

Problem Conceptualization as a Foundation of Data Analytics in Local Governments: Lessons from the City of Syracuse, New York

Felippe A. Cronemberger
University at Albany, State University of New York
fcronemberger@albany.edu

J. Ramon Gil-Garcia
University at Albany, State University of New York &
Universidad de las Americas Puebla
jgil-garcia@ctg.albany.edu

Abstract

The use data and data analytics (DA) has been attracting the attention of academics and practitioners in the public sector and is sometimes seen as a potential strategy for process and service innovation. While research on the many possible uses of data have clearly increased - open data, big data, data analytics-empirical research on the socio-technical process that local governments followed when using data analytics to improve services and policies is still scarce. Based on existing literature about data analytics in the public sector and the data lifecycle concept, this paper examines how data analytics is actually used in a local government and what are the main steps in this process. It analyzes the experience of a mid-size American city that had a dedicated task force to data analytics use to support decision making at the local level – Syracuse, New York. Findings suggest that data analytics as a process not only involves data analysis and representations (such as visualizations), but also data collection and cleaning. Further, it seems clear that the conceptualization of the problem is a critical step in producing meaningful data analytics, but also in thinking about innovations even when data is not readily available.

Keywords: Data Analytics, Innovation, Data Lifecycle, Problem Conceptualization, Collaboration, Government Data

1. Introduction

Data analytics (DA) has been attracting the interest of academics in practitioners in last few years. In different fronts, from business [1] to policy-making [2], such interest has been nurtured by emerging analytical technologies that can help to more effectively handle data and transform it into information for decision-making purposes. This fact is not new and it is tied to more than 20 years of information systems literature dedicated to maximizing the value of data in information environments [3]. The topic is also mentioned in the public sector literature, where its

coverage is more interdisciplinary and is tied to public administration concerns such as information management and governance [4][5], information sharing and integration [6], and information policy [7].

More recently, increasing attention has been dedicated to local governments, where the need to find “smart ways” to address ever growing public issues is pressing. In order to respond to specific needs, which include, for example, emergency preparedness [8] and transportation efficiency [9], reliance on data and evidence-based decision-making has become an important trend. In practice, such interest is reflected in dedicated efforts to data-driven policy making in smart city initiatives [10], in open data initiatives [11], and in better ways of using technology and consuming information [12].

Even though interest is clearly increasing at the practitioner side of DA use [13][14], research examining empirical cases understanding success and highlighting the limitations of existing DA practices is still scarce. Knowledge about local governments that systematically use data and DA is especially limited [15], with even less attention being given to small and midsize jurisdictions [16].

In particular, DA practices in those jurisdictions are yet to be explored as a socio-technical process in the context of data management and its lifecycle. The framework, already established in the realm of information systems implementation [17] and often referred as Data Management Body of Knowledge (DMBOK) comprises important practices in making data use more efficient and effective [18]. Those practices intend to add structure to the process for organizing and curating data, so they can then be analyzed to address a given problem. Later stages seem to display an overlap in literature with DA and data science and are commonly mentioned in the context of data warehousing and business intelligence [19].

In contrast with the mentioned established frameworks, practices focused on problem conceptualization seem to be overlooked, not explored empirically, or approached as a purely technical issue, instead of a strategic one [11][40]. This could be problematic to public organizations that are attempting to develop capabilities for data-driven problem-solving,

but whose problems are not necessarily approachable from an enterprise standpoint, either because of resource scarcity, lack of adequate data, or the “wicked” nature of the problems they have to deal with [52]. Therefore, the research question guiding this study is: How do local governments in small jurisdictions engage in data analytics practices and how these practices are related to data management concepts and processes?

Based on concepts of data management and the data lifecycle, this paper attempts to provide a more comprehensive socio-technical view of data analytics as a process and transformational practice to public organizations at the local level. The experience of Syracuse, New York, a city that engages in fostering data-driven practices in its policy-making endeavors, is used to illuminate the topic and define what is worth pursuing next. One of the main findings is the foundational importance of problem conceptualization as one of the first step in preparation for data analytics, but also as a way to frame an issue even if at the end the data analytics process is not feasible for that specific problem.

This paper is organized in seven sections, including the foregoing introduction. Section two includes a review of recent literature on data analytics in the public sector. It also includes a description and explanation of the data lifecycle and how it can be used as a framework to study data analytics as a process. Section three describes the research design and methods used in this study. This study is based on the case of Syracuse, which is a small-medium city in the state of New York and section four contains a brief description of the case. Section five presents the analysis and main results in terms of the main concepts proposed in the literature review. Section six discusses our findings and provides some implications for research and practice. Finally, section seven presents our conclusions and suggests ideas for future research about this topic.

2. Related Research

This section presents a review of recent literature with a focus on data analytics in the public sector and the data lifecycle as a way to frame our proposal of data analytics as a process.

2.1 Data Analytics in the Public Sector

Data analytics has been referred to in many ways and with relatively little consensus. Chen, Chiang, and Storey [20] state that “data analytics refers to the BI&A (Business Intelligence and Analytics) technologies that are grounded mostly in data mining and statistical analysis.” While not too many studies have been

dedicated to definitions in the public sector, it has been referred to in different ways, such as “policy informatics” [21], “policy analytics” [22], “computational social science” [23], and “government data analytics” [24]. Despite varying definitions, it seems the commonality is a concern with data use to address knowledge and organizational problems both in the private sector [25] and in the public [14].

