

A Context-Driven Data Visualization Engine for Improved Citizen Service and Government Performance

Un outil de visualisation de données guidé par le contexte pour améliorer les services aux citoyens et les performances de gouvernance

Feras Batarseh¹, Jay Gendron², Rafael Laufer³, Mythili Madhavaram⁴, Abhinav Kumar⁴

¹ College of Science, George Mason University (GMU), USA, fbatarse@gmu.edu

² The United Services Automobile Association (USAA), USA, gerald.gendron@gmail.com

³ Systems Planning and Analysis, USA, rafi.laufger@gmail.com

⁴ Volgenau School of Engineering, George Mason University (GMU), USA, mmadhava@gmu.edu/akumar19@gmu.edu

RÉSUMÉ. Chaque jour, le gouvernement génère et consomme de grandes quantités de données. Les agences fédérales sont maintenant à la pointe de l'analyse du *Big Data* pour résoudre des problèmes tels que l'amélioration du service aux citoyens. La tranquillité domestique, la satisfaction de la clientèle, la transparence et la fourniture de services de qualité comptent parmi les nombreux objectifs critiques que la plupart des agences gouvernementales visent à atteindre grâce au service aux citoyens. Les gouvernements fédéral, d'état, des comtés et des villes sont constamment confrontés à un nombre impressionnant de demandes de citoyens, d'organisations, de chercheurs, de médias et de nombreuses autres entités. Les gouvernements sont toutefois constamment confrontés à des problèmes de responsabilité qui peuvent remettre en question leur rôle et leur efficacité. Cet article, grâce à la science des données et du contexte, introduit un nouveau moteur qui pourrait être utilisé par toute agence fédérale dans le but d'améliorer le service aux citoyens, d'évaluer les paramètres de performance et de fournir des indications pour améliorer les taux de satisfaction. Le modèle utilisé dans ce moteur (appelé iGPS) est déployé à travers un certain nombre de jeux de données de taille réelle. Les visualisations de données significatives et exploitables sont présentées aux employés fédéraux en fonction de leur contexte pour les aider dans leur prise de décision. Le modèle a été testé pour la première fois (à travers six ensembles de données gouvernementales provenant de 46 agences), des modèles d'apprentissage automatique sont développés, des visualisations sont réalisées, le système est déployé au gouvernement américain et les résultats expérimentaux sont enregistrés et présentés.

ABSTRACT. Every day, the US government creates and consumes significant amounts of data. Federal agencies are finally riding the wave of Big Data Analytics for solving problems such as improving citizen service. Domestic tranquility, customer satisfaction, transparency, and providing quality service are among the many critical goals that most government agencies aim to achieve through citizen service. Federal, State, County, and City governments are constantly challenged by the overwhelming number of service requests from citizens, organizations, scholars, the media, and many other entities. Governments, however, are constantly facing accountability issues that can call into question their role and efficiency. This paper, through the use of *data science* and *context*, introduces a novel engine that could be used at any federal agency to improve citizen service, evaluate performance metrics, and provide pointers to enhance satisfaction rates. The model used in the engine (called iGPS) is deployed through a number of real-world citizen's complaints datasets. Insightful and actionable data visualizations are introduced to federal employees depending on their context. - The goal is to aid them in decision-making. The model is tested for the first time (through six governmental datasets from 46 agencies), machine learning models are developed, visualizations are built, the system is deployed at the US government, and experimental results are recorded and presented.

MOTS-CLÉS. Visualisation de données, Contexte, Gouvernement, Agence Fédérale, Service au citoyen.

KEYWORDS. Data Visualizations, Context, Government, Federal Agency, Citizen Service.

1. Introduction and Motivation

America's most formidable political challenge to providing effective citizen service is often at the core of its history. A cultural emphasis on individualism, the Founding Fathers' inherent suspicion of centralized power (by providing more power to states), and an incorrect implementation of principles

of a federal republic has resulted in a fragmented political structure. According to the American Customer Satisfaction Index (ACSI), American citizens are generally not content with citizen services at the federal government [1]. A decline in satisfaction rates has also corresponded with a general decline in trust in government accountability. The public's trust in government service has declined to 30% according to a study by PEW Research [1].

Compared to other nations, America's style of governance is highly decentralized [2] (i.e. Federal). Currently, there are more than 87,000 governments in the US on all levels (state, county, municipality, township, school, and special district). Relationships among different levels of government are usually not authoritative, and implementations of laws often involve a great deal of bargaining and compromise. Even "imposed mandates" by higher governments often have different local governmental interpretations. America's decentralized system of power, however, has many benefits: reduced risk of abuse by a central government, increased incentive for jurisdictions to compete, and greater opportunities for jurisdictions and states to take innovative initiatives without lengthy federal approval. Because of this political culture, Americans tend to place more trust in local governments than in the federal government in Washington, DC. However, this structure also adds to the complexity of administrative actions, adds to transaction costs, and frequently results in gridlocks.

Citizens often fail to recognize services that the government provides, or assume that the government performs functions that it doesn't. Several reform efforts, such as the Civil Service Reform Act of 1978 [3], did not achieve the desired results. These perceptions also directly relate to citizen satisfaction rates regarding government services. If the government is to function like a company, then measuring these rates is essential. Improving employee performance (to eventually increase citizen satisfaction) through performance-related pay was a key goal of the Civil Service Reform Act – especially for employees on the frontlines of citizen services. However, the implementation of the Act is not ubiquitous, with each agency, department, and level of government having their own interpretation. For instance, the State of Virginia tried to link incentives for a pay plan, but the general assembly did not pass the required funds.

State governments often have difficulty in identifying an objective performance rating metric, or offer too little in terms of bonuses to incentivize performance. The National Commission on Public Service (Volcker Commission) [4], made the first serious attempts to investigate issues with public service in 1989. The study concluded that, in a nutshell, the government is falling behind at all levels. Public service has not been able to retain the employees needed for effective operation of services. In addition, the study reported low morale among senior managers and career professionals. The Volcker Commission also raised concerns regarding the inability of the government to retain scientists and engineers, and the lack of good-quality applicants in the selection process. Fast forward to today, and the need for qualified data scientists and engineers in government is more crucial than ever before [4]. The increased collection of data within government, however, has its own side-effects. Officials across a range of agencies have echoed the conclusion that they have too many datasets, with little insight. This has led to incidents of flawed analysis. Independent investigations on cases such as the "underwear bomber" admit that the "crushing volume" of information is one of the factors for "hampered accurate analysis". The new skill requirements for government employees will involve locating, recognizing, understanding, and creating new systems with data. At the managerial level, choices about the application of technology to government will have to be made. With these facts in mind, it is evident that the government requires a concrete reform of its citizen service sector [5].

