infrastructure that will create incentives to provide data, resources to prepare data for use in analytics focused on high priority problems and demonstrate criticality of creating the necessary policy and management infrastructures to ensure and facilitate sharing of data across agency lines. Creating such enterprise capability requires full investment in data stewardship, including governance, and analytics capacity for the NYS workforce. To address this gap the CDO is launching a state-wide strategy development effort to produce a NYS Data Strategy. A multi-stakeholder engagement process will be used to define a draft strategy which will then be presented and refined during a two-day public summit. The multi-stakeholder engagement process will bring together a wide range of actors who can help inform the process with respect to the contextual differences found across the NYS government with respect to the wide variety of programs and policies about which the state collects, manages and uses data. These widely varying contexts represent different practices and standards as well as different requirements for how closely and carefully data is managed. In some cases, the quality of individual records is critical to a business process or decision; in other cases the overall quality of a data set is what matters. Understanding of the full range of conditions within which data is considered must inform the data strategy development process.

4. Challenges

The narratives presented above provide concrete illustrations of the challenges governments at all levels are facing as they consider and reconsider the need for enterprise level data management and explore new ways to design and execute accordingly. They reinforce the fact that beyond the technology used to collect, store and crunch the data, the data itself is a newly vital asset that requires new and innovative approach to building the necessary enterprise data management capability. Although we can be sure that the journey to enterprise data management will uncover much that we do not currently know, the stories of governments who are newly embarking upon this journey call our attention, even at this stage, foreshadow the issues that will require deeper inquiry.

4.1. Defining the Enterprise

The idea of an ecosystem is useful for enterprise level action since it focuses, first, on identifying the relevant components and, second, on the transactions among the components, in this case viewed as data exchanges. In each case presented, the enterprise, encompassing component organizations and data exchanges, must be defined. In the case of urban blight, the enterprise has been defined a priori through the decisions of these four cities to collaborate and their enumeration of the relevant city agencies. In the case of child welfare, and for NYS, more generally, the challenge will be to identify relevant stakeholders and forge agreements among them to collaborate in these novel ways. In this case, the enterprise may be more likely to emerge on the basis of the collaborative

agreements to which organizational leaders are willing sign on. This undertaking is likely to require leadership in the form of technical capabilities and expertise as well as leadership authority and legitimacy conferred by elected officials and legislators. Enterprise level action will never be initiated from the bottom of an organization.

4.2. It's all about data, context, and standards

The narratives also make clear that data is a top down and bottom-up endeavor. Data cannot be taken for granted and, indeed, enterprise data management must be built on the basis of data whose provenance and meaning is understood and whose quality can be assessed by its generators, stewards, and consumers. The decisions eventually made will only be as good as the data upon which they are based. Cai and Zhu [3] review the research on data quality emphasizing the criteria that can be used evaluate datasets; but ultimately whether or not data is fit for use depends on what the data consists of and what it is to be used for. This is not just an issue for the relatively mundane and limited process of collecting high quality property data, as in the example of cities addressing urban blight. It is a central issue for a CCWIS, since it is entirely possible that agencies will have their own context-specific definitions and ways of measuring concepts that they share. It is instructive to note, as Dawes and Helbig [8] point out, that the definition of a "family" varies from agency to agency involved in the provision of child care services. This is not just a matter of different norms; there are legal requirements that govern how vocabulary is used within an agency. It is, finally, also a problem for the analysis of sophisticated big data, should that be attempted by New York State or other government units, produced through digital transactions. That data may include online transactions, emails, videos, audios, images, click streams, logs, posts, search queries, health records, social networking interactions, science data, sensors and mobile phones and their applications" [23]. However, such data is highly dependent on the algorithms used by the companies that generate or collect it such as Google, Twitter, and Facebook and do not necessarily produce valid and reliable data for science, or perhaps for policy decisions [15]. The need to know and understand the value of a particular dataset used for a particular purpose is a requirement across context and across dataset.

Beyond this, diverse datasets must be integrated in shared platforms in order to be used for multiple purposes, requiring the recognition that data will be used for multiple and cross-agency purposes. Yiu [30] suggests that individual data owners must be incentivized to collect and manage data "across the board, and not just for the indicators that they (or their managers) deem important." Further, contending with legacy systems from different agencies presents the additional challenge of creating formats and standardized solutions for extracting useful information from datasets and for analyzing it [13]. Security in all of this activity is of the utmost importance. Unfortunately, as Kim, Trimi, & Chung [13] note, today's big data technologies, including Casandra and Hadoop, do not necessarily possess sufficient security tools.

4.3. Governance and the Chief Data Officer

The narratives described earlier cry out, in their individual ways, for the exercise of new forms of leadership within public sector organizations. Data management is not simply a technological issue. We have already noted that the support of elected officials and legislators for enterprise level data management projects is an absolute requirement. However, once the will of the relevant government units is established, these projects need to be led and managed on a daily basis in ways that are still very much in the process of emerging.

