San Jose State University
SJSU ScholarWorks

Master's Projects   Master's Theses and Graduate Research

Spring 5-2016

Implementing Data Analytics and Data Integration at the City of San José: An Ex-Post Comprehensive Evaluation of the Data Analytics Team (DAT)

Steve Guagliardo
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_projects

    Part of the Policy Design, Analysis, and Evaluation Commons, and the Public Administration Commons

Implementing Data Analytics and Data Integration at the City of San José
An Ex-Post Comprehensive Evaluation of the Data Analytics Team (DAT)

Submitted to the Graduate Faculty of the
California State University, San José
In Partial fulfillment of the
Requirements for the degree of

Master of Public Administration

By

Steve Guagliardo
B.A. University of California, Santa Cruz, 2008
May, 2016

Chapter 1: Introduction

Since the great recession of 2008, a recurring theme among municipal governments has been how to maximize efficiency and marginal utility. In an era with a heightened focus on public sector scarcity, the pressing question is how to ensure that the public's resources are being used as effectively as possible. In California, this has manifested through cost saving measures, which could be characterized as a prevention mindset (Heath and Heath, 2013). In the City of San José (CSJ), this approach not only resulted in a highly litigious variant of pension reform efforts, but also through service level reductions as a result of massive employee reductions and a reprioritization of the City's service deliveries. As CSJ slowly starts to regain its financial footing, it has begun to earnestly look at what internal operations can be improved in order to maximize the marginal utility of existing, and additional, programs. The careful analysis of current, and potential, programs represents a shift to a "promotional mindset" that seizes the opportunity inherent in the challenge of increased scrutiny and reduced resources (Heath and Heath, 2013). This project will perform an *ex post* comprehensive evaluation of CSJ's Data Analytics Team (DAT) which was implemented specifically to accomplish that goal.

In May 2015, CSJ instituted the Data Analytics Team (DAT) in the City Manager's Office. As part of the Mayor's June Budget Message for 2014-2015, as approved by the City council, CSJ authorized funding for two positions in the City Manager's Office to form the back-bone of DAT (Reed, 2014). The Mayor's Message provided multiple prongs: first, the team would allow the City Manager's Office to execute data analysis projects directly, without relying on staff in departments, and, secondly, this team would "undertake a review of all City departments and their existing analytic capability" (Reed, 2014). The proposal also called for the team to develop best practices that would guide departments that need improvements and support in the use of data analytics (Reed, 2014).

Significant work went into the development, refinement, shaping, and ultimate creation of the DAT so that it would be positioned to follow emerging best practices in a nascent sphere of public administration. Although previous work involving data analytics and data integration[1] had been conducted on a piecemeal basis in CSJ, the holistic integration of data analysis across the City's operations was as yet unachieved at DAT's inception. Indeed, the mandate creating DAT could easily be interpreted both as a call for independent analysis to be conducted of the data from the line departments and as reinforcement that the departments need support and guidance in developing best-practices for using data analytics to drive decision-making.

Although the initial groundwork to create DAT took the better part of a year, that groundwork included critical elements to ensure that the team would be able to be successful from the beginning. CSJ's DAT consists of two exempt policy, as opposed to technically expert, staff (1.0 Senior Executive Analyst and 1.0 Executive Analyst) who currently sit in the City Manager's Budget Office and nominally report to a Deputy Director of the Budget Office. Although they are not in technical positions, the unique work of DAT is substantively different than the day to day generalist work of other Analyst classifications throughout CSJ, with a greater emphasis on quantitative techniques (Miller, 2016). The City Manager's Budget Office in San Jose has long been a central repository for analysis throughout the City due to its unique location and role in the organization, and the inclusion of DAT into that office has some natural

---

[1] For the purposes of this paper, data integration will refer to breaking down barriers for access to information the organization is already collecting, but perhaps by a different department or arm of the organization, and data analytics will refer to using that same data to inform decisions and examining data to reach conclusions. These terms were developed by the City of San Jose DAT to establish a common nomenclature for discussing, and distinguishing, such projects (Miller, 2016). Furthermore, in this paper data analytics and data integration will refer to internal-facing processes and open data will refer to the external facing processes, although there is often significant overlap.

synergies. Those synergies and where DAT fits best into the structure of CSJ will be explored

throughout this comprehensive evaluation.

Parallel to the idea of data analytics and data integration is the idea of open data. At its

core, open data is the premise that (almost) any data set that a City collects and produces should

be disseminated to the public. This could include everything from a master address database to

employee salaries (which are public under California law) to code enforcement violations. The

idea behind open data is that not only could it facilitate greater data integration within the City, a

practice that is sometimes referred to as "dogfooding" since it is an internal consumption of a

product produced for external consumption in the same manner of dogfood manufacturers, but

that third-party interests could access that information and produce valuable insight from it.

Although the Mayor's March Budget Message for 2014-2015 also included a mandate for an

open data architect, the hiring for that position lagged significantly behind DAT's creation

(Miller, 2016). Instead, the creation of an open-data policy and other facets related to open data

included in the Mayor's March Budget Message for 2014-2015 have been adopted as goals by

DAT.

An important note regarding the creation of DAT is that it is not necessarily a new

concept. Although the methodologies, technologies, and strategies used by DAT to accomplish

its goal may be novel in the public sector, the point and purpose of DAT, at its core, is

fundamentally policy analysis and program evaluation. That idea will be explored more

thoroughly throughout this *ex post* comprehensive evaluation, but for purposes of framing the

discussion it should be noted that DAT is the logical extension of a formalized policy analysis

and program evaluation approach. Its premise is to isolate the signal from the noise amongst

large data sets in order to accurately and concisely interpret results and recommend actionable intelligence to policy makers.

Whether or not DAT can achieve that lofty goal has yet to be seen. Indeed, one of the most difficult aspects of government programs is evaluating those programs, and ensuring that "policy outputs" align with "policy impacts" (Nachmias, 1979). The policy outputs are the creation of the team as authorized by the Mayor's March Budget Message, but that output alone does not suffice to accomplish the goals of the program. The goals are what will constitute the policy impact. The goals, as described above, are to conduct independent data analysis projects, evaluate existing capacity in departments, develop best practices for use of data analysis, and institute an open data policy.

Now, almost a year after DAT started, CSJ is faced with overlapping information needs. The primary information need that CSJ has is evaluating whether or not the program worked in order to inform potential choices about the future of the program; this information need should be addressed by a formal *ex post* comprehensive evaluation of DAT's performance. In some regards, this touches on choice information since it appears to be "binary", a simple choice between "whether or not" DAT should continue (Heath and Heath, 2013). Accordingly, the lens of analysis should be widened to explore other alternatives in order to pursue a real choice. After such a choice is created and analyzed, the real crux of the evaluation becomes not simply whether DAT should continue, but whether DAT is effective in its current incarnation and/or should any aspects of its organization and structure be changed to increase its effectiveness? In order to address this implication arising from a widened lens, and options beyond simply continuing DAT in its current form or dismantling it, the secondary information need that CSJ has is exploratory: how can the program be made more effective?

An *ex post* comprehensive program evaluation of CSJ DAT's first year of existence can facilitate answers to these questions and inform CSJ's decisions about the future of DAT. The precise role of DAT as an evolution of a formalized role for policy analysis and program evaluation makes it an especially good candidate for a program evaluation. DAT itself should be evaluating programs and analyzing policy as one of its core functions and a program evaluation of DAT can therefore serve as a sort of 'meta' program evaluation, encompassing the program evaluation performed by DAT in order to understand its effectiveness as a central office. Assessing whether the "policy output" of creating DAT aligns with the "policy impact" of maximizing operational effectiveness is the core function of this program evaluation.

To accomplish that task, this evaluation will consist of both "process evaluation", assessing whether or not DAT was implemented in a manner consistent with its mandate, and of "impact evaluation", assessing whether or not the creation of DAT has moved the City in the intended direction (Nachmias, 1979). As such, it constitutes what could be considered a "comprehensive evaluation" which accomplishes both process evaluation and impact evaluation by employing the appropriate techniques for each (Freeman, 1970). These techniques will be explored more thoroughly in the literature review section in order to establish a sound methodology for the evaluation.

Literature Review

Although there is a significant body of work on the processes associated with conducting a program evaluation, including the various ways it can be done, there is relatively scant research on data analytics programs. This is primarily due to the emerging nature of data analytics programs in government entities. Some institutions like John Hopkins' have started initiatives such as the Center for Government Excellence (GovEx) to address this lack of documentation,

but such efforts are in their infancy. Therefore, this literature review consists of two elements. The first concerns itself with a defensible approach to conducting a program evaluation of CSJ DAT. One plank of such a program evaluation is a literature review itself, and the balance of this literature review is devoted to exploring the wider context of data analytics in municipal government. Although the information is largely exploratory and of a background nature it nonetheless offers important context for a comprehensive evaluation of CSJ DAT.