Interest in DA seems to be related to the realization that good use of data and information, increasingly perceived as being more abundant than ever before, has the potential to facilitate problem-solving in topics where it has never been used before. In research, such interest has been materialized in studies that either acknowledge DA as a mainly technological endeavor to produce information products or address it at a more fundamental level as a transformational initiative with profound implications for organizational routines [26].

In the public sector, where data and information needs and use are known to be especially challenging [27], research on data analytics practices has been broad and sparse. Promising uses of data have been focused on making data more accessible through open data [28] and in technical means of extracting data from such open or big data [29]. While the research on open data does not often establish a link between open data and data analytics, it can be the case that efforts in opening data may potentially add value by expanding data resources for analysis [29]. More commonly, research has focused on the technical infrastructure concerns for data management in the context of the Internet of Things [27] or on the study of how people collaborate around data [28].

2.2 Local Governments and the Use of Data Analytics

For local governments, research has explored technological improvements that make cities smart [32]. Noticeably, however, most of these research could be classified as technological determinism; a paradigm that may overlook other necessary capabilities for local governments that are trying to take data and data analytics use to the next level.

Those capabilities may be especially necessary in a context where multiple stakeholders may need to collaborate to understand public issues and design a solution [33]. That logic characterizes the concept of living labs, which are initiatives dedicated to innovation processes that are public and open [34] and endeavors focused on gathering support to public projects, such as crowdfunding [35]. While not directly referred to as data analytics in local government literature, research suggests that collaborative efforts appear to play a crucial role in helping public leadership and citizens to

collaborate around data [31] or even to help collecting it and producing it for subsequent use [36].

Research that more directly refers to analytics does not necessarily cover public sector experiences in specific, but has clearly highlighted the importance of investigating data analytics capabilities in the light of “human” and “intangible” resources [37]. That would include concerns with finding the right talent [38], and assessing relationships with factors like “organizational culture” and “top management commitment” [39].

While most previous research has focused on developing a vision and identifying factors that can be used to build that vision, few studies have considered that cities may be at different stages of development in their data analytics agenda [40]. Different organizational realities, as different local governments across the world are likely to represent, require specific assessments on existing capabilities. As they evolve and develop towards becoming more data-driven at their own pace, their ability to leverage data analytics may follow [41], but empirical data on how local governments become data savvy are still limited.

2.3 The Data Life Cycle as a way to Understand Data Analytics

Data use is a topic that is pervasive across multiple disciplines and domains. Due to the variety of applications and contexts in which data can be used, the study of data use spans experiences in business, public administration, and a variety of disciplines where particularities of data manipulation use could be observed. The inter-disciplinarity of the topic makes it more challenging, but also enriches information science and other disciplines around the use of data.

Data lifecycles frameworks are commonly used to study practices and processes in which data are used. Those frameworks help to design goals, define stages for data manipulation, and map expected outcomes for data use. Research has approached the topic in basically two ways. The first way encompasses comprehensive research, mostly dedicated at refining established models and enhance the explanatory power of existing theories in the light of new empirical evidence. That is accomplished through the proposal of new frameworks that are focused on providing a more holistic view [42]. The second approach involves the scrutiny of specific stages of the data lifecycle, where research explores and expands the view on particular concepts such as data quality [43] or data collection [43].

Most frameworks outline similar stages and definitions of stages for data use. Those stages are generally referred to as 1) data collection, generation or creation [44]; 2) data cleaning and curation [45], 3) data

analysis [46], and 4) data management [46]. As explained before, each one of the stages is explored in research with varying levels of depth and examined in specific contexts from distinctive theoretical angles. Such endeavors suggest that, given the multitude of constructs and meanings present in a dynamic and still evolving research domain, examinations of data analytics are typically fragmented or not explicitly made.

The socio-technical complexity of the topic and the relative scarcity of research on data use in the public sector [40], suggest that studying DA in the context of the data lifecycle in the public sector could become an important research agenda. Much is yet to be learned from different stages and experiences, particularly those in which data use is directly linked or referred to data analytics practices and technologies. Ku and Gil-Garcia [13], for instance, found that collection is a critical part of data analytics practices in local governments and argue that only when the necessary and adequate data are collected, the analysis becomes feasible and more useful.

2.4 Data Analytics as a Process in Local Governments

Efforts to analyze DA practices more systematically are still emerging and are often based on case studies. Those efforts involve connecting established literature on data management to emerging concerns on whether existing infrastructure can in fact produce value in the public sector [40]. While some research discusses the importance of data management from a leadership and governance perspective [40], others explore the role of collaboration and coordination in what could or should be contextually done with the data [31].

Mergel, Rethemeyer and Isett [16] argue that analytical capability development in the public sector can be linked to three goals: a) manage and process of large accumulations of unstructured, semi-structured, and structured data; b) analyze that data into meaningful insights for public operations; c) interpret that data in ways that support evidence-based decision making. According to the authors, those capabilities are especially needed in “small jurisdictions” [16].

Research on data analytics has focus on both the relevance of data analytics to achieve smart governments and the factors that may influence the success of data-driven initiatives [14]. Across different articles, the general view is that, prior to successful execution of analytics at the technical level, when all the data is in place and ready to be used, organizational factors, such as leaders support and governance efforts to collect data, are critical [47].