Legislations that are intended to bring more transparency to citizen service include the Government Performance and Results Act of 1993 (GPRA) [6] and the National Performance Review. Both of these attempted to enhance "performance-based decision-making" and provide greater accountability for government-funded programs and agencies [7]. The challenges to implementing performance metrics are partly "technical", but also organizational and political. The main technical difficulty in implementing performance metrics is a Big Data management and measurement process [8][9]. For

example, the Workforce Innovation and Opportunities Act (WIOA) and the Bureau of Labor Statistics (BLS) both measure employment rates [10], but use different definitions of employment. BLS count unpaid family workers as employed, while the WIOA only counts individuals with wage records as employed [10].

Some common organizational problems include different informational needs in every agency, ambiguous mission objectives, and “multiple and contradictory agency goals”. Agencies such as the Environmental Protection Agency (EPA), for example, report to over “seventy different committees and subcommittees” with differing goals and interests [11]. There are also political complications. Congressional races are focused on short-term re-elections. As a result, initiatives for updated performance metrics and data analytics are not interesting to voters (compared to other legislations). A study by Jones and MaCaffery [12] has found that many Congressional leaders have “little knowledge and interest” in the long-term implementation of performance measurement initiatives. As a result, congressional interest in improving performance metric techniques in government is also likely to be low [13].

Despite technical, organizational, and political challenges, data analytics can be used as a tool for government for effective citizen service. Data analytics can assist agencies by collectively putting scattered data on a unified platform. Examples have already starting to emerge. Medicaid programs now have state and local agencies sharing data to look at healthcare trends and outcomes. Another example is the Smart Policing program, implemented in 38 different American police departments, which funds and empowers data-focused crime prevention tactics. Analysis incorporates information from GIS mapping and predictive analytics to search for “trends in criminal activity” and to determine its root causes [14]. For example, from 2012 to 2013, Big Data Analytics helped the Philadelphia Police Department to identify at-risk areas for crime and send more regular patrols to those places. The result was a “39% reduction in home burglaries”.

Another example of the successful use of data analytics is the Office of Management and Budget (OMB) [15], which cited a drop in government-wide improper payment rates due to the use of analytics and improved government oversight. Other examples include the USDA, which successfully reduced the rate of food-stamp trafficking from around 2.5 percent of food stamp benefits to about one percent through Big Data analytics. However, there is still a major lack of a standardized system that can consume, organize, and visualize Big Data in the government (especially within different contexts).

The model presented in this paper (Intelligent Governmental Performance and Services: iGPS) is a viable *context-aware* solution for government agencies to implement data analytics, improve customer and complaint services, increase citizen satisfaction, and extract pointers for enhancing performance metrics. Citizen Service is critical to citizen satisfaction, however examples from court requests and citizen complaints are very relevant to the method presented in this paper. When a complaint comes in, federal analysts have no direct means to quickly access the right data and find insights that would aid the citizen. The method presented in this paper aims to do so with the use of context as it is extracted from data.

This paper is structured as follows: the next section introduces related work and existing methods in the government, and section 3 presents the main contribution of this paper: the *context-driven iGPS*. Experimental setup, datasets, visualizations and results are given in section 4, and finally, conclusions and future work are presented in section 5.

2. Related Work and Initiatives Over the Past Century

2.1. *The Evolving Confluence of Big Data and Performance Management in Government*

Advances in data science and computational power are helping to bring to fruition the visions of 20th-century thought leaders and helping the government to achieve performance management goals more fully.

2.1.1. *The Early Years – Industrial Statistics for the Second Industrial Revolution*

The period between 1870 and World War I is often called the Second Industrial Revolution. It differed from the first, which occurred some 100 years earlier, by changing the relationship between the knowledge of nature and its effects on technological change [16]. This heralded a new way of thinking about progress and organizational operations. As the industrialized nations emerged at the turn of the 19th century, wise and pragmatic statisticians working in and alongside government sought to apply mathematics as an influencer of technological change. Sir R. A. Fischer developed techniques for Analysis of Variance and the Design of Experiments in the mid-1930s while working on data from agricultural experiments at the Rothamsted Experimental Station in England. He left perhaps one of the greatest legacies to data science from a mathematical perspective – he connected statistics to solving governmental problems [17].

Meanwhile, in that same decade, W. Edward Deming was working in the United States on statistics for industry. He studied under Fischer and later, while working in agriculture, he brought Shewhart's work of statistical process control into manufacturing, management, and enterprise processes. His contributions improved industrial-level measurement and decision-making, which drove change on production floors in near-real time. Deming notes that “most of us have thought of the statistician's work as that of measuring and predicting and planning, but few of us have thought it the statistician's duty to try to bring about changes in the things that he measures” [18].

2.1.2. *The 1960s and 1970s – Statistical Computing and Visualization Comes of Age*

As computational devices emerged from companies like IBM, academia and then industry began developing programs to automate statistical computation. Among the first leaders in this emerging space were SAS and SPSS. By the 1970s, work produced by Tukey [19] introduced the notion of exploratory data analysis and changed how scientists looked at their data. His work occurred around the same time as the release of the S statistical programming language by Bell Labs in 1976 (the precursor to the R statistical programming language). Together, exploratory and visual analysis became more automated and harnessed the work of early leaders, such as Sir Fischer, to find trends based on a scientific approach.

2.1.3. *The 1980s and 1990s – an Era of Total Quality*

After World War II, Deming's methods were adopted in Japan as they rebuilt their country and their economy. This led to high levels of economic growth for Japan through the 1980s. Interestingly, American industry did not adopt Deming's approaches like their Japanese competitors – that is, until the United States realized that they were falling behind in terms of manufacturing quality and product sales. Total quality management (TQM) emerged after corporate America took the rise of Japan's economy seriously. According to Bemowski [20], the US Navy coined the term TQM and focused on statistical process control to leverage data, improve quality, and assess process performance. A key contribution of TQM was bringing quality and statistical process control into government.