Methodology Literature Review:

In order to perform a true "comprehensive evaluation" of CSJ DAT, both the implementation and the performance of DAT should be evaluated. As discussed above, the "policy outputs" of the creation of the team and tracking against the mandate are important, but the "policy impact", or whether or not DAT is performing the way it was expected to and producing the desired results is the ultimate purpose of this comprehensive evaluation (Nachmias, 1979). Additionally since "programs are rarely so unidimensional that they warrant a single research methodology" a program evaluation is a good candidate for multiplism (Sylvia, Meier, and Gunn, 1985).

Synthesizing both an outcome evaluation and a process evaluation is warranted in this case due to the myriad facets of DAT that should be examined for a truly comprehensive evaluation, such as such as its implementation, its individual effectiveness, its ability to influence departmental outcomes, and its impact on the organization as a whole. Adopting and developing an outcome model will enable goals to be mapped against program elements and outcomes of DAT. Since DAT was created with clear goals, an emphasis on the latter two phases of process evaluation ("Implementation, and Feedback evaluation") of DAT could be used to extrapolate information about its performance (Sylvia, Meier, and Gunn, 1985). In outcome evaluation,

program outputs "are assessed against clearly defined program goals" but since only four goals, articulated above, were mandated that process is a relatively straightforward exercise (Sylvia, Meir and Gunn, 1985). For process evaluation, the evaluation should "use the goals to analyze the organization of agency activities, or to assess the allocation of organization resources among various activities, or to assist in the redesign of delivery systems in line with the goals" (Sylvia, Meier and Gunn, 1985). This is especially important as alternatives are considered for the continued implementation of DAT and makes process interview methodologies appropriate for this comprehensive evaluation.

Interviewing, survey research, case studies, and secondary data analysis are all valid techniques to be used to "reduce uncertainty for decision making in public settings" (Haas, Springer, 1998). However, the "problem should dictate the methods, not vice versa" (Patton and Sawicki, 1993). As such, borrowing from not only outcome evaluation and process evaluation, but also policy analysis and program evaluation is suited to this particular topic. For the most part, the stages of an evaluation model seem appropriate, but since an "evaluation model does not always prescribe the comparison of various policy or program alternatives" such a comparison needs to be factored into the analysis in order to widen the lens as discussed above (Haas and Springer, 1998).

Finally, due to the emerging nature of DAT, and of data analytics programs in general, a formalized and rigid quantitative analysis would not yield significant returns. Instead, qualitative methods, providing a study of selected issues in depth and in detail, are most appropriate. In order to resolve the exploratory information needs discussed above, "interviews with selected individuals…surveys of selected populations…on-site observation and site visits…analysis of existing data…[and] literature reviews" are all appropriate research methodologies (Haas and

Springer, 1998). This exploratory information will assist in satisfactorily addressing the ultimate information need of CSJ regarding DAT regarding how to proceed with DAT. The remainder of this literature review will turn its attention towards addressing that information need.

Review of Data Analytics Literature:

Municipalities across the country are collecting more and more information on a daily basis. What local governments do with that data, if anything, is emerging as a critical area for improvement. The previous efforts that cities conducted in recent years to make that data available with the stated end goal of "transparency", as a value in and of itself, are no longer sufficient (Nemani, 2013). The current calls for reform and improvement are primarily related to both gleaning actionable recommendations from the data itself and publishing the data in its rawest form rather than in a format that cannot be disaggregated. These new calls for improvements serve two purposes: (1), they allow the public and others to examine and analyze the data themselves, possibly through different methods than those used by the promulgating jurisdiction, and (2) they enable independent verification of the work that was already done. The other dimension of this work is that the same data could be used to improve internal operations. Neither of these paths to improvements is being implemented on a systematic basis across cities and so there is no existing best practice playbook for implementing them in cities.

Although the process of using large pieces of information, and running calculations and algorithms on that information, to reach conclusions and insight is not an entirely new process for government, its use in the public sector still lags significantly behind the private sector. Private industry has developed it to a far further extent, and has linked it to every facet of business development (Shueh, 2014). In fact, there are still many misconceptions around governments' use of analytics and their non-use of data analytics. Many governments "still

operate on rote bureaucratic procedures that don't use a lot of data mining or analytics to prioritize how government employees do their work and when they do what they do" (qtd. in Shueh, 2014). This is not necessarily unique to data analytics, but applies to many government programs. These bureaucratic procedures often do not include formalized evaluation, and policies/programs often continue merely out of consistency with past practices. This "status quo" bias evidences itself as organizational inertia that can be difficult to overcome, especially for a new program designed to change the organization's trajectory (Heath and Heath, 2013).

Formalized evaluative evidence is just one factor that policy-makers consider when they make decisions. It is contextualized, and often marginalized, against other factors including "political values and ideologies" as well as "vested interests" (Nachmias, 1970). However, it is becoming increasingly apparent that in the absence of formalized evaluative evidence, those other factors will certainly prevail. As stewards of the public good, government employees have a certain responsibility to lay out a clear analysis driven by the facts. The outcomes, and impacts, that result from the policy-makers' decisions may not always be supported by that analysis, but the analysis needs to be presented cohesively and coherently in order to be able to influence policy-makers appropriately. If it is presented in that manner, data analytics can fulfill one of the promises of policy research, namely that it "can improve the quality of public debate by making evidence a part of the decision calculus" (Haas and Springer, 1998). If this is done at an institutional level, significant benefits can be realized.

Some cities are pioneering the path to institutionalize open data, data analytics, and data integration. So far, analytics platforms have been "systems that send, receive and interpret information from all parts of an organization" (Shueh, 2014). These cities offer key insights into emerging smart practices in this space, and how data analytics and data integration can be

successfully implemented, and their efforts will be compared against CSJ DAT's efforts later in this paper. Although these programs are on the bleeding edge of innovation and are constantly shifting their approaches, a brief comparison of factors that affected their different approaches is critical to understanding what contributed to the success of the program.

Using data to make decisions is not a new concept, but the use of technology to more effectively leverage the collection of data has only begun in earnest in recent years. Line departments are collecting more data on a minute-by-minute basis, effectively representing a "fire hose" of information (Flowers, 2013). However, like any fire hose, in order for it to be effective it has to be handled properly and pointed at the right problem. Additionally, the flow to the "fire hose" is sometimes severely limited by the many filters and screens that are put in place by proprietary and legacy data systems in different departments. If those filters and silos were removed, there would be an even greater flow of information, which could be even more effective at reducing duplications of effort and increasing efficiencies from a process level. The clear flow of communication, and knowledge about who in the city is using what system, seems like a straightforward practice for local government that should already be in place.

Unfortunately, the reality is there can be significant breakdowns in communication across an organization of any size, but especially in a City with over 5,000 employees. For example, a recent audit of facilities maintenance by the San Jose City Auditor found that while both the Airport department and the Public Works department use the same database software for asset management, the Airport uses it far more effectively (Yotam & Hedges, 2014). However, until the audit was conducted Public Works had no knowledge that other CSJ departments were using the same database software as them, or that they were not maximizing the use of their database. Merely by realizing that there are other elements of the City using it, Public Works can now

reach out to Airport for training, development, and documentation of how the Airport uses the database.

Managing the flow of information and increasing the volume of data in order to enhance data analytics drives the idea of 'Big Data'. Although 'Big Data' means myriad things, it usually implies that looking through a larger data set can yield more significant results or introduce other factors as potentially correlative and thereby enhance the analysis. The oft-quoted example of this is the of google search terms to predict flu trends in 2009 through correlation rather than causation (Mayer-Schonberger & Cukier, 2014). This launched a movement that causation was secondary to correlation, which resonates with some of the trailblazers of data analytics (Flowers, 2014). From an analytical perspective, the "why" can be subordinated to the "what", provided that "what" can be sufficiently improved to realize operational efficiencies. This is heretical when compared to previous public policy work that demanded causation to be established in order to establish success. If all of the City of San José's information was collected in a single, machine-readable format the City could conceivably run countless algorithms and calculations in order to reach conclusions from the data. Some private companies even offer to run an analysis that will produce an algorithm that can be used for predictive modeling for any given scenario (Harris, 2014). The potential utility of analysis based on predictive modeling is difficult to overstate.

New York City's Data Analytics Team developed a predictive model to assess the risk threat for illegal conversions in different buildings throughout the City. The group combined building identification information with geocoding software, in order to reconcile the disparate location information used by various City Departments (Goldsmith & Crawford, 2014). They then imposed filters based on interviews with the subject matter experts; in this case building

inspectors who had their gut level reactions distilled to measurable items. They found that "two of the most powerful predictors of high-risk conversions are whether the building's property taxes have been paid…" and "whether a bank has foreclosed on the property" (Goldsmith & Crawford, 2014). They also found that recent permits for recent brickwork suggested that a building that might otherwise be considered at-risk was probably not at-risk since it suggested that the owner was more invested in the building. (Mayer-Schonberger & Cukier, 2013The hit-rate for inspections sky-rocketed, and vacate orders went from being issued in 13 percent of site visits to 70 percent (Mayer-Schonberger & Cukier, 2013). This represents a quintupling of vacate orders issued, and since fires in illegal conversions are "15 times more likely to result in injury or death for firefighters" "than other fires" this has a meaningful impact on not only residents but also New York City firefighters.