Although the model is simple and closely based on the data lifecycle perspective, we argue that it offers clarification about DA as a process, its main stages, and some of the factors that affect this process (see Figure 1). As research in socio-technical systems suggests, many of the factors and conditions in the process are dynamic. Besides changing over time for external reasons, they are also affected by the interactions and interrelationships among them. Further analysis may help clarifying complexities in the light of empirical evidence.

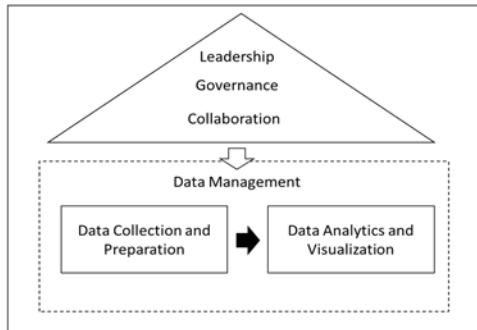


Figure 1. Data Analytics as a Processes in Local Governments

The model reflects elements identified in the literature as being crucial to data analytics practices in local governments. Data analytics practices involve two data management processes: a) data collection and preparation, at the input side of DA; and b) data analytics and visualization, as the output of information products for consumption. Those practices are influenced by three key constructs: leadership, governance and collaboration. Leadership is known to foster strategic use of information in the public sector. Collaboration is a foundational principle, playing a central role in enabling inter-organizational knowledge and information sharing. Finally, governance is considered to play a central role in orchestrating socio-technical efforts in terms of rules and arrangements that maximize benefits from the strategic use of information.

3. Research Design and Methods

This paper is based on a case study, which are a good research strategy for relatively new topics [57]. Within this overall approach, several data collection methods were used. First, semi-structured interviews with stakeholders involved in data analytics initiatives in the selected case were the main source of information. The profile of respondents included analysts, policy-makers, and leaders of initiatives. The questions in the interviews were related to what factors respondents

perceived to have an impact on their ability to use data analytics. They were asked to elaborate on their opinions about data, technical, organizational, institutional, and contextual aspects. Generally, answers were coded according to each dimension as pertaining to specific variables. However, following grounded theory procedures [58][59], respondents were also encouraged to provide additional details and topics they considered relevant, and, in those circumstances, new variables and narratives emerged.

Second, a session with community leaders involved in code enforcement for housing provided a topic-centered example of how City Hall goes about approaching problems of public interest analytically and with a participatory approach. The session was observed using a protocol that included (1) paying attention to the discussion of the topic being conducted and (2) how people involved in data analytics approached the issue being discussed. Finally, perceptions were supplemented with document analysis in openly available electronic sources such as the What Works City [48] website and, in specific, the i-Team (Innovate Syracuse) [49].

The conceptual model presented at the end of the literature review section is to be used to guide the exploration of the case. Understanding the main constructs in the model, such as the role of leadership and collaboration, will be the focus of the empirical analysis. The data were analyzed based on pre-established categories based on the literature, particularly the proposed conceptual model. However, new concepts and relationships that emerged from the qualitative data were also considered, as pertaining to the particularities of local governments as a research setting. Although open data is one of the resources used in Syracuse for data analytics, the focus of this paper is on the process of analyzing data from any source, whether open or restricted to internal use.

4. Brief Description of the Case

Syracuse, New York engaged in a nationwide program called Bloomberg's What Works Cities initiative (WWC). Through the program, the city became part of a network of 100 municipalities that committed to "enhance their use of data and evidence to improve services, inform local decision-making and engage residents" [48]. This endeavor is interested in fostering "best practices" across selected cities, "helping local leaders identify and invest in 'what works'" [48]. Chosen cities are formally certified, receiving funding and support to foster data-driven initiatives from practitioners and researchers affiliated with research centers such as the Government Performance Lab at

Harvard Kennedy School of Government and non-partisan organizations such as the Sunlight Foundation.

The WWC program uses a framework that is grounded in four pillars: 1) Commit with goals; 2) Measure; 3) Take a stock; and 4) Act. Under each pillar, several steps to achieve goals set for each pillar are outlined. Participants are expected to embrace the WWC mission and comply with such standards. Under those guidelines, DA sponsor and champions from each city reconvene at a yearly summit where practices and experiences are shared and discussed. Cities have their individual goals and priorities. For instance, Syracuse's involvement with WWC aimed to "improve open data practices and establish and improve performance management programs to improve results for residents." [48].

In 2015, the City of Syracuse established the Innovation Team (i-Team). Under the motto "Changes call for innovation, and innovation leads to progress," the Innovation Team focuses on initiatives to address public infrastructure problems and foster local economic development. Initiatives include the installation of street quality electronic devices that address infrastructure issues such as street potholes and the development of an early detection system for water infrastructure problems. The city has piloted several smart city technologies and is advancing steadily in data collection and DA practices. The i-Team also created a blog to share information about current and past projects. In a segment of a post, the Innovation Team acknowledged that the City of Syracuse "will need much more than to simply buy technology" to become a smart city [49].

5. Analysis and Results

This section presents the main results of our analysis. Overall, data analytics in Syracuse could be understood as a process in which data analysis and visualizations are important aspects. However, data collection and cleaning also emerged as a large part of data analytics efforts. Finally, it seems clear that the conceptualization of the problem is a very important step in the whole process and something useful to do even when data is not entirely available for conducting a more data-oriented analysis of the situation.