Leadership of such projects necessitates structures for data governance, which encompasses agreements about the decision rights and responsibilities of organizational actors involved in managing data assets [20]; the creation of an entity to provide data governance is central task for any enterprise data project. Some recommend the creation of a governance council that can adjudicate among the interests of different stakeholders and ultimately make binding decisions. Such a council might be composed of both data owners, who are accountable for the correctness and consistency of data, and data stewards, who create rules for handling data and are thus accountable for data management [20]. The potential for increasing demand to know how such governance structures work and what factors affect their success is an area of research for e-government scholars that are likely to become critical.

But leadership for such projects must also be situated in an individual who can guide the entire effort, pointing participants in desirable directions, setting goals for the enterprise project as a whole, and supporting the values that should guide decision making. Corporations engaged in enterprise data management have increasingly placed individuals in the role of Chief Data Officer (or similar positions) for such leadership [1]. As illustrated in the narrative

about NYS, the ability to forge agreements between state agencies, accustomed to autonomous action, may be possible only by an individual empowered with authority and legitimacy by the government's chief executive. Interestingly, President Obama created and staffed the position of Chief Data Scientist late in the second administration, although it is currently empty. New York State, as well as New York City, Philadelphia, Boston, Baltimore, Los Angeles, and Chicago, have appointed individuals to the position of Chief Data Officer. The duties of some of these positions have more to do with managing open data portals than with enterprise management. In New York, however, the occupant is responsible for leading an effort to define a state-wide data strategy.

What further actions or duties might lie within the province of a Chief Data Officer? The first duty might be to cultivate growing capabilities for government employees to become better data stewards. Is everyone now a data steward? What needs to be done to teach employees to be more sensitive to the role of data? What skills do employees need to learn; what training must they receive? Is there an ethics of data management?

A second duty might be to lead the efforts to create and maintain data governance mechanisms, such as data quality and governance councils, and take responsibility for their effectiveness [16]. As a leader for these efforts, the Chief Data Officer rises above data managers and stewards as the individual who is best positioned to advocate for enterprise data management to the highest level program and policy leaders within various collaborating agencies.

Finally, a third duty, but by no means the least important, might be to take actions that increasingly close the gap between data readiness and data use. This would require the imagination to conceive of what questions of interest might be relevant to policy makers and leaders within their government context and take the actions needed to make available and ready the relevant data for analysis. For example, former Chief Data Scientist DJ Patil focused, among other projects, on acquiring the data needed to advance President Obama's signature Precision Medicine Initiative to build the largest and richest database of genetic information, partnering at the time with the National Institutes of Health, the office of National Coordinator for Health IT, the National Science Foundation, and the Department of Energy. This kind of responsibility directly addresses the issue of what kinds of issues are most worth addressing through the creation of enterprise data management systems.

It is clear that there is much to be conceptualized in the evolution of this type of position. Bean [1] suggests that the rise of the chief data officer may reduce the need for chief information officers, while others may wonder if the position has arisen due to inadequacies in our understanding of what chief information officers should be doing. Lee, et al. [16] modeled the various perspectives that a chief data officer might take in enacting this role. However, there are important research questions to ask about those who fill this role and about the effects of their activities on the success of enterprise data management systems. Interestingly, the title of Chief Data Officer has yet to appear in the academic literature of e-government; we suspect that this is a situation that is destined to change.

5. Conclusion

The issues of shared data, interoperability, data quality, data standards, and governance are well established topics within the literature of e-government. But our contemporary recognition of the power of data in its linked, open, large, and big forms and the ability to use this data to generate undiscovered insights through analysis has made conversations about data management salient and urgent. Time is of the essence as the potential for data to create public value is real. The emergence (or reemergence) of the phenomenon of enterprise data management of the types we have described here and, no doubt, in many other forms and contexts in the future, suggest an understanding of the need to data seriously, an appreciation of the urgency to do so, and the courage to innovate in ways that have the potential to stimulate substantial changes in the processes, practices, structures, and strategic management of public sector organizations.

Earlier experiences of corporations with enterprise system development suggest that we may expect to see confusion, missteps, and perhaps even failure as public sector employees struggle to discover how to put together the pieces of a puzzle, when the pieces are not quite apparent. But success leads to the possibility of government organizations that are structured, at least in part, around data and data-driven decision making. It is not clear that this will be a completely positive outcome, but it could certainly lead to rejuvenated and innovative government practice. Here too there is much opportunity for e-government scholars to become involved in both research and practice that is fruitful and mutually informative. While many of the relevant issues and concepts are well known to us, we need to apply and test them in contemporary projects, generating perhaps novel understandings of what they

mean and how they are relevant to new projects. We also need to be flexible in our applications of these concepts to the new contexts in which they are deployed. And above all, we need to be sensitive to the appearance of new practices, new models, and innovative solutions as public sector employees experiment with the complexity and the promise of enterprise data system environments of the future.