As useful as it is to have such predictive models, getting all of the information into a single format is a daunting task, and no City has yet done it, making the utility of doing so sound in theory but untested in practice for the public sector. Although lagging behind the private sector, the public sector has started to turn the corner and break down the walls and silos that have kept data in proprietary systems to understand the data and see how it can inform actionable insight.

Predictive modeling, information integration, sentiment and affect analysis, anomaly detection, and content analytics have all been identified as outcomes of data analytics programs in government with the highest priority being predictive modeling (Chiang, 2012). Others have pointed out that government and the public sector are now "primary motivators for big-data analytics", but that government data is largely in the semi-structured and real-time categories, which makes it difficult to conduct analysis on (Kambatla, 2013). Nevertheless, as uniform

standards for government data emerge, the scale for big data analytics to impact will increase. Until then, the "janitor work" or scraping together existing datasets into a larger aggregate and using filler information for the non-applicable portions is a major hurdle to adopting data-analytics programs (Lohr, 2014). Often times, between 50% and 80% of Data Analysts time is spent on collecting, organizing, and standardizing data (Lohr, 2014.) That means that only between 20% and 50% of their time is available for actual analysis. Conceivably, if their time could shift from the janitor work, it would shift to analysis. This also supports the approach that data analytics should be one prong of a larger information organization scheme in order to yield the most meaningful results.

However, the results that are gleaned from the analysis of scrubbed data might not always be significant or important. So far, research and the track record for big data "suggest gains of 1 percent to 5 percent from data driven decision-making" (Lohr, 2014). This track record is important to local government as it not only shapes the expectations that must be managed, but because in an era of heightened resource uncertainty, gains in efficiency of 1 percent to 5 percent can be tantamount to dozens of jobs being kept. It is critically important that decisions be made on accurate data instead of unsubstantiated anecdotes (PFPS, 2011). In order to reach those gains, data analytics programs have to be carefully constructed and pointed at the proper targets. "New data sources, new processing technologies, and new analysis" can "provide more and better decision support and analytics" but only if the data is used in analysis (Power, 2013). If garbage data is used as an input for the analysis, then garbage will be produced as an outcome of the analysis. If the data analysis component is not built into process work flows and incorporated into business as usual, the data can be discarded and the analysis becomes fruitless. If data

analysis is demanded by policy-makers as an input used to drive decisions, then it becomes an integral part of the process instead of something hastily tacked-on to the process.

With the emphasis on technology as the facilitator for the next wave of data analytics it is important to keep in mind that people are still the most important element. "The tools aren't as important as the insights they foster through analysis and discussion" (PFPS, 2011). Furthermore, "for this kind of analytic thinking, there exists no formula" (Behn, 2014). The analysis requires "creative people who are willing to muck around in the data, rummaging for patterns that others don't see, foraging for insights, dissecting apparent coincidences for something more than pure randomness" (Behn, 2014). This is not a traditional professional competency for bureaucrats in local government. Instead, it is an "obscure" skillset that needs to be developed, nurtured, and applied (Shueh, 2014). Unfortunately, since it is not a traditional role for a bureaucrat, it is also difficult to find leaders who can effectively speak to the importance of data analytics programs and run effective teams.

Another element that leadership must exhibit in order to make effective data analytics programs is the ability to change the minds of individuals who want to continue to own their data and resist sharing it with the rest of an organization. "In government" there are "a lot of barriers where people feel that they'll be undermined or lose power if they share their data" (qtd. in Shueh, 2014). The need to identify and articulate strategies to overcome that element and convince people it is more than just a flash-in-the-pan slapdash effort is paramount. Additionally, reassuring people that the point of a data analytics program is to increase the effective use of data, not to cut jobs, is one of the important tasks an organization must accomplish to be successful. A distillation of the data analytics paradigm is that it is not a mandate to do more with less, but rather a mandate to do more with the resources already being provided.

If data analytics, indeed any mandate, is not approached from a subordinate leadership model, wherein the organization asks the component participants how it can help, it can be met with skepticism and opposition. In December of 2010, Deputy Mayor of New York City Stephen Goldsmith announced that the Department of Sanitation would be streamlined by cutting four hundred jobs and demoting a hundred supervisors to line staff. This resulted in unsubstantiated "accusations of wildcat strikes and work slowdowns by sanitation workers" (Townsend, 2013). This ultimately caused parts of Queens to stay unplowed for days after a Christmas blizzard hit New York City. This lesson is instructive; it demonstrates what happens when a top-down, hierarchical approach is forced on to an organization without sufficient organizational buy-in, especially when it calls for "greater efficiency" even as it trims the head count.

Since there is no established guide, there is also no clear-cut way to organize a data analytics program. However, one thing that is clear is that "agencies don't have to start an analytics program on a grand scale" (PFPS, 2011). Even using in-house systems and Microsoft Office can grow programs as agencies and cities gain experience in performance management (PFPS, 2011). Additionally, disseminating the information back out to departments once it has been aggregated is as important as collecting the information in the first place. The goal should be to have as few barriers between the data and its actionable use as possible. This enables end users to do whatever they need to without imposing additional work. It creates capacity in end users instead of creating work for them to do.

Expanding on the points above related to incorporating data analysis into existing processes and projects is the idea that once the data has been curated, it should refresh automatically. This is the driver behind the idea of "extract, transform, load" wherein data is extracted from a legacy system, transformed into a common standard, and loaded into another

system, combined with other data, where it can be easily analyzed by end-users (Guarino, 2014). This is also related to the 'janitor work' mentioned above, and ensures that once the 'janitor work' is done once, it can be easily replicated without repeating the leg-work.

Finally, use cases and how they are defined are more important than technical solutions. Big Data is a tool, and like any tool it is not necessarily appropriate for every situation. Instead, it is more applicable to some instances than others and instead of "wasting valuable time and resources implementing a broad program with only marginal value added" a clear and actionable plan forward with a specific target should be in place (Govloop, 2014). Discrete problems and business requirements are critically important to successful data analytics programs everywhere, and especially in local government.

Managing for data analytics programs is "at the heart of an agency's ability to know how well it is performing and helps it determine what it can do better" (PFPS, 2012). "Managers need to integrate [data analytics] into how they manage, how they reward, how they build it into the culture of the organization" (PFPS, 2012). The actual operationalization of data analytics are beyond the capacity of any individual data analyst; in order for a data analytics team to be successful it needs strong, dedicated guidance and champions throughout the organization with executive support.

Literature Review Conclusion:

Although Data Analytics is a relatively nascent development in Public Administration there are lessons that have been learned from other cities and jurisdictions. Goals beyond transparency as an end goal in and of itself are necessary to inform decision-making. In order to be a part of the decision making, data analytics needs to be incorporated into the decision making process. Furthermore, a holistic effort to integrate systems through data integration can pay

dividends for data analytics, as seen in New York City's building inspection example. The management of expectations related to Data Analytics programs is also important, since gains in efficiency may be as low as 1%. These gains can be increased if the amount of time spent on "janitor work" is decreased, but requires an emphasis on standardizing data sets. Finally, the people are more important than the technology. The analysis requires a particular skill set able to work with technology, but also work collaboratively with departments and outside stakeholders in order to convey the 'value-add' of data integration and data analytics. Data programs can also start small and scale up, using in-house systems and Microsoft Office, in order to convey meaningful information back to decision-makers. Each of these elements offers context for a comprehensive evaluation of DAT in the City of San Jose.

As seen through this Literature Review, the need to examine both policy outputs and policy impact for a comprehensive evaluation has been established. A methodology based on an outcome evaluation and process evaluation is the best way to perform that analysis, with a special focus on the implementation and feedback evaluation of DAT in order to "assess the allocation of organization resources" in order to "assist in the redesign of delivery systems in line with the goals" (Syliva, Meier, and Gunn, 1985). A qualitative toolbox was explored, and an appropriate methodology will be discussed in the next section.

Methodology:

Qualitative assessment was the primary driver of original research for this paper. Ironically, comprehensive data to quantitatively analyze data analytics programs and their relative effectiveness does not yet exist. Instead, utilizing the techniques identified by Haas and Springer in the literature review above drove the research. The research questions necessary to perform a comprehensive evaluation are detailed below.