5.1 Data Analytics as a Data Management Process

Syracuse seems to have a highly collaborative and interdisciplinary team that understand data management and analytics issues. Self-starters and with a can-do attitude from leadership to policy-design and analysis,

the team benefits from a combination of skills: data analytics practitioners and researchers from the academia who works in a consulting capacity are coordinated by a Chief Innovation Officer and a Chief Data Officer. Syracuse's Innovation Team operates inside the City Hall with a structure that is similar to a think-tank, with people approaching problems with autonomy while conceptualizing and tackling data problems and acting towards priority policy-goals.

At the time this study took place, housing issues were considered a major concern. Later as a "housing instability" issue, interviews revealed that DA endeavors were mostly dedicated to collecting housing and code enforcement data. In this context, for example, one key goal was to understand causes of "high frequency of forced moves" faced by citizens and mitigate consequences such as chronic homelessness and damaging financial and health impacts to the community [50]. Clear understanding of priority goals were critical to start applying DA to problem-solving and Syracuse have a process approach to DA.

5.2 Data Collection and Preparation

Data analytics endeavors could be divided in two processes. First, collecting statistical data on occupation and eviction rates. Work contained data sampled across different regions but mostly within the city of Syracuse. Such data was expected to help with problem definition and point to directions to be followed and specific policy alternatives.

Second, in a transition from the "what" to the "why" questions, session with code enforcement personnel would help understanding the reality of citizens. In order to listen to the "voice of the residents", the team would hold meetings with community members in their office or visiting sites to personally collect data and inform themselves about the issues that needed to be addressed. They seem to be actively involved with in-person data collection efforts, often producing data analytics products and consuming them as they learn about the problems. According to respondents, such data was scattered across different local governments organizations, often remained uncollected or inaccessible to those who needed to use them. Much of the Innovation Team efforts went to learning what kind of data exists and to what extent inter-organizational partnerships could help them access those datasets. This is consistent with previous research on the role of data availability and collection [13][18] and more traditional digital government research on collaborative networks and their role in information sharing and use [51][52][53].

5.3 Data Analysis and Visualizations

Data analytics practices appeared to rely on multiple iterations so desired results could be achieved. Data analytics practitioners seemed to value the tools that helped them with the numbers, but their tasks appeared to be based on a dynamic collection-analysis routine. In that sense, the more data they had or analyze, the more they needed to dig deeper on the questions and issues being investigated. Such part of the work appeared to be both qualitative and quantitative, with different proportions across distinct DA roles. For instance, analysts working on similar or on the same projects were found to have quantitative and qualitative skills, and using them as needed and as the project evolved.

One aspect that seem to explain the dynamics of the organizational practices in terms of data analysis is that not only information for every problem did not appear to be easily accessible, but, when retrieved, they were not easily manipulated. At the analysis side of the process, hurdles with formats and missing data seem to require manual adjustments, as well as constant validation with the data sources. Analysts were often holding meetings and collaborating with relevant stakeholders so information could be corrected or better contextualized for subsequent analysis. As one interviewer said, it is a matter of verifying if their understanding of the information they have access to is correct, and, if not, identifying what is missing, and who could help addressing that.

Finally, it must be highlighted that DA practices in Syracuse were not centered at any specific type of technology. Rather, as most interviews claimed, they were centered at both structured and unstructured data, as well as on people's understanding of it. Data science technologies were used by analysts and leaders, but mostly in an *ad hoc* fashion. As three interviewees stressed, for instance, many answers are really on data, and most of it are either in non-computerized form or yet to be collected. Such understanding, as pointed out by both leaders and analysts, is what justifies DA analytics initiatives as field work, where investing in ties with communities as invaluable sources of data becomes a fundamental part of the their analytical endeavor.

The visualization portion of data analytics deserves especial attention, since it does not always have to be digital. Post-it notes, charts, graphs and maps were spread across the walls and were constantly referred to when discussing specific problems. As supporting elements for storytelling and scenario construction, they acted as boundary objects [54][55], helping analysts in the routine assessment of their challenges and goals by being openly displayed in their work environments. In addition, those elements were not only visual references for data-driven problem solving, but also artifacts that

stimulate debate and reflection. Analysts would constantly revisit those visual elements to check their collective understanding on issues and occasionally make new observations as their thinking evolves. Through the flexibility of this iterative, investigative process, data analytics practitioners could expand upon existing data and extend the scope of their information sources.

5.4 The Importance of Problem Conceptualization

As far as their technical experience with DA use, one important aspect relates to the qualitative nature of their DA work. Interviewing and coding of unstructured data were common and a central aspect of the conceptualization phase of the problems. In addition, data analytics in Syracuse appeared to be highly interdisciplinary, human-centered and focused on multiple sources for data collection and analysis. However, it is now clear if this is the case all the time or if it depends on the nature of the topics and the specific people involved.

Much of the data collected has the potential to inform policy-design as well as processes through which more data could be collected. That was observed through a session with code enforcement personnel which was held to not only conceptualize the problem, but also to identify gaps in their data collection processes. Following a debate-mediated structure, the I-Team team attentively listened to concerns and opportunities for improvement and took notes on perceptions. Notes were later used to define what next steps for data collection and use could or should be.