Each of the military services within the Department of Defense (DoD) later adopted some form of TQM, and this later spread into other government organizations. In 1987, the US Government, under the Department of Commerce, established the Baldrige National Quality Program [21]. It recognized the first cohort of recipients in 1988, which included Motorola (manufacturing), Westinghouse (manufacturing), and Globe Metallurgical Inc. (small business) [21]. These three companies exemplified the quality movement that was gaining momentum in the US and met the rigorous requirements of the Baldrige Award. Since then, the US President has awarded this honor annually. As quality programs became part of corporate culture through the end of the last century, the Department of Commerce changed the name of the award to the Baldrige Performance Excellence Program – recognizing that progress had been made in quality and thus shifting the focus to performance excellence [22].

Figure 1 highlights seven categories, as shown by the hexagons and the foundational category “Measurement, Analysis, and Knowledge Management”. Moreover, the framework highlights key topics that appear in this paper; namely, providing **Results** to benefit **Customers**. Those results come from **Leadership** by blending **Operations** and the **Workforce**, and are manifest through **Measurement** and **Analysis**. In accordance with the Baldrige Program, the GPRA act of 1993 (discussed previously) was created to achieve six specific purposes.



Figure 1. The Baldrige Government Excellence Framework [21]

These purposes are summarized in Table 1, which expresses the objectives and approach for each one. Because of the GPRA, Agencies provide the Congress with five-year performance plans; annual performance plans establishing goals, measures, and initiatives; and reports to assess success in meeting targets. Beginning in 1996, the commercial sector and academia added to this growing body of knowledge in performance management by writing consultant-style guidance in the form of books and articles. The Balanced Scorecard approach emerged from work in the early 1990s to measure organizational performance [23]. This approach was highly acclaimed and captured the essential elements of GPRA, even though it was developed for industry use. Niven [24] later adapted the balanced scorecard framework specifically for government and nonprofit agencies. In addition to revising the underlying framework, Niven provides rules of thumb to keep the number of performance measures at a manageable level. Otherwise, agencies run the risk of having so many measures that the results are uninterpretable.

Achieve these objectives	Using these approaches
Improve confidence in capability of government	Hold Agencies accountable for reaching goals
Initiate program performance reform	Set goals, measure performance, and report progress publicly
Improve program effectiveness and public accountability	Focus on results, service quality, and customer satisfaction
Help managers improve service delivery	Require plans to meet objectives; provide information about results and service quality
Improve decision-making	Provide objective information on: a) achieving objectives, and b) effectiveness and efficiency of programs and spending
Improve internal management of the Federal Government	<i>No approach specified in GPRA</i>

Table 1. *Purposes of GPRA: objectives and approaches, excerpted from the Act*

2.1.4. The 2000s – Transparency and Technology in Governance

The 2000s marked a resurgence in machine learning which began in the 1970s but was limited in utility because of computational capabilities available at that time. By 2009, a buzz of activity was emerging from both industry and government to harness data science and citizen science.

- Microsoft released a freely available data science eBook (under the creative commons license 3.0) entitled *The Fourth Paradigm*, which contains essays inspired by the prophetic insights of researcher Jim Gray and his vision of a paradigm based on data-intensive science [25].
- The Open Government Initiative, also called the Open Government Directive, was released by the White House. It lays out the three principles of transparency, participation, and collaboration as the centerpiece of open government. Together, they promote accountability, enable idea sharing from citizens, and improve government effectiveness [26]. Getting citizens involved in the work of government helps it to run at a higher level of effectiveness.
- The US President appointed the first US Chief Technology Officer, Aneesh Chopra, as well as the first US Chief Performance Officer, Jeffrey Zients. Creating two new positions on Mr. Obama’s management team helped to “promote technological innovation to help achieve our most urgent priorities” and “streamline processes, cut costs, and find best practices throughout our government”. These appointees would work closely with the US Chief Information Officer, Vivek Kundra. According to Mr. Obama, together they were “responsible for setting technology policy across the government, and using technology to improve security, ensure transparency, and lower costs. The goal is to give all Americans a voice in their government and ensure that they know exactly how we’re spending their money – and can hold us accountable for the results” [27].

The first decade of the 21st century was marked by a shift. Government had become a partner with industry and academia to enlighten the decade to follow – arguably the most significant decade for data science.

2.1.5. The 2010s – Data Science Responds to the Growth in Government Data

In 2011, Davenport and Patil's landmark article, entitled "Data Scientist: The Sexiest Job of the 21st Century", appeared in the *Harvard Business Review* [28]. It captured the hearts and imaginations of tens of thousands of people and impacted data science to the present day. Soon after, Provost and Fawcett [29] released their own epic on the heels of the famous Davenport and Patil article. The book, entitled *Data Science for Business*, took up the charge by exploring data science more deeply in a business context. It was not the first book on the use of data analysis for business, but it was novel in that it communicated this burgeoning field by focusing not on the algorithms, but rather on their utility. While articles and books helped plug the emerging gap for practicing data science professionals, the government was engaged in filling the technological gap by using high-performance computing.

The Office of Science and Technology Policy, in concert with the Networking and Information Technology Research and Development Program, began planning in 2012 to establish a national high-performance computing (HPC) resource. By 2015, a solicitation emerged to create four regional Big Data hubs to enable data science innovation and create enduring partnerships among academia, government, and industry. The guiding principle is that the impacts of HPC should not aid a single campus or research center. Rather, the resources must accelerate the use of HPC for Big Data across the nation [30].

In the White House, D.J. Patil, the first US Chief Data Scientist, named in 2015, discussed another first to support the Open Government Initiative as promulgated in Executive Order 13642, which mandated open and machine-readable data as the new standard for government information [31]. Patil used his address to articulate his roles:

- "Creating nationwide data policies that enable shared services and forward-leaning practices to advance our nation's leadership in the data age."
- "Working with agencies to establish best practices for data management and ensure long-term sustainability of databases."
- "Recruiting and retaining the best minds in data science for public service to address these data science objectives and act as conduits among the government, academia, and industry" [32].

Among Patil's focus areas, he highlighted: precision medicine and usable data products – the data products that enliven the insights of open, machine-readable data. All in all, the 21st century has brought amazing momentum to data science from diverse sources. Additionally, it has helped bind together a century of work and progress toward harnessing data to improve the performance of government. It sets up a new playing field – a new frontier – to more ambitiously tackle the challenges of using Big Data at scale in a governmental context.