Research Questions:

1) What was the legislative mandate to implement DAT in San Jose?
    a. Was DAT implemented according to that legislative mandate?
    b. Was DAT implemented in a manner that will work for San Jose?
    c. Was DAT implemented according to emerging practices that have worked elsewhere?
    d. Could the implementation of DAT be refined in order to improve its success?
2) Is DAT achieving its goals?
    a. Is DAT conducting independent analysis?
    b. Is DAT analyzing existing capacity within Departments?
    c. Is DAT developing best practices for Data Analytics and integration of data into decision-making?
    d. Is DAT developing a policy and best practices related to Open Data?
3) Is DAT succeeding overall?
    a. What can be done to increase DAT's impact across the organization?
    b. Are there any other efforts being pursued by the City of San Jose that DAT can be included in to maximize its impact?

The first two research questions will be addressed through the findings and the third and final research question will be addressed through recommendations. In order to conduct a "comprehensive evaluation" that assesses both the implementation of DAT against its legislative intent and the "policy impact", beyond mere "policy outcomes", semi-structured interviews with selected individuals and analysis of existing data related to CSJ DAT were conducted. The policy outcome must be evaluated against the mandate that created it in order to assess the internal validity of the policy impact assessment. The implementation is therefore the *a priori* research need as demonstrated through the first research question.

It is also important to understand the context of the City of San José in order to make recommendations about improving DAT Simply copying an existing program from another City without adapting it to the nuances of the City of San José would not be useful. To that end, deciding what to improve is driven in part by what the City is already doing, and what it has the capacity to take on. Semi-structured interviews with staff from the City Auditor's Office, the

Information Technology Department, and DAT were used to shape the scope of what could be recommended based on what can be reasonably accomplished with available resources.

These interviews and analyses informed the creation of an outcome evaluation table, essentially an outgrowth of a logic model, related to DAT's long-term outcomes (or its long term "policy impact") that was mediated by program elements. The logic model is ideally suited to this aspect of an outcome evaluation, since it can capture theoretical expectations, the program goals as articulated by the interview subjects, program elements as arrived at by program staff, and proximate indicators/measures (outcomes). The measures for the proximate indicators are necessarily qualitative.

Table 1 – Outcome Evaluation Table of CSJ DAT

Legislative/Program Goals	Program Elements	Proximate Indicators/Measures
Conduct Independent Projects	Provide independent analysis through collaboration with departments on high priority projects	Departments operationalization of recommendations
Analyze Existing Capacity within Departments	Conduct data inventory to understand what data the City already collects and identify knowledge gaps	Map of existing data sources
Develop best practices for data analytics and integration of data into decision-making	Provide recommendations for improved use of existing data	Recommendations for improved use of existing data
Develop policy and best practices for Open Data	Development of Open Data Policy	Adoption of Open Data Policy
	Internal Data Champions Team	Departmental use of concepts discussed during Data Champions Meetings
	Conduct Service Delivery Evaluations	City Council adoption of recommendation of Service Delivery Evaluation Recommendation

The program elements identified in the outcome evaluation table were compared and contrasted against other cities and jurisdictions based on secondary document review, existing

data, and a purposive sample survey. The purposive sample survey was disseminated to jurisdictions that sent a representative to the Code for America summit, a civic innovation and civic technology conference in 2015, as well as participants in Bloomberg Philanthropies' "What Works Cities" initiative (discussed in greater detail in the latter portion of this comprehensive evaluation). Due to the low response rate associated with the survey, it can only be used for qualitative and descriptive purposes in conjunction with other existing data. Analysis of current published program plans and a review of their project charters all contributed 'smart practices' to benchmark against CSJ's DAT as discussed in the findings section.

This methodology ensures that a sound *ex post* comprehensive evaluation can be performed, analyzing not only a policy outcome evaluation but also a policy impact evaluation. Whether or not DAT is moving the needle on its stated goals to conduct independent data analytics projects, analyze existing department capacity, develop best practices in CSJ, and its adopted goal of developing an Open Data policy will all be addressed in the findings section. Recommendations based on the findings that emerged from the original research will be addressed in the conclusion section of this comprehensive evaluation, to widen the lens of analysis and enable options beyond the simple "whether or not" choice to continue DAT in its current incarnation.

Findings

Through the semi-structured interviews discussed in the methodology above and analysis of existing data an outcome evaluation table (Table 1) was developed. That table is displayed above, and moves from the Legislative and Program Goals of the program, through the Program Elements, to the Proximate Indicators/Measures (Outcomes). Each of these will discussed more thoroughly throughout the rest of the findings Section.

In order to conduct a comprehensive evaluation, both the implementation and the process of the program must be analyzed. To ensure the validity of the process evaluation portion, the *a priori* research task was to ensure that the policy outcome (the implementation) was consistent with legislative intent and the mandate that was originally issued for its creation. The first finding satisfies the first research question related to what the legislative mandate of DAT was, and whether or not it was implemented according to the legislative intent behind it.

Finding #1: The implementation of the Data Analytics Team (DAT) within the City of San Jose is consistent not only with the legislative intent that created it, but also with emerging   practices proven to be effective elsewhere.

The Mayors' June Budget Message for 2014-2015 that authorized the creation of the Data Analytics Program laid out the following mandate:

> "To increase the use of data analytics throughout City operations, as well as assess the current analytic capacity in individual departments, the City Manager is directed to allocate $195,000 ($209,000 in ongoing funding) to provide a data analytics team consisting of 1.0 Senior Executive Analyst and 1.0 Analyst to allow the City Manager's Office to: (1) execute data analysis projects directly, independent from City departments; and (2) undertake a review of all City departments and their existing analytic capability. Through the independent projects, this team will develop best practices that can be utilized in guiding those departments needing improvement and support in the use of data analytics." (Reed, 2014)

The actual policy impact of DAT will be discussed later in this findings section, but the implementation itself is fairly easy to evaluate against the textual mandate. Interpreting the legislative intent introduces another layer of complexity to the evaluation, but DAT's implementation still passes muster. Additionally, the implementation of DAT was predicated on emerging practices and its structure and placement within the organization are modeled on other successful programs.

The legislative intent behind the creation of DAT is evidenced in the goals that are set out, even the ones that are not explicitly stated as goals. "To increase the use of data analytics throughout City operations" suggests that there is insufficient use of data analytics currently and that the lack of data analytics should be remedied; this conclusion is further bolstered by the call for the development of best practices. Furthermore, the call to assess the current analytic capacity in individual departments could be interpreted as an indication that the departments vary significantly in their capacities and capabilities to conduct defensible data analytics projects. The establishment of ongoing funding indicates that this is not a pilot program, which is important for the sustainability and longevity of the effort. The ongoing funding is also significant because it warrants an extra level of scrutiny not only from an allocation of resources perspective, but because a pilot program would contain an inherent "tripwire" (Heath and Heath, 2013). Pilot programs are evaluated after an initial term, and kept, refined, or abandoned. Ongoing projects face no such scrutiny, making this *ex post* comprehensive evaluation all the more important.

One discrepancy between the text of the mandate and its implementation that should be identified is the distinction between an Analyst and an Executive Analyst. In the City of San Jose, the classification of "Analyst" is not exempt from the Fair Labor Standards Act (FLSA), whereas the "Executive Analyst" classification is exempt. The mandate called for 1.0 Senior Executive Analyst and 1.0 Analyst, but over the course of researching best practices and developing an implementation model, the City Manager's Office decided to request a reclassification from Analyst to Executive Analyst (Miller, 2016). This would enable the Executive Analyst to be driven by the tasks at hand rather than a set schedule, facilitating flexibility, and also add a distinguishing characteristic to the position. The City Manager's Office is the only department in CSJ with Executive Analysts, so the change in classification would

mark the incumbent as having the authority and backing of the City Manager's Office (Miller, 2016).

Additionally, the creation of a central team that would conduct a city-wide inventory of capabilities and develop best practices by leveraging existing departmental resources was predicated on past practice from other cities. These cities were alluded to in the literature review, but San Francisco, San Diego, Austin, Houston, New York, Chicago, Boston, Louisville Metro, and others have all implemented Data Analytics teams with varying degrees of success (Kicking Off a city Hall Analytics Shop, 2015). New York's program is the best publicized, and its team started off as a team of one, with Mike Flowers being the only staff member, that became a team of two and slowly scaled up. Mike Flower's first task was to take an inventory of all of the data that New York City was collecting and find a way to get it all into a single format (Goldsmith & Crawford, 2014).. He conceptualized this as a "data bridge" and posited that it would facilitate a greater understanding not only of the interconnections of the data, but of the correlations that would be indicated by apples to apples rendering of disparate data sets. Mike Flowers first hire was a quantitatively minded recent graduate, and Mike's job was to clear the administrative and bureaucratic road blocks that could otherwise prevent the analysis, the presentations of that analysis, and actions driven by that analysis (Goldsmith & Crawford, 2014). This represents efforts in both data integration (data bridge broke down silos) and data analytics (since they used their new large data sets to perform analysis).