Another evidence that suggests the foundational role of problem conceptualization comes from the fact that interviewees were very clear about the scarcity of the necessary information. To many of them, that meant that parts of the problem to be solved were not fully understood and needed further investigation. They attempt to better conceptualize the problem and characterize the process as "learning-on-the go", which involved interviewing citizens and organizing workshops in the city hall to obtain more information about a particular problem. It is not clear if these efforts are enough to complete the necessary data, but they could be considered a useful first step.

Finally, the information scarcity also created incentives to conduct in-person visits to partnering agencies, where key staff members held important data sources or have access to them in legacy IT systems. According to respondents, the opportunity to understand information generation processes also contributed to add context to available sources and, in doing so, facilitate problem conceptualization. Data analytics

outputs would benefit from this effort because a more informed analysis could lead to richer perspectives on limited data.

5.5 The Impact of Leadership, Governance, and Collaboration

Interviewees mainly mentioned the role of leadership in two ways. First, at a more macro, political level, leadership support helped kicking the WWC initiative off, sending a strong message to the public and the public servants on the extent to which the city government would prioritize data-driven approaches to policy-making. Most interviewees indicated that the very existence of the i-Team depended on that.

Second, leadership was understood as a proxy to being “self-driven” and “entrepreneurial” towards data analytics practices. Given the often stated limitation of resources and the difficulties in having access to them – information, included -, respondents seemed to hold a high level of independence and accountability in the projects they were conducting. Interviewees revealed that they enjoyed the freedom to go after the data they need by visiting communities, by holding interview sessions, and by thinking about innovative ways of addressing problems. While many confirmed that such level of engagement was encouraged from the top, a few respondents observed that that required a collective sense of involvement with the problems the team were working on.

Collaboration was often mentioned as a crucial element for data analytics work, which is consistent with research on collaborations around data [31][33]. Respondents stated that the nature of their work is collaborative and relies on participation of both analysts and stakeholders involved by the issue being addressed using data. By having people collaborating around data, new understandings of existing data and of the problem emerged. While mentioned as being key, such collaboration was not referred to as coming naturally. Rather it is a result of concerted efforts to bring people to the same room and get them involved. That effort was not without hurdles because people and the data needed to address a particular problem were in different organizations. Therefore, their participation required leadership and resources, a condition that was at least partially addressed, mitigated or fostered by the WWC initiative.

Finally, governance did not emerge as important to people directly involved in data analytics. However, two elements observed by respondents could be indirectly related to governance. First, the need to have better institutional mechanisms that can bring people together to make data available for use and data analytics more effective. Second, coordination emerged as being an

important factor. Respondents observed that the unstructured nature of their work demanded some level of self-coordination and a go-get attitude towards data and relevant information for subsequent analysis.

6. Discussion and Implications

This section discusses some of the main results of this study and presents a few implications for research and practice. Overall, results suggest that data management is very important for data analytics [17][40]. DA as a process is the result of a dynamic practices that unfold and adapt over time. Such development occurs as data analytics practitioners explore existing and new sources of data. With that exploration seems to come a greater level of understanding on the problem being solved. Data management practices in data analytics may be more reliant on the ability to collaborate and deal with unstructured information, than with assuming that all the data needed to solve a particular problem will be available [40]. This is consistent with previous research about the role of information sharing for successful digital government projects [6][51].

6.1. The Fundamental Importance of Problem Conceptualization for Data Analytics

Syracuse’s experience is enlightening for several reasons. First, evidence suggests that DA analytics is clearly not centered at technological artifacts only [11], but also at interactive practices through which raw data and information on public problems are shared and used. In this process, the conceptualization of problems help defining the guiding steps for data collection and, subsequently, to data analytics outputs. Such process is markedly iterative, with multiple sessions dedicated to knowledge and information sharing, what leads to a more complete understanding of issues.

There is also evidence that problem conceptualization is a central practice and, could also be considered a foundational step of DA as a process, which has not been specifically highlighted in previous DA research [13]. Such effort are normally qualitative in nature, relying on multiple iterations through unstructured data, and facilitate collective sense-making and alignment with regards to what exactly is the problem that needs to be addressed.

As per the findings, the problem conceptualization phase of DA needs examination at different levels, including the role of political leadership and the public in setting priorities for data analytics use. Externally oriented initiatives such as WWC, seem to speak loudly and have political endorsement at the local level.

Internally, a focus on problem conceptualization and data needs seems to prevail over technology, but also seems reliant on soft and organizational aspects, including the ability to bring different stakeholders to collaborate around data and produce innovative and useful information.

6.2. Towards an Expanded Model of Data Analytics as a Process in Local Governments

A theoretical view that positions DA both in the macro view of technological frameworks for data and information use and in the conceptual explorations of data use as a socio-technical practice, may lead to a variety of models. An expanded model of DA use based on Syracuse experience should be attentive to such paradigm (Figure 2). Considering the empirical evidence from our case, two key adaptations to the proposed initial framework are necessary.

First, in line with literature in innovation and smart governments [56], leadership is confirmed as a driving force in setting the direction for data analytics use. As observed, such leadership starts with political commitment with the DA agenda and is supported along the way by WWC partners and City Hall champions involved in the effort. Since members of the Innovation Team are part of a highly collaborative environment and appeared to enjoy autonomy in their responsibilities with the project, leadership in DA analytics could be interpreted as a force triggered by a few actors, and stewarded by collaborators inside and outside the City Hall.