2.2. New Frontiers: Harnessing Big Data and Context at Scale in Government

When it comes to the decision-making process, we seek answers from a reliable source. Modern technologies give us the ability to describe data, to process data, and to share data, but the basic human process is no different to how it was millions of years ago. It is a condition of the human spirit to find the cause of a phenomenon to explain it and answer questions. The pursuit is noble if the journey is sound. The situations presented in the introduction are data-centric approaches to measuring government performance. Using data to address the problem is a natural evolution of open government and performance management, but requires the appropriate application of data science by coalescing

the people involved, the processes they manage, and the technologies available to better utilize the sea of data available – this is the main premise of this paper and the model presented (iGPS).

Job postings for government and contracted positions among federal organizations have begun to indicate a growing demand for data science specialties. This trend is a step in a positive direction; however, creating and organizing data science teams in government poses unique challenges. In this time of organizational change, a simplification of the human dimension is helpful. Gendron *et al.* [33] present a framework of archetypes to capture the four essential data science functions that government leaders should assemble when building a data science team.

This framework not only simplifies the concept of a data science team’s functions, but also reinforces the notion that data science is more than just machines and business processes. Archetypes allow for personification of the underlying data science functions. This can help map the necessary data science functions to job descriptions that already exist within an organization. For instance, project and product managers have skills that align with the Strategist archetype, whereas people with backgrounds in statistics, economics, and computer science are more apt to fulfill the role of The Designer [34]. This structure helps with any data system in government, such as the iGPS system presented in section 3.

Ask someone at Google “What is the data science process?”, and you will likely get many answers. Processes develop organically and reflect the environment of the organization applying those processes. In a democratic republic, governmental organizations have a unique process challenge in that they are accountable to society for the conduct of their business. This means that they need to maintain the public trust and answer to the public directly or to their representatives. Government agencies assess their bottom line in public perception. This has implications for establishing government data science processes.

Sewadeh *et al.* [35] make the point clearly: responding to the desire for data in a rapidly changing technological environment requires new processes. Table 2 highlights four challenges that government must overcome in reengineering their operating processes. Even after building the right data science teams and infusing the organization with new processes, that workforce will need the proper tools and resources to execute the new processes. Batarseh *et al.* [36] note that “there are many ‘missed chances’ by the government, ones that could be easily utilized to exponentially maximize the efficiency and effectiveness of governmental processes”. Specifically, they present three areas where governments may focus attention and tune their technology infrastructures to the questions they pose and the answers they seek: Knowledge-based Systems, Context Management, and Machine Learning.

Challenge	Implication of challenge	Possible solution
Data Silos	Inconsistent metadata, variations in data definitions, variations in dataset aggregation, and poorly communicated data quality	Reviewing the cohesiveness of legacy data systems. In some cases, discontinue their use in favor of adopting new data dissemination processes
Lack of Metadata Standards	Inability of users to find data and, when data is found, the likelihood of misunderstanding its context	Structuring data for consistency across mixed platforms to aid in providing a semantical understanding

Understanding User Needs	Dissatisfaction among users who demand more tailored and custom data in a seamless manner	Shifting the organizational culture by focusing on user needs and taking actions based on feedback
Communicating Uncertainty	General inability among most people in understanding the effects of variance in data used for decision-making	Communicating the aspects of uncertainty when providing information as open data

Table 2. Summary of data process challenges, their implications, and possible solutions

Governments will be required to surmount the technology challenges if they hope to avoid ‘missed chances’. The struggle is real and must not be underestimated. However, similar challenges existed in the private sector as well. Industries such as finance, medicine, and insurance must all adhere to high standards and regulatory requirements. They mimic governmental challenges in many ways because ‘simply’ using open source tools, for instance, poses a large problem for auditable IT infrastructures. Yet, they have found a balance in extracting information from their data while preserving the integrity of their businesses. They may provide many lessons for government implementations. All the challenges mentioned beg for the answer to a very important question: how can *data science* and *context* be used to improve government accountability and enhance citizen service? The next section addresses this question through the contributed model.

3. Intelligent Governmental Performance and Services (iGPS)

To address the challenges presented above, an Intelligent Governmental Performance and Services (iGPS) model is presented. The goal of the model is to provide the federal employee/analyst with data that is relevant to the present decision that they must make (context). The data aim to provide better responses to the citizens, in order to eventually improve government accountability. The major differentiator for iGPS is that it uses data to define context, which is a notion that few researchers in context have studied.

In this model, the collection and management of context evolves through three types:

- Domain context: derived from data that are relevant to the current request, as well as the domain knowledge of the federal employee. The iGPS system presents this by extracting the correct data from the SQL server and the Hadoop cluster.
- External context: derived from data that are irrelevant to the current request (situation at hand), but might be useful in other steps of the process or other requests. Those are the datasets that stay in the Hadoop cluster and SQL databases, and are not pulled toward the visualization engine.
- Procedural context: data that show the variables that effect how the analyst will (and should) handle the request (actionable context). The iGPS system provides this through a data visualization. The visualization allows the analyst to quickly identify locations of weakness or issues in citizen support that need to be handled first.

In data science, preparing the data for analysis is usually a very expensive process; most agree that the majority of the effort is in preparing the data. In cases where the data are in fact ‘Big Data’, that challenge is even more prevalent. In the case of government, a federal analyst needs to browse through millions of records to be able to understand the “history” of a certain complaint or use case. More importantly, though, the analyst is not able to find relevant complaints that could benefit from the same solution, or find similar issues that were solved prior. Historically, it is common that the federal employee’s *experience* is what usually drives such decisions, and not actual data that could be visualized and become part of the decision process. In the three types of contextual data management, the employee’s experience is used to “filter” the data that define the domain context (using Power BI). Using data visualization tools, the analysts can choose which columns of data might be relevant to the request. That step leads to procedural context (which is the third type, and is actionable in nature). The interesting part, however, is that such data are available. Through the method presented in this paper, the official procedures corresponding to a specific complaint are contextualized in different practices that are perfectly tailored to any specific demand. There is no lack of data in government, but there is a serious lack of automated intelligent systems that can quickly provide meaningful insights to the federal employee.