In San Francisco, the Mayor's Chief Data Officer, Joy Bonaguro, has a staff consisting of one other full-time employee, and they take much the same approach as CSJ, but with a greater emphasis on open data and data integration (Miller, 2016). Their goals are closely aligned with those of San Jose, with their primary job being to raise the tide of data-use to drive decision

making throughout the City and County of San Francisco. However, San Francisco has end users in each department who are already fairly savvy regarding the use of data in decision-making, leaving their central office free to focus on the marginal utility of data integration efforts (Miller, 2016). Most other cities with data analytics programs also focus on data integration due to the natural synergy between the two elements, and CSJ therefore follows the emerging smart practice by similarly merging the functions (Kicking Off a City Hall Analytics Shop, 2015).

Another distinction that warrants further exploration is the placement of CSJ's DAT in the organization. In most of the examples mentioned above, the data analytics team is in the Mayor's Office, although it is sometimes in the Information Technology Department or a Department of Innovation. In those cities, that placement makes the most sense since they are governed by a strong-mayor system (Kicking Off a City Hall Analytics Shop, 2015). In a strong mayor system, the Mayor serves as the Chief Executive Officer and has significant sway over resource allocation. In the City of San Jose, the City is governed by a Council-Manager form of government, wherein the Mayor serves more as a first among equals among Council Members, and the Chief Executive Officer of the City is a council appointed City Manager (City of San Jose: Government, 2016). In a strong mayor system, department heads report to the Mayor. In a Council-Manager form of government, department heads report to the City Manager. This additional layer of bureaucracy has its benefits and its drawbacks, but it has a very important impact on the placement of DAT in CSJ. In order to align the role of DAT with similar teams in other cities, it makes the most sense to have it in the City Manager's Office since that is the Chief Executive Office.

The enabling legislation called for DAT's placement in the City Manager's Office, but this means that in order to work in conjunction with the Mayor's Office there is an additional

level of coordination that would not be necessary if DAT was in the Mayor's Office. Administrative overreach is an oft-cited concern related to the Mayor directly intervening or requesting work of CSJ staff, a theme that will be explored more in the policy impact portion of this evaluation. As a consequence of DAT's placement in the City Manager's Budget Office, the Mayor's Office has hired its own innovation staff, including a strategic innovations position and a Chief Innovation Officer that have both helped facilitate the success of the DAT from outside the City Manager's administration (Sammeta, 2016). Finally, the placement in the Budget Office is further complicated by the unique skill sets and expertise required for Data Analytic Projects. DAT has worked closely with the Information Technology department and, even in the absence of a formal reporting structure between DAT and IT, would not be able to accomplish its tasks without that collaboration.

The implementation of the City of San Jose's data analytics team is therefore not only consistent with the legislative mandate and intent that created it, but also aligns with the implementation of similar programs in other cities. Although there is not a single silver bullet, or packet of "best practices" that apply to all cities implementing data analytics, predicating the implementation of CSJ's DAT on methodologies that have been refined from other cities may enable DAT to learn from their pitfalls and successes. An informal network of practitioners from New York, San Francisco, and Chicago already communicate on an irregular basis, and through initiatives like Bloomberg Philanthropies' "What Works Cities" some of those communication channels are being formalized (Miller, 2016).

Finding #2: DAT has made progress towards each of its legislative and program goals in its first year.

As discussed above, the four legislative goals of DAT were to conduct independent projects, analyze existing capacity within departments, develop best practices for use of data analytics, and it adopted the goal of working on the Open Data policy. Interviews with the program manager, the Chief Information Officer, and analysis of various documents that DAT produced will be discussed in this section in order to provide answers to the second research question, addressing whether or not DAT is meeting its goals. As seen in the program outcome model, each of these goals was addressed through various program elements that DAT worked through during its first year. However each of these program elements has been limited by the assigned resources; there is a finite volume of work that can be accomplished by only two individuals. DAT has proactively worked to identify factors that could be employed as force multipliers, and was the main point of entry for a partnership with Bloomberg Philanthropies What Works Cities (WWC) initiative (Miller, 2016).

WWC is a nation-wide effort to improve the administration of local government in cities that are between 100,000 and 1,000,000 in population by partnering cities with experts in various areas. The Center for Government Excellence at Johns' Hopkins is the primary implementer, but they have outsourced different elements of "smart city" development to third party subject matter experts. Some of the expertise is Data Analytics, through Mike Flowers, who now works for the Center for Government Excellence, some is how to best frame messaging in order to "nudge" people to the desired outcome through the Behavioral Insights Team, and some is on Open Data Policy and implementation through Sunlight Foundation (Miller, 2016). Although there was initially trepidation on the part of the administration, the Mayor's Office was a strong advocate for the City of San Jose's participation, and helped work through some of the initial hurdles and roadblocks (Sammeta, 2016). This highlights some of the intricacies associated with Council-

Manager forms of government that are not present in strong-mayor systems. Partnering with

What Works Cities has significantly increased the bandwidth and capacity of DAT by providing

expert advice and consulting services that would otherwise not be available. Additionally, DAT

is currently exploring a partnership with a third party machine learning vendor, DataKind,

mediated through Microsoft, related to the Vision Zero project to analyze traffic fatality data will

be discussed more below (Miller, 2016). The synergy between the Mayor's Innovation Staff

(discussed above) in seeking out additional resources that can support DAT has been appreciated

by DAT since it also increases their capacity and helps develop relationships for future

partnerships with outside organizations (Miller, 2016). Thus, DAT has proven adept at using

outside resources in order to increase its impact on the organization, especially since it only has

two full time staff members.

The first goal of DAT was to conduct independent analysis. Although DAT has yet to

conduct a truly independent project, that outcome is by design and the legislative intent of

conducting analysis on behalf of the City Manager's Office has been accomplished (Miller,

2016). DAT's projects are in collaboration with departments, but its analysis is independent.

Through various projects that DAT has conducted it has worked collaboratively with

departments, essentially following the playbook for policy analysis and program evaluation, to

identify relevant questions and information needs that could conceivably be addressed by data. In

each of these projects, DAT has worked through iterations of the scoping and the analysis with

the departments to make sure that those information needs are being addressed and that

actionable recommendations that can be implemented in departments without creating

significantly more work are the final outcome (Miller, 2016).

DAT has worked on numerous projects that have crossed departments and stakeholders and has already made an impact through its analysis. DAT has worked with the Department of Transportation (DOT) on the Vision Zero Initiative, which seeks to completely eliminate traffic fatalities and significant collisions, and helped DOT interpret the Data it was already collecting (Miller, 2016). DAT also worked with the Code Enforcement division to ensure an even distribution of work amongst its code inspectors based on a new tiered inspection schedule. This was done through weighting the different inspection tiers and examining how many of each was in the previous workloads. The redistribution of the work will ensure that Code Inspectors are able to visit the most high-risk properties on a more frequent basis while still conducting the routine inspections of lower-risk properties (Miller, 2016).

DAT is also working with Code Enforcement to synthesize additional data sets originating from outside Code Enforcement, such as ones from the Fire Department, to establish regression models that predict the highest risk properties, even within the "high risk" tier. This effort is predicated on the similar effort that took place in New York City that increased the "hit rate" of significant violations found by code inspectors by triaging which properties they visited. This represents a perfect example of a project that marries the ideas of data integration (since it merges existing data sets that currently are disconnected from other) with data analytics. DAT also engaged with the cross-departmental unit assigned to illegal dumping. Through a hybrid data integration and data analytics approach, DAT merged independent data sets from partner departments in order to generate a comprehensive holistic "heat map" of sites that were especially impacted by illegal dumping. This enabled proactive enforcement in those areas, which limited the dumping (Miller, 2016). Through each of the examples above, it is clear that

the DAT has conducted independent analysis on projects in collaboration with departments, satisfying its first goal.

In each of these projects, DAT has first looked for examples of similar projects conducted by other cities to use as a model, and then adapted the approach for San Jose (Miller, 2016). This addresses the second element of the development of best practices, and DAT is developing a play book for potential data analytic techniques that could be used on different problem sets. Additionally, DAT has been institutionalizing their techniques. For example, through their collaborative work within DOT on Vision Zero to identify trends within the high intensity corridors, they have facilitated DOT's ability to continue that analysis on an ongoing basis without relying on more time from DAT. This approach is especially significant, since it ensures that DAT's efforts don't depend on ongoing maintenance from the central team (Miller, 2016). By decentralizing and spreading the knowledge of appropriate techniques to the City as a whole, the use of data to drive decision making becomes more common place in the organization, thereby improving the decision-making calculus of the policy-makers.