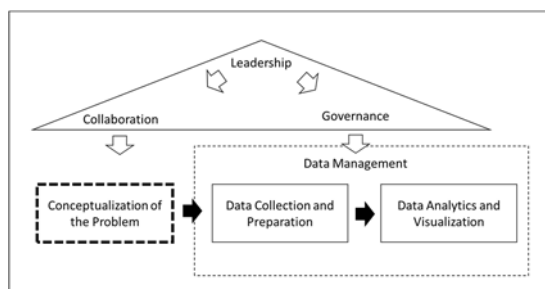


Figure 2. An expanded model on Data Analytics as a Processes in Local Governments

Second, as mentioned before, the problem conceptualization phase should be clearly identified as an important initial step of the data management cycle and for defining what needs to be addressed through DA [11][18]. It also seems appropriate to link collaboration, a construct that is not new to digital government literature and continues to be studied, and conceptualization efforts. Since research in problem definition and conceptualization in the context of data analytics is still limited [13], those theoretical linkages

could be further explored empirically. One possibility is to further divide conceptualization in data-driven problems, defining, for instance, which stakeholders and what conditions should be in place to ensure and effective DA process in local governments.

As suggested by the Syracuse experience, the existence of *ad hoc* approaches to DA indicates that some operational flexibility is needed to accommodate multiple ways to define and address problems with data. The relatively newness of the topic [1][2], particularly in terms of local government experiences, could benefit from flexibility in defining models for DA use. Also, more categories or lenses could be added in the light of advancements. New concepts and their relationships should also be considered for the purpose of revising our proposed model, but also, more generally, expanding our knowledge about DA and its impact on government innovations in services and policies.

7. Conclusion

This paper contributes to research and practice by highlighting the fundamental role that problem conceptualization plays in data analytics. This emphasis was not found in existing literature and was extremely important in our case study. Since research in DA is fragmented and multidisciplinary, constructs such as the one highlighted in this paper may be observed from different perspectives or simply overlooked. Therefore, it is important to pay attention to emergent concepts in innovative research. Furthermore, it must be observed that hopes associated to producing more information with data had not been adequately addressed, and existing research has not captured all relevant factors, observed direct consequences, or considered whether satisfying results are being obtained. This gap is to a certain extent surprising, given the recent focus on smart cities and data-driven decisions in the public sector. A focus on important constructs, such as problem conceptualization, has been missing in research about DA in local governments, so this paper specifically contributes to data analytics research and practice, by providing evidence of the importance of problem conceptualization.

Understanding DA as a process affected by multiple and emerging capabilities requires taking a socio-technical perspective. The contrast between a futuristic, and sometimes unrealistic, view on how a data-driven city could function and the much needed process-oriented perspective on how DA actually operates in governments is an important contribution of this study too. Particularly in the case of local governments, the lack of lessons and guidance about the importance of data management for data analytics suggests that practitioners, especially in the short-term, may benefit

from a growing number of studies where similar experiences with DA and specific technologies are analyzed. In this respect, this study contributes to the perspective that DA in local governments is not centered at technological artifacts only, but also includes data and problems in which multiple stakeholders are involved.

A single case study is always enlightening, but the results are far from being generalizable, for experiences are context-sensitive and are never to be applicable unambiguously to different contexts. However, results and implications of this study could be useful to similar realities, particularly, to small and medium cities already using or planning to use DA as a strategy to promote innovations in services and policies.

In addition, the final conceptual model is only a set of propositions that could be empirical tested by using quantitative data or by conducting additional case studies and this could also be seen as a limitation of this study. Future research should do that and explore different context for DA use in local governments. Given the focus of our study, there were several topics that we not studied in depth. In fact, more research could be dedicated to answering questions such as: What has been done in terms of DA in other cities? Are there any differences and similarities when referring to larger or smaller jurisdictions? Is problem conceptualization always as important as a foundation for DA in local governments? Why or why not? Responses to these and other related questions would help to better understand DA as a process in local governments and how a socio-technical perspective based on the data lifecycle model and its stages could be useful to generate new insights into this topic.