3.1. Data Analytics Tools Used within iGPS for Analysis and Context Definition

iGPS is a *software package* that takes advantage of existing powerful tools. The power of iGPS is in its interactivity and context awareness. iGPS uses multiple contextual elements to help the analyst to identify datasets that need to be visualized for decision support. Most government employees have access to Microsoft products, such as Power BI (the US government is one of the most prominent Microsoft customers), and other pieces of software are open-source tools. Within iGPS, tools such as Power BI, Hadoop (open source) and Spark (open source) are leveraged. Before presenting the model, the following tools are presented:

- Microsoft Power BI (used for iGPS Data Visualization): Power BI [37] is a suite of business analytics tools that are used to analyze data and share insights. Power BI’s dashboards provide a 360-degree view for their users with their most important metrics in one place, updated in a real time and available on all their devices. Although iGPS allows for the use of Tableau or Excel to build graphs or visualizations, as the experiment shows, Power BI proved to be the fastest and most suitable. Power BI is fast, easy, and free. Using their gateways, the tool can connect SQL server databases, analysis services, and many other data sources to the same dashboards. If the user already has reporting portals or applications, then the user can embed Power BI reports and dashboards for an enriched, unified experience. It helps the user to combine data from disparate databases, which is a major use case within government. Power BI has a focus mode which means that the user just clicks on the button of the visualization in which they are most interested, and Power BI simply zooms in to fill up most of the screen until the user clicks out of it. It also displays some extra meta-information that might be useful after the original interactivity. Additionally, it has the added feature of data-driven alerts (the capacity to alert the federal analyst upon certain events, via its own web-based notification center).
- Apache Hadoop (used for iGPS Data Management): Apache Hadoop [38][39] is an open-source platform that is used for distributed storage and processing of very large datasets on computer clusters built from commodity hardware. Its services include data storage, data processing, data accessing, security, data governance, and operations. Within iGPS, besides the Microsoft SQL

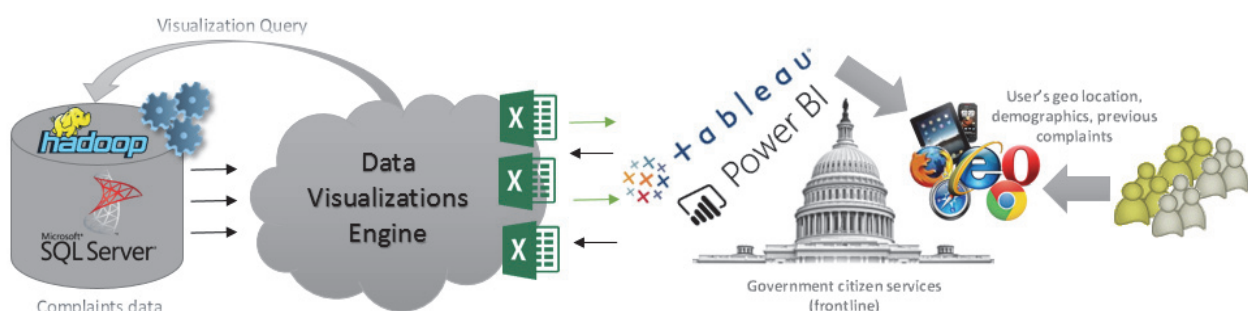
server, Hadoop is used to manage the data repository. The benefits of using Hadoop include: scalability, reliability, and low cost.

- Apache Spark (used for iGPS Data Modeling): Apache Spark [40] is a fast, in-memory data processing engine with elegant and expressive development APIs, which allows the data workers to efficiently execute streaming, machine learning, or SQL workloads that require fast iterative access to datasets. With Spark running on Apache Hadoop's YARN [39], federal developers could create applications to exploit Spark's power, derive informative insights, and enrich their data science workloads within a single, shared dataset in Hadoop. Apache Spark consists of Spark core and a set of libraries. It is a distributed execution engine and the Java, Scala, and Python APIs offer a platform for distributed application development.

3.2. Context-Driven Data Visualizations

iGPS provides *data visualizations on demand* depending on the context. The system reacts to a request based on: (1) the type of request received from a citizen or an institution, (2) previous citizen satisfaction levels, and (3) available historical data (these three points define the domain context of the request, as discussed previously). The federal analyst initiates the process by inputting the complaint or citizen case number, which includes their geographical location, previous complaints, type of complaint, and many other parameters. The *data visualization engine* (DVE) then fetches the Hadoop database for data that has a similar **context**. Complaints from the same geo-region or of the same type (i.e. category of complaint) are collected and queried from Hadoop to the DVE. If there are any Spark models built inside a Hadoop cluster, the results of these models are also sent to the engine (note: building Spark models is done manually in this process, and it is expected to be performed by a data scientist at the government. A frontline complaints employee is not expected to have such skills). Afterwards, the data is exported to Power BI, and it is presented to the federal employee through a data visualization of their choice.

iGPS is illustrated in Figure 2. Defining context is difficult. Context as an idea has infinite dimensions. For the method presented in this paper, context is driven from data; however, it is challenging to specify the limitations of data collection, and what types of data to collect [41]. When context is abstract, it is unclear how to identify the constraints of the data needed to develop the required visualizations, given that data to define that context can be infinite [42]. Parts of the data will be irrelevant to the context, while other parts will be highly relevant; therefore, it is important to pre-identify data that fall within the scope of the context. These are identified as contextual elements. Each element has different instantiations. The decision made by the federal analyst is dependent on the different variations of elements. The type of request, historical data, and user satisfaction are the elements of the context in iGPS.



iGPS is developed using C#, and is tested with a number of *real* government datasets. Multiple visualizations are built through the system. The next section presents the datasets and experiment, along with the data visualizations developed to evaluate the presented data system.

4. Experimental Work

This section presents all the datasets presented in the evaluation of iGPS. Data mining models are also provided to illustrate how data science and context could be used to improve decision-making at the frontlines. Finally, example dashboards are presented.

4.1. Governmental Datasets (Inputs to the Experiment)

This section presents six major datasets that are used in the experiment. The datasets are available to the reader: please contact the authors for a download link. In iGPS, the datasets are placed in one Hadoop repository, and data columns can be extracted on demand. Consumer Complaints Dataset: this dataset [43] was published by the Consumer Financial Protection Bureau (CFPB), which is responsible for consumer protection in the financial sector. CFPB jurisdiction includes banks, credit unions, security firms, payday lenders, mortgage-servicing operations, debt collectors, and other financial companies operating in the United States. This dataset has a list of all the complaints received from the consumer's end about financial products and services. This dataset contains information such as date received, products, sub-products, issue, sub-issue, consumer-complaint narrative, state, zip code, tags, consumer consent provided, submitted via, date sent to company, timely response, consumer disrupted, and complaint ID. The most important fields of this dataset are:

- The date-received columns, which note the date on which a consumer complaint was received.
- The products field, which notes the type of product that the consumer had an issue with.
- The sub-products field, which notes the area to which the product belongs. For instance, a consumer complaint was received on 02/04/2015 for debt collection, which is categorized in the area of *mortgage*.
- The issue field, which notes what problem or difficulty the consumer was facing with the product that they mentioned.
- The submitted via field, which notes the mode of submission for every customer in order to know how they turned in their complaints.
- The date sent field, which notes the date when the consumer sent their complaints.
- The complaint ID field, which notes the unique identification number for each complaint generated.