Another area of independent analysis that DAT has been instrumental in addressing is Alternative Service Delivery Evaluations (ASDEs). CSJ has a very particular process for evaluating which programs can be outsourced to private companies or achieved through a hybrid service delivery. The process mirrors a traditional program evaluation, and so it is a nearly perfect fit for DAT (Miller, 2016). Through an examination of what is currently being accomplished, the costs associated with that service delivery, and an exploration of alternatives, the ASDE process facilitates an informed recommendation by the administration to the City Council. In the past, these ASDEs were administered by the City Manager's Office, but were largely developed by the department that was currently delivering the service. This presented an

inherent conflict of interest, since the department had a vested interest in the status quo bias of continuing the current service delivery. The City Manager's Office usually added some additional points to the analysis, but it was largely predicated on the original report from the department. Unsurprisingly, few programs were completely outsourced. DAT has contributed to a revised ASDE for the administration of the City's Workers' Compensation program which was hybridized on a pilot basis in 2013-2014. Until DAT got involved, there was not much traction on a revised recommendation due to a logjam between the Human Resources' Department and the City Manager's Budget Office on various issues related to the interpretation of the data. DAT has helped move the process forward for a recommendation to City Council about the future of the Workers' Compensation program (Miller, 2016). The institutionalization of best practices regarding data integration and data analytics represents progress against DAT's second stated goal of developing best practices for those areas.

The role of independent arbiter is one that carries special import for DAT. One of the things that the team had to overcome was initial trepidation by departments. This was discussed in the Literature Review, and explicitly encountered by DAT in its first year. Since DAT is located in the City Manager's Budget Office, department heads expressed concern that DAT's analysis would be used to increase efficiency through reductions. DAT had to have similar conversations with a number of departments to establish their credibility and reiterate that their purpose was to enable Departments to do more with what they had, rather than make them do more with less (Miller, 2016). This trepidation may have been avoided entirely if DAT was located elsewhere in the City Manager's Office, as discussed more fully in the conclusions section of this document. Nonetheless, through its performance on the individual projects discussed above and its efforts with departments related to open data, DAT was able to overcome

the organizational inertia and change the mindset of departments to make them understand the value-add of collaboration with DAT.

Although the development of work related to open data was not initially assigned to DAT, it is the clearest example of a policy outcome that was successful, although its policy impact cannot yet be determined. DAT's work on open data started off as a "tertiary goal" but over the course of DAT's first year it became "increasingly integral" to the work of DAT not only for the value of the output (a Data Inventory that identified over 600 data sets that CSJ maintains) but for the conversations it enabled DAT to have with departments (Miller, 2016). Often times it was through the context of these conversations that DAT was able to convince departments of the utility of data integration and data analytics projects, and establish the collaborative nature of the relationship between DAT and the department (Miller, 2016).

Through their work on open data, DAT has finished a data inventory. Again, DAT predicated the data inventory on previous efforts from other cities (Miller, 2016). Other data analytics teams had identified the knowledge gap that surrounded the ability to know what data their city had access to in the first place, and the critical need to resolve that knowledge gap. Mike Flowers' data bridge efforts were discussed earlier, and DAT set out to similarly break down the walls between different siloed data sets in San Jose (Miller, 2016). Through interviews with department contacts, they identified hundreds of disparate data sets that the City collects and maintains. They then cross-referenced those sets against the departments' responsibilities in order to ensure that they were capturing everything, and worked with City Manager's Budget Office analysts to ensure there were no gaps (Miller, 2016). In the end, they identified 600 disparate data sets that are collected and maintained by various departments. The next work related to open data is a collaborative effort between DAT and the Mayor's Innovation staff to

identify the highest value data sets to be loaded into the City's open data portal. There are a

number of criteria that will be used to evaluate the different data sets, including the likely impact

of the data on public policy, public desire for the data, ability of the data to reduce public

records' act requests, and others (Miller, 2016). As these efforts continue, the data sets that are

identified as high value will be standardized and loaded into the City's open data portal

(data.sanjoseca.gov) which will enable individual departments, and the public at large, to

conduct analyses on the data (Sammeta, 2016). As part of evaluation of data sets, legal issues

surrounding the different data sets will also be explored in conjunction with the City Attorney's

Office to ensure that things are released in a legally defensible manner.

The release and publication of these data sets, and their use in routine City operations,

constitutes the policy impact of DAT's work on open data. However, they also performed a great

deal of the work on crafting the open data policy that was just recently adopted by the San Jose

City Council (Miller, 2016). The development of the policy provides the framework that

facilitates the release of that data (Sammeta, 2016). That framework ensures that there is a

collective vision going forward for how the data should be structured, released, and what it can

be used for. This means that the administration has clear sign posts for the release of data sets,

and will not need to return to the City Council every time it wants to post a data set.

Additionally, the policy is modeled on the most aggressive open-data policies of other cities,

which "default to open". This creates a clear default option for data sets: they will be published

unless there is a legal reason prohibiting their publication. Additionally, the policy ties in future

policy implications by mandating that new information systems should "have the ability to report

data in a manner consistent with the open data standards" (Sammeta, 2016). As discussed

extensively in the literature review section of this document, the output of generating an open

data policy does not guarantee the desired policy impact of departments actually defaulting to open data standards, but it does lay out the framework and the mandate to do so. A policy impact evaluation of the open data policy has important implications for the success of DAT, but is beyond the scope of this comprehensive evaluation.

The final goal that has not yet been addressed in this finding is the goal of "analyzing existing departmental capacity" to conduct data analytics projects. This is arguably the most nebulous of the goals, but nonetheless DAT has made significant progress towards not only achieving it but towards increasing departmental capacity. Through their interviews with every department about data sets, they also ascertained how departments were using the data they were already collecting, and made recommendations for how the use of that data could be integrated into departmental operations (Miller, 2016). Additionally, DAT has started an informal monthly meeting of "Data Champions" that is well attended by departmental staff from throughout the organization (Miller, 2016). This informal meeting frequently includes guest speakers and gives a specific department the opportunity to talk about their data analytics efforts. It represents a bottom-up approach (most of the attendees are junior staff) that nicely complements the top down approach of the legislative mandate. Furthermore, it creates a pipeline of individuals who may not ever work directly with DAT, but will incorporate best practices for data usage into their daily work (Miller, 2016). This opt-in approach has proven successful for raising the standards of data usage throughout the City, although its long term impact has yet to be seen. This monthly meeting also replicates in function, if not in form, San Francisco's Data Academy SF, which similarly seeks to disseminate knowledge about use cases for data integration and data analytics. If the 'Data Champions' choose to continue their careers with CSJ, the organization will surely

benefit from a data integration and data analytics perspective. DAT also has a page on CSJ's intranet site, where it posts best practices that it has developed.

Throughout its first year of existence, DAT has made progress towards each of its legislative and program goals. It has been significantly helped in these efforts by the use of outside partnerships, including What Works Cities, Microsoft, and others. It has worked on specific projects that cross departments, jurisdictions, and stakeholders, including the Vision Zero project, Illegal Dumping, and Service Delivery Evaluations of various programs including the Workers' Compensation Program. Through those independent projects it developed collaborative relationships with departments that have contributed to DATs success and facilitated the incorporation of DAT's analytic efforts into departmental work. DAT's work on open data has resulted in the adoption of the City's first ever Open Data policy, as well as an inventory of over 600 disparate data sets. This will be instrumental to continued data integration efforts, and the City's open data portal will serve as a public-facing version of New York City's Databridge. Finally, DAT has not only analyzed existing departmental capacity through its data inventory, but it has actually increased departmental capacity through publishing its best practices, interesting projects, and hosting an informal monthly meeting for interested staff members. The next section will talk about potential improvements that could be instituted in order to make DAT even more effective, which would in turn magnify its impact throughout the organization.

Recommendations:

Although DAT has made progress towards accomplishing each of its legislative goals in its first year, it could be modified for further operational effectiveness. The Mayor's March Budget Message for 2016-2017 includes some additional direction related to Data Analytics,

representing not only some additional options that were not available before but also additional

resources being allocated specifically for Data Analytics Projects. Additionally, the Mayor's

Innovation Staff has laid out a "Smart City" vision in March 2016 that institutionalizes some of

the efforts that DAT spearheaded. Each of these elements will be contextualized as part of

recommendations that could be implemented in order to increase DAT's effectiveness.

Recommendation #1: DAT should be moved from the City Manager's Budget Office to the City Manager's Office of Innovation and Digital Strategy *(Finding #1, Finding #2)*

The Mayor's March Budget Message for 2016-2017 mandates the creation of an Office

of Innovation and Digital Strategy (OIDS) in the City Manager's Office to implement the Smart

City vision laid out by the Mayor's Office. The form and structure of this Office is not specified,

but DAT is a natural fit for such an Office. Furthermore, the movement of DAT into OIDS

would signal the point that DAT has made since its inception, which is that its primary purpose is

not to cut the budgets of departments. Instead, tying it to Innovation and a larger digital strategy

would ensure that stakeholders associate it with a proactive approach rather than a preventative

one. Additionally, some of the elements of the smart city vision, such as "Integrate digital tools

into the daily work of the City to promote a transparent, data-drive culture of innovation and

continuous improvement" echo much of the work that DAT is already doing (Smart City Vision,

2016).