6. References

[1] Bean, R. (2013). Organizational alignment is key to big data success. Sloan Management Review.

[2] Beath, C., Becerra-Fernandez, I., Ross, J., & Short, J. (2012). "Finding value in the information explosion." *MIT Sloan Management Review*, *53*(4), 18.

[3] Cai, L. & Zhu, Y., (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal. 14, p.2.

[4] Cook, M. & Sutherland, M. (2015). A shared information resource to fight urban blight: challenges and opportunities in NYS. Proceedings of the 16th Annual International Conference on Digital Government Research. ACM Digital Library.

[5] Davenport, T. H. (1998). Putting the enterprise in the enterprise system. Harvard Business Review.

[6] Davenport, T. H., Barth, P., & Bean, R. (2012). How big data is different. *MIT Sloan Management Review*, *54*(1), 43.

[7] Dawes, S. S. (2009). Governance in the digital age: A research and action framework for an uncertain future. *Government Information Quarterly*, *26*(2), 257-264.

[8] Dawes, S S., and Helbig. N.,"Information strategies for open government: Challenges and prospects for deriving public value from government transparency." *International Conference on Electronic Government*. Springer Berlin Heidelberg, 2010.

[9] Eggers, W. D. , & O'Leary, J. (1995). Revolution at the roots: Making our government smaller, better, and closer to home. New York: The Free Press.

[10] Department of Health and Human Services (2016). Comprehensive child welfare information system. Federal Register, 81(106). P. 35450.

[11] Harrison, T. M., Pardo, T. A., & Cook, M. (2012). Creating open government ecosystems: A research and development agenda. *Future Internet*, *4*(4), 900-928.

[12] Joseph, R. C., & Johnson, N. A. (2013). Big data and transformational government. *IT Professional*, *15*(6), 43-48.

[13] Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78-85.

[14] KPMG Government Institute (2016). KPMG Federal CCIS final rule spurs the next generation of child welfare information systems KPMG Government Institute – Issue Brief July 2016).

[15] Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), 1203-1205.

[16] Lee, Y., Madnick, S. E., Wang, R. Y., Wang, F., & Zhang, H. (2014). A Cubic Framework for the Chief Data Officer: Succeeding in a world of big data. MIT:ESD. ESD-WP-2014-34.

[17] Markus, M.L., & Tanis, C. (2000). The enterprise system experience – From adoption to success. In Framing the domains of IT Research: Glimpsing the future through the past 173(2000), 173-207.

[18] Mervis, J. (2012). Agencies rally to tackle big data. Science, 336 (6077), p. 22. doi: 10.1126/science.336.60777.22

[19] Olson, David. (2009). Enterprise information systems: Contemporary trends and issues. Hackensack, NJ: World Scientific Publishing Company.

[20] Otto, B. (2011). A morphology of the organization of data governance. ECIS 2011 Proceedings. Paper 272. http://aisle.aisnet.org/ecis2011/272

[21] Pardo, T. A., Gil-Garcia, J. R., & Burke, G. B. (2006). Building response capacity through cross-boundary information sharing: The critical role of trust. *Exploiting the knowledge economy: Issues, applications, case studies*, 507-514.

[22] Pardo, T.,Gil-Garcia, R. and Burke. B. (2007). Sustainable Cross-Boundary Information Sharing (Chapter XXI, pp. 421-438). In Hs. Chen, L. Brandt, S. Dawes, V. Gregg, E. Hovy, A. Macintosh, R. Traunmüller, and C. A. Larson (Eds). *Digital Government: E-Government Research, Case Studies, and Implementation*. Springer.

[23] Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (pp. 42-47). IEEE.

[24] Saleski, A. (2017). Unlocking the value of a chief data officer. CIO Review.

[25] Schilling, Joseph & Pinzón, Jimena (2016). "The basics of blight. Recent research on its drivers, impacts, and interventions". VPRN Research & Policy Brief No. 2. Washington DC: Vacant Property Research Network.

[26] Simon, H. (1997). Administrative behavior: A study of decision making processes in administrative organizations, 4th ed. NY: The Free Press.

[27] Sivaprakasam, S. R. (2010). Enterprise Data Management. Infosys Viewpoint.

[28] Susha, I., Janssen, M. and Verhulst, S. (2017a), "Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: taxonomy development", *Proceedings of the 50th Hawaii International Conference on System Sciences in Waikoloa*, *HI*

[29] Susha, I, Janssen, M, and Verhulst, S. (2017b) "Data collaboratives as "bazaars"?: A review of coordination problems and mechanisms to match demand for data with supply", Transforming Government: People, Process and Policy, Vol. 11 Issue: 1, pp.157-172.

[30] Yiu, C. (2012). The big data opportunity. *Policy Exchange*, 1-36.