Oregon Consumer Complaints Dataset: this dataset [44] has a list of all the consumer complaints from the State of Oregon that are registered with the Oregon Department of Justice, which are derived from the consumer contacts during the past three years. The Oregon Department of Justice's main goal is to ensure a safe and fair marketplace in Oregon. This dataset contains information such as: reference number, status, date opened, date closed, respondent, address 1, address 2, city, state, zip code, business type, complaint description, closing description, and geographical location (geo data location). For example, a consumer complaint with reference number FF321-14 was opened on 09/03/2014 and closed on 09/11/2014. The respondent for this complaint was *Walgreens* with the

address of Kelly Ware – Regulatory Law, Litigation & Regulatory Law Department; Deerfield, IL, 60015. The type of business was Pharmacies & Drug Stores. The complaint description was Private Class Action, and was closed with an action of Class Action Fairness Act Notice. This complaint currently holds the status of closed, as it has already been attended to.

Consumer Services Mediated Complaints Dataset: this dataset [45] was published by the City of New York. It features consumer complaints against businesses that were mediated by the DCA Consumer Services Division during the past few calendar years. This dataset has information about business names, their industry types, complaint type, mediation start date, mediation close date, complaint result, satisfaction, restitution, business building location, business address, business city, business state, zip code, complainant zip code, longitude, and latitude. For example, the business LIL PETES AUTOMOTIVVE INC of industry type *Tow Trucks* had a complaint for a billing dispute. The mediator started action on 4/7/2017 and ended it on the same day. The complainant was satisfied with the response given. The responsible business's address is noted as 635 S Columbus Ave, Mount Vernon, NY, 10550, and the complainant's zip code is 10701.

Attorney General Consumer Complaints Dataset: this dataset [46], published by the State of Washington, consists of complaint data from consumer complaints filed with the Consumer Protection Division. The Consumer Protection Division is composed of attorneys and professional staff, which enforces the act. The agency's goal is to help keep Washington's marketplace free of unfair and deceptive practices. This division provides information and education to businesses and the public on consumer issues and issues alerts. It provides press releases to warn consumers and businesses about fraudulent or predatory activities.

The North American Industry Classification System (NAICS) classifies business establishments for collecting, analyzing, and publishing statistical data related to the US economy. The NAICS industry codes define establishments based on the activities in which they are primarily engaged. This dataset consists of information related to the complaint's opened date and year, the status of the complaint, estimated and actual savings, business category, NAICS code and name, business name, business street address, city, state and zip code, business location, business identification (ID) number, and complaint ID. For example, a complaint was opened on 06/03/2016 with the Consumer Protection Division under the business category of Government Agencies and was given a NAICS tag of 921190-Government (General) and had a business name of Department of Revenue based in Olympia, WA, 98504. The business's ID number is noted as 58567 and the Complaint's ID is noted as 486363.

CGB Consumer Complaints Dataset: the dataset [47] published by the Federal Communications Commission (FCC) has individual informal consumer complaints filed with the consumer help center beginning October 31, 2014. The FCC is an independent agency of the United States government created by the statute to regulate interstate communications by radio, television, wire, satellite, and cable. It is funded entirely by regulatory fees and works towards goals in the areas of broadband, competition, the spectrum, media, public safety and homeland security, and modernizing itself. This dataset has information about the ticket's identification number, date of ticket creation, date of issue, time of issue, form, method, issue, caller ID number, type of call or message, advertiser business number, city, state, zip code, and location. For example, a ticket with ID 1000296 was created on 05/25/2016 at 11:15:29 AM. This issue was said to be reported on 05/1/2016 at 1:00 PM. This complaint was registered with the consumer health center by phone and the caller ID number was 866-410-0458. The address of the business is noted to be based in Plantation, FL, 33324.

MMR Agency Performance Indicators Dataset: this last dataset [48][49], published by the City of New York, has agency-specific performance indicators from the FY16 and FY17 Mayor’s Management Report (MMR). The MMR is mandated by the City Charter and is served as a public account of the performance of city agencies, measuring whether or not they are delivering services efficiently, effectively, and expeditiously. The full fiscal MMR is published each September. This dataset has information about different agencies. It contains information about the performance of each agency, has a tab on all of the performance indicators for the fiscal years 2012–2016, and shows its target levels for 2016 and 2017.

Based on the past performance indicators, it also predicts its desired growth for the next five years. For example, for both of the fiscal years 2016 and 2017, there are statistics for agencies such as ACS and BIC, and their critical analysis in terms of true or false has a performance indicator such as “Abuse and/or neglect reports responded to within 24 hours of receipt from the State Central Register” and has its statistics for the years 2012 – 2016. It also has an estimated percentage for its targeted years of 2016 and 2017. This dataset has information about all the NYC-based agencies and their measure of performance indicators. It also captures MMR goals, their ratios and a list of minimum and maximum values. For example, the agency NYPD has around 68 performance indicators and about nine MMR goals. This dataset is also available through www.data.gov. The six datasets presented (representing 46 governmental agencies) are used to build data models and data visualizations through iGPS. This is presented in the next section.

4.2. Experimental Results

All of the datasets presented above were migrated into the Hadoop cluster of iGPS. They were all used, but not all of them had obvious actionable insights. A variety of visualizations are created, and some of them are descriptive of the data while others are predictive in nature, and aim to provide clues as to the future. This section also provides a review of the results that are created through data visualizations. The goal of the visualizations is to give the government employee a valuable way in which to communicate important information. For example, the *MMR Agency Performance Indicator* dataset (section above) is used to create a visualization in the form of a scatter plot that underscores certain trends.