Aligning DAT with OIDS would also remove some of the bureaucratic layers between

stakeholders. Currently, DAT reports to a Deputy Director in the Budget Office, who reports to

the Deputy City Manager/Budget Director, who reports to the City Manager, who reports to the

Mayor. Although the structure of OIDS is not specified, if it is led by a Deputy City Manager,

that would remove one of the layers of reporting. Additionally, the ad hoc reporting of DAT to

IT, at least for advice and consulting, would be resolved. The Mayor's March Budget Message

for Fiscal Year 2016-2017 intimates that IT needs to be restructured from a line department, so it

is very conceivable that it would fall under the purview of OIDS as well. When this information

is coupled with the announced retirement of the IT Director in July 2016, moving IT into OIDS

seems like a very strategic move to ensure that executive leadership is in place (Sammeta, 2016).

This would also make DAT more flexible and agile in its collaboration with IT, and enable to

more quickly connect with the Mayor's Office since the chain of command would be reduced

somewhat. As a side note, unrelated to DAT, the Chief Innovation Officer position in the

Mayor's Office seems like it would be redundant alongside OIDS and it would be logical to

bring that position into the fold of OIDs as well. This structure would reflect work by other cities

that have similar Offices of Innovation that merge data integration efforts with data analytics

efforts, as well as other functions such as information technology and performance management.

There is significant overlap between those functions, and it makes sense to have them collocated

to streamline operations.

This recommendation explicitly addresses issues identified in Finding #1 and Finding #2.

In Finding #1, it was evidenced that the current reporting structure for DAT through an Assistant

Director, through a DCM, through the City Manager, to the Mayor's Office has caused

administrative and bureaucratic complexity that could be streamlined. Furthermore, the tension

that DAT has encountered through its positioning in the Budget Office could be avoided outright

in the future if DAT is moved into OIDS.

Recommendation #2: DAT should have positions added in order to enhance its service

delivery to the City of San Jose (*Finding #1, Finding #2)*

There is an old adage in government that you can always do more with more. DAT has made significant strides towards accomplishing its program goals in its first year, but it is limited by the incredibly small number of dedicated staff. DAT consists of 2.0 FTEs, while the entire City has more than 5,000 FTEs. This significantly limits the scope of work that DAT can work on at a given time. Although this has its benefits, since it has forced DAT to work collaboratively with departments and made them focus on projects that are of the absolute highest priority, more could be done if additional staffing resources were provided. The mandate to create OIDS represents an interesting nexus for DAT since it would presumably be the policy staff for that Office. This would increase its scope even beyond the original goals, and the adopted goal of open data, of DAT, to include all of the goals associated with the Smart Cities initiative. If that work was simply added to the existing workload of DAT, without additional resources, the ability of DAT to focus on its core priorities would suffer. Accordingly, it is recommended that DAT should receive 2.0 more FTEs at the Executive Analyst level in order to complete more data analytics projects, data integration projects, and the release of data sets on the City's open data portal. DAT has demonstrated that a lean model can make incremental progress, but it would be able to do even more if it had additional resources. Additionally, the inclusion of more positions would enable the development of succession planning for DAT. If the Senior Executive Analyst incumbent left that role, there would be more options to choose from for the ongoing leadership of the program or at least to continue operations while a new search was conducted.

DAT represents an opportunity for central policy analysis and program evaluation to be conducted independently by the City Manager's Office. Although their efforts are in collaboration with departments, their analysis is independent. As such they are uniquely suited to provide actionable information that does not have a hidden agenda to policy-makers; their job is

simply to summarize the data in a clear and concise manner. This skill benefits the City as a whole, and the addition of more positions in DAT would also enable a bench that could potentially move to roles, ideally in leadership positions, in departments. Some departments are already starting their own efforts to better incorporate data into their decision making. Parks, Recreation, and Neighborhood Services (PRNS) recently posted for a program manager to manage their business intelligence efforts. Through conversations with DAT, they realized that the staffing model they were working with did not provide the most effective analysis. If DAT had a deeper bench in place, one of their analysts would be perfectly suited to run a single department's data operations.

This recommendation addresses elements of Finding #1, that DAT currently only has 2.0 FTEs but has leveraged outside resources to do more impactful work. With more personnel, DAT's impact on the organization would be even greater, especially if it was partnered with similar services in OIDS. As seen in Finding #2, DAT has made progress towards its goals, but it has had to prioritize, and it could take on more projects faster if it was up-staffed. Additionally, growing the analytic capacity through informal and formal channels to build a bench would benefit the organization as a whole. A similar structure exists for the San Jose Budget Office, wherein a central office has analysts assigned to various departments. A central DAT with a comparable number of analysts would ensure a small enough portfolio for each analyst to deliver meaningful insights and analysis to their respective departments.

Recommendation #3: OIDS should be the central office for coordinating special technology projects and technology procurements *(Finding #2)*

As part of the Mayor's March Budget Message for 2016-2017, there are a handful of technology related proposals. Some of these are related to special technology projects, including

predictive analytic software based on social media for PD. Another project calls for $125,000 for a Master Address Database project. Although the Master Address Database project enumerates the marginal utility of such a database for first responders, that same database could also yield dividends for Code Enforcement. This is mentioned somewhat in the proposal as is the "reducing of wasteful city staff work" (Liccardo, 2016). This proposal is very revealing, since it indicates the promulgation and adoption of DAT's efforts into the fabric of the City as a whole and especially so in the Mayor's Office, which could be used as measure of DAT's success. The civilianization of the Bureau of Technical Services in the Police Department is also recommended as part of the Mayor's March Budget Message, and represents another facet that OIDS should be involved in.

Finding #2's information related to open data and the presence of over 600 different data sets underscores the current fractured nature of the City's operations. This recommendation would represent a strong step forward towards a cohesive and holistic approach to information management. As DAT continues to move through its evaluation of those data sets, it will undoubtedly find redundancies and overlap. It is imperative that CSJ does not backslide once that inventory has been completely mapped, and one of the methods to maintain that progress is to have all special technology projects, and technology procurements, flow through OIDS at some point. This would prevent duplication of effort, and enable better management of software that may already be in use across the City. This would solve the instance identified in the facilities audit, where two different departments were using the same database software but unaware that any other stakeholder in the City was using it. The operational efficiencies and gains that would be realized by this high level coordination are uncertain, but it could prevent the ad-hoc purchases of systems and solutions to address problems that the City has already solved.

Currently, all IT procurements must be approved by the IT department, but the incorporation of DAT (and OIDS) into that review would allow for an even higher level of strategy to be applied.

Each of these recommendations is interconnected and would yield the most significant returns if they were implemented together. However, each would also benefit the City if it was implemented independently. Adding more positions to DAT will enable it to accomplish more in a narrower timeframe, allowing CSJ to realize the gains of the program sooner rather than later. Moving DAT into OIDS will signal the administration's stated intent to use DAT to improve service operations, not cut budgets. Running all special technology projects, such as the development of a master address database or the civilianization of PD's Bureau of Technical Services, through OIDS will ensure that duplication of effort is minimized and that a coherent comprehensive strategy is applied across the board. Each of these recommendations would help DAT move closer to accomplishing its stated goals.

Conclusion:

Overall, DAT has been successful at achieving its legislative and program goals and is therefore a success from both a policy output and policy impact perspective. When the Mayor's March Budget Message for 2014-2015 mandated the creation of DAT, it had a relatively narrow construction. Through leveraging outside resources, in conjunction with the Mayor's Innovation staff, DAT was able to achieve far more in its first year than would have otherwise been possible with just two employees. By predicating projects on best practices that are emerging from comparable programs in other cities, DAT ensured that it did not stray too far into the deep end of the pool before it was ready. It has performed independent analysis in collaboration with departments, analyzed existing capacity of CSJ departments and made recommendations for incorporating the use of data into daily operations to each department, developed best practices

that it has disseminated through both an informal working group and its intranet page, and was instrumental in the research, writing, and ultimate adoption of the City of San Jose's first Open Data policy.

DAT had to overcome initial trepidation by internal stakeholders and convince departments that they were there to work collaboratively to improve service delivery instead of cut budgets. It was able to accomplish each of these by iterating through specific projects with stakeholders and institutionalizing the practices that they developed. This has ensured that the work DAT performed can be replicated even without DAT, which is incredibly important to the sustainability and longevity of data-driven decision making in CSJ. The more departments use data in their day to day operations, the more that data will be used to inform decision making.

For a comprehensive evaluation, both the policy output and the policy impact have to be analyzed and evaluated. The *a priori* research question of whether DAT was implemented according to its legislative mandate was relatively straightforward to assess given the narrow construction of the program. The reclassification from an analyst to an executive analyst was consistent with the intent of the program, and has facilitated greater prestige for DAT within the organization. The policy impact is evidenced not only by the projects that DAT has conducted, its work with What Works Cities, and the incorporation of Service Delivery Evaluations into its portfolio, but ultimately by the Mayor's March Budget Message for 2016-2017 and the mandate issued with the Smart Cities vision. The allocation of additional resources for data that will drive decision making is an important measure of the success of DAT's efforts to make data a part of the decision calculus.