8. References

- [1] N. Côte-Real, P. Ruivo, T. Oliveira, and A. Popovič, "Unlocking the drivers of big data analytics value in firms," *J. Bus. Res.*, vol. 97, pp. 160–173, 2019.
- [2] J. R. Gil-Garcia, T. A. Pardo, and L. F. Luna-Reyes, "Policy Analytics: Definitions, Components, Methods, and Illustrative Examples," in *Policy Analytics, Modelling, and Informatics*, Springer, Cham, 2018, pp. 1–16.
- [3] T. H. Davenport and D. D. Dreyfus, *Information ecology: Mastering the information and knowledge environment*. New York: Oxford University Press, 1997.
- [4] J. E. Fountain, "Building the virtual state," *Information technology and institutional change*, pp. 61–82, 2001.
- [5] D. F. Andersen and S. S. Dawes, *Government information management: A primer and casebook*. New Jersey: Prentice Hall, 1991.
- [6] J. R. Gil-Garcia and D. S. Sayogo, "Government inter-organizational information sharing initiatives: Understanding the main determinants of success," *Gov. Inf. Q.*, vol. 33, no. 3, pp. 572–582, 2016.
- [7] S. Braman, "Defining information policy," *J. Gov. Inf.*, vol. 1, pp. 1–5, 2011.
- [8] S. Chaudhari, A. Bhagat, N. Tarbani, and M. Pund, "Dynamic Notifications in Smart Cities for Disaster Management," in *Computational Intelligence in Data Mining*, 2019, pp. 177–190.
- [9] C. Kaptan, B. Kantarci, T. Soyata, and A. Boukerche, "Emulating Smart City Sensors Using Soft Sensing and Machine Intelligence: A Case Study in Public Transportation," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–7.
- [10] V. Moustaka, A. Vakali, and L. G. Anthopoulos, "A systematic review for smart city data analytics," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, p. 103, 2018.
- [11] G. Puron-Cid, J. R. Gil-Garcia, and L. F. Luna-Reyes, "Opportunities and challenges of policy informatics: tackling complex problems through the combination of open data, technology and analytics," *International Journal of Public Administration in the Digital Age (IJPADA)*, vol. 3, no. 2, pp. 66–85, 2016.
- [12] R. Matheus, M. Janssen, and D. Maheshwari, "Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities," *Gov. Inf. Q.*, 2018.
- [13] M. Ku and J. R. Gil-Garcia, "Ready for data analytics?: data collection and creation in local governments," in *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, 2018, p. 36.
- [14] F. Cronemberger and J. R. Gil-Garcia, "Big Data and Analytics as Strategies to Generate Public Value in Smart Cities: Proposing an Integrative Framework," in *Setting Foundations for the Creation of Public Value in Smart Cities*, M. P. Rodriguez Bolivar, Ed. Cham: Springer International Publishing, 2019, pp. 247–267.
- [15] J. Aguilar, "Tutorial: Data analytics in the domain of smart cities and e-Government," in *eDemocracy & eGovernment (ICEDEG), 2016 Third International Conference on*, 2016, pp. 1–4.
- [16] I. Mergel, R. K. Rethemeyer, and K. Isett, "Big data in public affairs," *Public Adm. Rev.*, vol. 76, no. 6, pp. 928–937, 2016.
- [17] J. P. Ohr, C. J. E. Chandler, and G. B. O. Bunker, "System and method for policy-based data lifecycle management," 7680830, 16-Mar-2010.
- [18] T. M. Harrison *et al.*, "The Data Firehose and AI in Government: Why Data Management is a Key to Value and Ethics," in *20th Annual International Conference on Digital Government Research (dg.o 2019), June 18–20, 2019, Dubai, United Arab Emirates*. ACM, New York, NY, USA., Dubai, United Arab Emirates, 2019, pp. 1–6.
- [19] D. Larson and V. Chang, "A review and future direction of agile, business intelligence, analytics and data science," *Int. J. Inf. Manage.*, vol. 36, no. 5, pp. 700–710, Oct. 2016.
- [20] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," *Miss. Q.*, vol. 36, no. 4, pp. 1165–1188, 2012.
- [21] S. S. Dawes and N. Helbig, "The Value and Limits of Government Information Resources for Policy Informatics," *Governance in the Information Era*:

- Theory and Practice of Policy Informatics*, p. 25, 2015.
- [22] J. Höchtl, P. Parycek, and R. Schöllhammer, "Big data in the policy cycle: Policy decision making in the digital era," *Journal of Organizational Computing and Electronic Commerce*, vol. 26, no. 1–2, pp. 147–169, 2016.
- [23] R. M. Chang, R. J. Kauffman, and Y. Kwon, "Understanding the paradigm shift to computational social science in the presence of big data," *Decis. Support Syst.*, vol. 63, pp. 67–80, 2014.
- [24] E. Kalampokis, E. Tambouris, and K. Tarabanis, "Linked open government data analytics," in *International Conference on Electronic Government*, 2013, pp. 99–110.
- [25] T. H. Davenport, J. G. Harris, D. W. D. Long, and A. L. Jacobson, "Data to knowledge to results: Building an analytic capability," *Calif. Manage. Rev.*, vol. 43, pp. 117–138, 2001.
- [26] C. Loebbecke and A. Picot, "Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda," *The Journal of Strategic Information Systems*, vol. 24, no. 3, pp. 149–157, 2015.
- [27] E. W. Welch, M. K. Feeney, and C. H. Park, "Determinants of data sharing in US city governments," *Gov. Inf. Q.*, vol. 33, no. 3, pp. 393–403, 2016.
- [28] A. Zuiderwijk, M. Janssen, and Y. K. Dwivedi, "Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology," *Gov. Inf. Q.*, vol. 32, pp. 429–440, Oct. 2015.
- [29] M. Janssen, D. Konopnicki, J. L. Snowdon, and A. Ojo, "Driving public sector innovation using big and open linked data (BOLD)," *Inf. Syst. Front.*, vol. 19, no. 2, pp. 189–195, Apr. 2017.
- [30] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the internet of things using big data analytics," *Computer Networks*, vol. 101, pp. 63–80, 2016.
- [31] I. Susha, M. Janssen, and S. Verhulst, "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, no. 1, 2017.
- [32] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of things and big data analytics for smart and connected communities," *IEEE access*, vol. 4, pp. 766–773, 2016.
- [33] I. Tucker, J. R. Gil-Garcia, and D. S. Sayogo, "Collaborative Data Analytics for Emergency Response: Identifying Key Factors and Proposing a Preliminary Framework," in *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance*, New Delhi AA, India, 2017, pp. 508–515.
- [34] M. Gascó, "Living labs: Implementing open innovation in the public sector," *Gov. Inf. Q.*, vol. 34, no. 1, pp. 90–98, 2017.
- [35] S. Hong and J. Ryu, "Crowdfunding public projects: Collaborative governance for achieving citizen co-funding of public goods," *Gov. Inf. Q.*, vol. 36, no. 1, pp. 145–153, Jan. 2019.
- [36] A. Meijer and S. Potjer, "Citizen-generated open data: An explorative analysis of 25 cases," *Gov. Inf. Q.*, vol. 35, no. 4, pp. 613–621, Oct. 2018.
- [37] M. Gupta and J. F. George, "Toward the development of a big data analytics capability," *Information & Management*, vol. 53, no. 8, pp. 1049–1064, 2016.
- [38] H. Ekbja *et al.*, "Big data, bigger dilemmas: A critical review," *Journal of the Association for Information Science and Technology*, vol. 66, no. 8, pp. 1523–1545, 2015.
- [39] S. F. Wamba, A. Gunasekaran, S. Akter, S. J.-F. Ren, R. Dubey, and S. J. Childe, "Big data analytics and firm performance: Effects of dynamic capabilities," *J. Bus. Res.*, vol. 70, pp. 356–365, 2017.
- [40] T. Harrison, T. Pardo, M. Gasco-Hernandez, and D. Canestraro, *The Salience and Urgency of Enterprise Data Management In the Public Sector*. 2018.
- [41] M. Janssen, H. Van der Voort, and A. Wahyudi, "Factors influencing big data decision-making quality," *J. Bus. Res.*, vol. 70, pp. 338–345, 2017.
- [42] M. H. Ofner, K. Straub, B. Otto, and H. Oesterle, "Management of the master data lifecycle: a framework for analysis," *Journal of Enterprise Information Management*, vol. 26, no. 4, pp. 472–491, 2013.
- [43] B. Otto, K. M. Hüner, and H. Österle, "Toward a functional reference model for master data quality management," *Information Systems and e-Business Management*, vol. 10, no. 3, pp. 395–425, Sep. 2012.
- [44] V. Poe, S. Brobst, and P. Klauer, *Building a Data Warehouse for Decision Support*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.
- [45] H. V. Jagadish *et al.*, "Big data and its technical challenges," *Commun. ACM*, vol. 57, no. 7, pp. 86–94, Jul. 2014.
- [46] Y. Wang, L. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technol. Forecast. Soc. Change*, vol. 126, pp. 3–13, Jan. 2018.
- [47] A. T. Chatfield and C. G. Reddick, "Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services," *Gov. Inf. Q.*, Nov. 2017.
- [48] "What Works Cities," *What Works Cities*, 2017. [Online]. Available: <https://whatworkscities.bloomberg.org/>. [Accessed: 20-Feb-2017].
- [49] "Smart Cities," *Innovate Syracuse*. [Online]. Available: <http://www.innovatesyracuse.com/blog/smart-cities>. [Accessed: 23-Apr-2018].
- [50] "What is housing stability?," *Innovate Syracuse*. [Online]. Available: <http://www.innovatesyracuse.com/blog/whatishousingstability>. [Accessed: 19-Feb-2019].
- [51] J. R. Gil-Garcia, T. A. Pardo, and D. S. Sayogo, "From bureaucratic machines to inter-organizational networks: Characterizing the response to the World Trade Center crisis," *Transforming Government*, vol. 10, no. 4, pp. 568–590, Oct. 2016.
- [52] S. S. Dawes, A. M. Cresswell, and T. A. Pardo, "From 'Need to Know' to 'Need to Share': Tangled Problems, Information Boundaries, and the Building of Public Sector Knowledge Networks," *Public Adm. Rev.*, vol.

- 69, no. 3, pp. 392–402, May 2009.
- [53] S. S. Dawes, “Interagency information sharing: Expected benefits, manageable risks,” *J. Policy Anal. Manage.*, vol. 15, no. 3, pp. 377–394, 1996.
- [54] L. F. (orcid:000000020852404x) Luna-Reyes *et al.*, “Modeling and Simulation as Boundary Objects to Facilitate Interdisciplinary Research: Modelling and Simulation for Interdisciplinary Research,” *Syst. Res. Behav. Sci.*, Nov. 2018.
- [55] F. A. Cronemberger, D. S. Sayogo, and J. R. Gil-Garcia, “Examining Boundary Objects in Inter-Organizational Information Sharing (IIS) Success,” in *Proceedings of the 18th Annual International Conference on Digital Government Research*, 2017, pp. 167–176.
- [56] S. Mellouli, L. F. Luna-Reyes, and J. Zhang, “Smart government, citizen participation and open data,” *Information Polity*, vol. 19, no. 1, 2, pp. 1–4, 2014.
- [57] Robert K. Yin. (2014). *Case Study Research Design and Methods* (5th ed.). Thousand Oaks, CA: Sage. 282 pages.
- [58] Strauss, Anselm and Juliet Corbin. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage Publications.
- [59] Glaser, Barney G. and Anselm Strauss. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine Publishing Co, 1967.