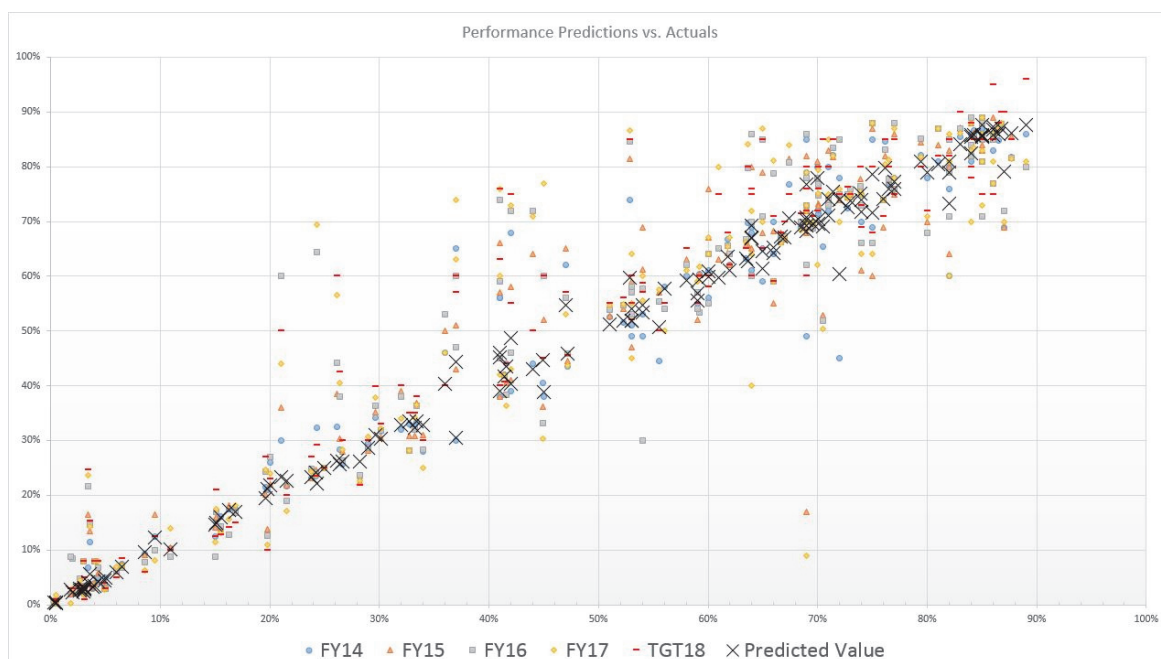


Figure 3. Predicted vs. Actual Government Performance

Figure 3 shows all of the reported performance metrics: the x-axis shows task completeness, and the y-axis shows the percentage goal set by the agency. The graph aims to show the predicted completeness in comparison to what the agency expects to be completed by 2018. Based on historical trends, a linear regression model is built in Spark to provide the forecast. The black “x”s are the predicted values based on the model. Each data point is a task, and its shape/color represents its fiscal year.

On Figure 4, all of the data points in the white area (area above the trend line), represent tasks that are predicted to miss their deadline. This is a model that shows how some deadlines can be underestimated, and some tasks have not shown sufficient historical progress to indicate that they will improve or be accomplished in the future. Such predictions can be leveraged by readjusting the expectations, and can minimize frustrations and difficult deadlines within a federal agency.

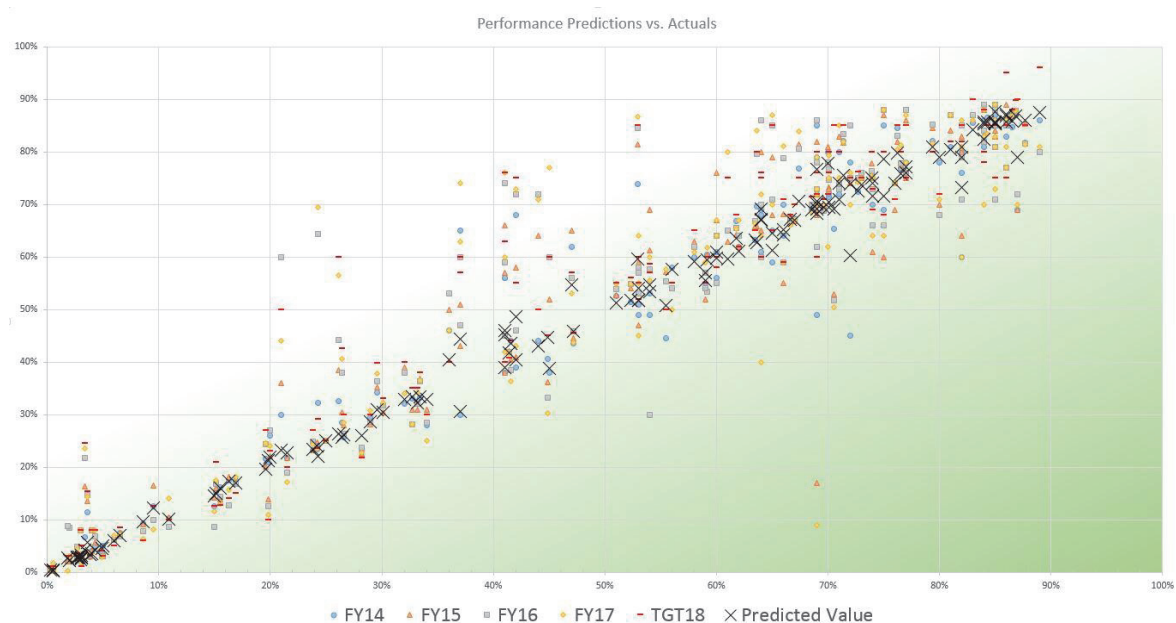


Figure 4. In the White Area: Data Points that are Predicted to Miss the Deadline

Figure 5 shows the performance indicators for all of the agencies that are mentioned in the datasets. It quickly shows how each agency fares in their performance and how fast they are in resolving consumer complaints. The Department of Commerce seems to be doing very well based on this dataset.