DAT whet the appetite for data-driven decision making in CSJ. Both the Smart City vision and the Mayor's March Budget Message for 2016-2017 call for widened mandates that

will amplify the work that DAT has started. The creation of an Office of Innovation and Digital Strategy within the City Manager's Office is a logical iteration from DAT's work. Accordingly, DAT should be incorporated into OIDS organizational structure, where it would likely be the linchpin of that office. Doing so would reduce the administrative burden on DAT, and make it more agile for responding to requests and initiatives from the Mayor's Office by stripping out a layer of bureaucratic reporting. As OIDS supports a new committee tasked with implementing the Smart City Vision, the tide of data-driven decision making across the organization is likely to rise. Departmental use of data, even outside of DAT, is increasing, as evidenced by the addition of the PRNS data manager position. Ideally, each department would have its own internal expert on the use of data for decision making, which DAT could coach and develop centrally. A similar structure exists for administration of the departmental budgets, with a designated departmental contact and coordination and central administration in the City Manager's Budget Office. An iterative comprehensive evaluation of OIDS, from both a policy output and a policy impact perspective, would be an excellent process to revisit, especially since its formal structure is not specified in its mandate.

Recommendations to make DAT more effective include provisioning additional resources dedicated to DAT. Two more FTEs would enable DAT to take on more work sooner, develop a bench for future leadership within DAT and in departments, and would signal that DAT is more than just a token effort. Running all special technology projects and technology procurements not only through IT, but through OIDS, would enable a greater depth of strategy to be applied and ensure that duplication of effort was avoided. Redundant systems and solutions would not need to be purchased and departments would be able to leverage expertise of other departments more effectively with regards to specific systems.

The Mayor's Smart City vision calls for San Jose to be "the World's Most Innovative City by 2020" and DAT represents an integral part of that plan (Smart City Vision, 2016). Its policy impact has already paid dividends and allowed for data driven decision making with regards to illegal dumping, traffic death mitigation efforts, code enforcement workloads, and code enforcement risk assessments. As DAT continues, it will likely continue to strengthen the use of data to drive decisions in CSJ. Ultimately, DAT is a small central team that is trying to effect organizational change. As such, its best hope of accomplishing that organizational change is continuing its collaborative efforts with departments, including the informal 'Data Champions' working group. However, like any other element focused on policy analysis and program evaluation, DAT merely provides one element to policy –makers: actionable recommendations based on analysis of data. What policymakers do with that data and the recommendation is at least partially dependent on other factors. Regardless, DAT represents a great stride forward for formalizing policy analysis and program evaluation through data analytics. DAT utilizes a host of models and techniques to inform its analysis, but it is essentially just the next logical step for using information, instead of anecdotes, to drive analysis and decisions.

As other cities realize the potential applications for data analytics programs to incorporate data into business operations and decision-making it is likely that formal best practices will develop and emerge. Even though these best practices will not be a panacea that can be applied in every situation, they will nonetheless provide a starting point. DAT has contributed to those smart practices by iterating projects that other cities already performed, including a comprehensive map of existing data sources in the city. Integrating this data from disparate systems presents an opportunity for further analysis and may reveal previously unrealized correlations that could further enhance service delivery. Although DAT has made progress

towards achieving its goals in its first year, there is still plenty of work to be done in CSJ.
Moving DAT from the City Manager's Budget Office into OIDS represents an important
development that will facilitate DAT's continued success and an even greater policy impact over
the course of the program, and would warrant additional study to understand what other factors
could be improved in order to maximize DAT's operational improvements to CSJ.

Works Cited

Behn, Robert. 2014. The PerformanceStat Potential: A Leadership Strategy for Producing Results. Brookings Institution Press.

Chen, H., Chiang, R., Storey, V. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly. (36:4), 1165-1188

City of San Jose: Government. 2016. Accessed online at https://www.sanjoseca.gov/index.aspx?NID=105 on 4/2/16

Freeman, Howard E., Evaluation Research and Public Policies in Gene M. Lyons, ed., Social Research and Public Policies. The public Affairs Center, Dartmouth College (Hanover, N.H.: University Press of New England, 1975), p. 146

Flowers, Mike. 2013. Beyond Transparency: Beyond Open-Data: The Data Driven City. Code For America Press.

Goldsmith, S, Crawford, S. 2014. The Responsive City: Engaging Communities Through Data-Smart Governance. John Wiley and Sons.

Govloop. 2014. The Govloop Guide to Big Data: Examining the Big Data Frontier

Guarino, Dave. 2014. ETL For America. http://daguar.github.io/2014/03/17/etl-for-america/ Accessed on 11/3/2014

Hammack, Steve. 2014. Parks Division Business Intelligence Status Report. http://www.sanjoseca.gov/DocumentCenter/View/28743 Accessed online on 10/2/2014

Haas, Peter and Springer, Fred. 1998. Applied Policy Research: Concepts and Cases. Garland Publishing.

Harris, Derrick. 2014. This startup says it can find the algorithm that defines your data. https://gigaom.com/2014/04/09/this-startup-says-it-can-find-the-algorithm-that-defines-your-data/ Accessed online on 10/5/2014

Heath, Chip and Heath, Dan. 2013. Decisive: How to Make Better Choices in Life and Work. Crown Business.

Kambatla, K, Kollias, G., Kumar, V., Grama, A. 2013. Trends in Big Data Analytics. *Journal of Parallel and Distributed Computing.* (74), 2561-2573.

Works Cited (continued)

Kicking off a City Hall Analytics Shop. 2015. Code For America Summit Breakout Session. Accessed online at https://www.youtube.com/watch?v=IbmFYtvKhNs on 12/20/2015

Liccardo, Sam. 2016. Mayor's March Budget Message for Fiscal Year 2016-2017. Accessed online at http://www.sanjoseca.gov/DocumentCenter/View/55150 on March 20, 2016

Lohr, S. 2014. For Big-Data Scientists, 'Janitor Work' is Key Hurdle to Insights. *New York Times.* :http://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html Accessed online on 10/2/2014

Lohr, S. 2014. Making Big-Data Think Bigger. *New York Times*. Accessed online at: http://bits.blogs.nytimes.com/2014/09/20/making-big-data-think-bigger/ Accessed online on 10/2/2014

Mayer-Schonberger, Viktor, Cukier, Kenneth. 2013. Big Data. Houghton-Mifflin.

Miller, Kevin. 2016. Semi-structured interview conducted on 4/8/2016

Nachmias, David. 1979. Public Policy Evaluation: Approaches and Methods. St. Martin's Press.

Nemani, Abhi. 2013. Beyond Transparency: Afterword: What's Next? Code For America Press.

Partnership for Public Service. 2011. From Data to Decisions: The Power of Analytics.

Partnership for Public Service. 2012. From Data to Decisions II: Building an Analytics Culture.

Patton, Carl V., and David S. Sawicki. 1993. Policy Analysis and Planning: Theory and Practice. Prentice-Hall.

Power, Daniel. (2013). Using 'Big Data' for analytics and decision support. Journal of Decision Systems. (23:2), 222-228.

Reed, Chuck. 2014.Mayor's June Budget Message for 2014-2015. http://sanjoseca.gov/DocumentCenter/View/37382. Accessed online on 10/2/2014

Rosenberg, Marc. 2014. San José Fire 911 Response Times Getting Worse. http://www.mercurynews.com/crime-courts/ci_25624061/san-jose-fire-911-response-times-getting-worse Accessed online on 10/2/2014

Sammeta, Vijay. 2016. Semi-Structured Interview conducted on 4/8/2016

Works Cited (continued)

Sammeta, Vijay. 2016. Open Data Policy.
http://sanjose.granicus.com/MetaViewer.php?view_id=&event_id=2130&meta_id=5651
40 Accessed online on 4/7/2016

Shueh, Jason. 2014.   "Big Data Could Bring Governments Big Benefits".
http://www.govtech.com/data/Big-Data-Could-Bring-Governments-Big-Benefits.html
Accessed online on 10/2/2014

Smart Cities Vision. https://www.sanjoseca.gov/index.aspx?NID=5001 Accessed online on
4/7/2016

Townsend, Anthony. 2013. Smart Cities: Big Data, Civic Hackers, and the Quest for a New
Utopia. W.W. Norton and Company.

Yotam, Avi and Hedges, Cheryl. 2014. Facilities Maintenance Audit: Process Improvements are
Possible, but a Large Deferred Maintenance Backlog Remains. Accessed online at
http://sanjoseca.gov/DocumentCenter/View/37404 on 1/30/2015