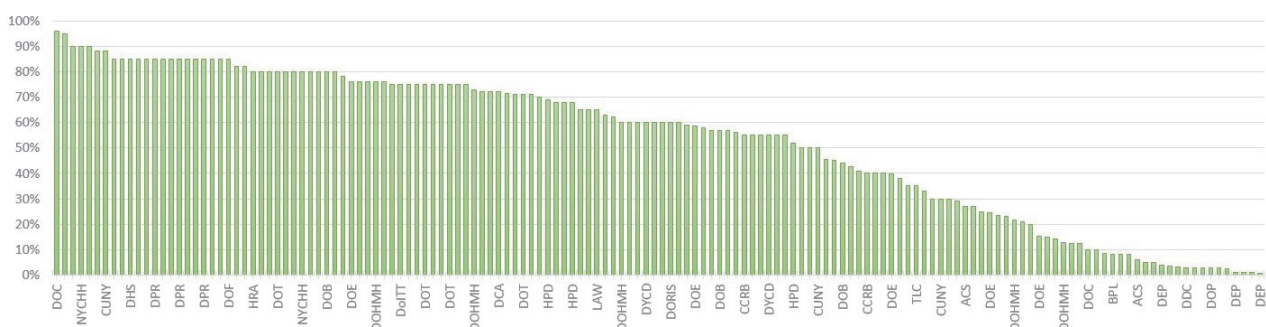


Figure 5. Performance Metrics Success Percentage for Multiple Agencies/Departments

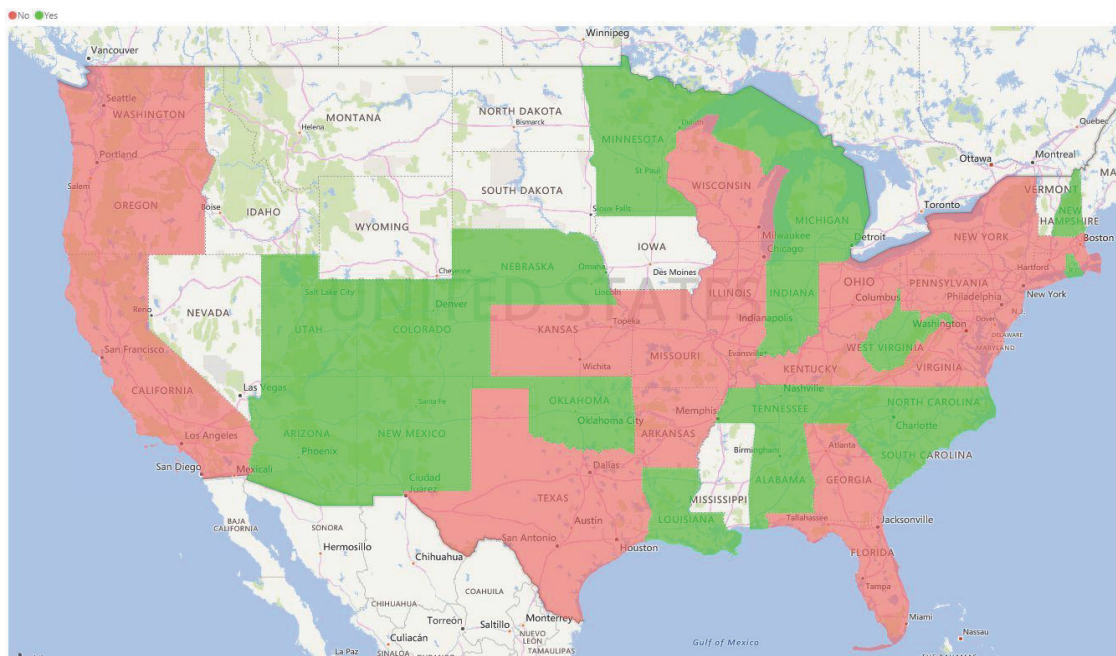
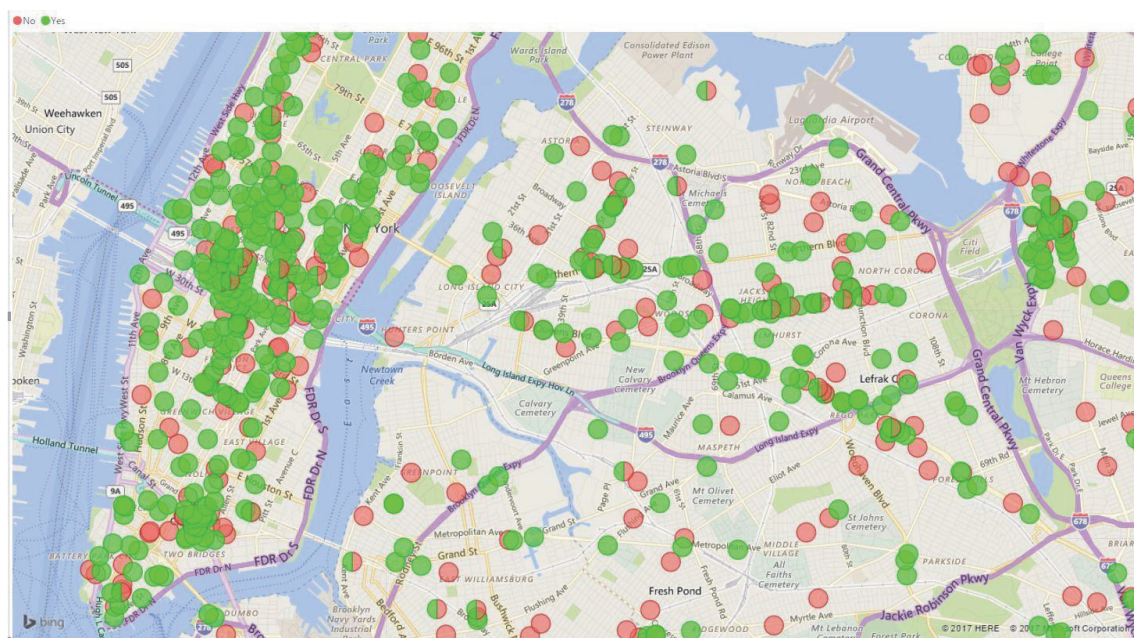


Figure 8 shows satisfaction rates per zip code. Let us take New York City as an example: the analysts note that in midtown Manhattan, there seems to be a general acceptance of citizen service, while in other parts, acceptance is on the negative side. It is important to note that geography is only one way of showing a detailed distribution of citizen service feedback; many other visualizations can be developed through iGPS to help the federal employee make better decisions. The following section presents the conclusions and points toward future work that would improve iGPS, enforce its successful aspects, and eliminate some of its shortcomings.



5. Conclusions and Future Work

As public trust in government services declines, reform is desperately needed to address the challenges of citizen service. The difficulties of citizen service within government include managing

citizen service mandates among different levels of government, retaining technically qualified managerial and labor staff, dealing with an overabundance of data leading to flawed analysis, and adjusting to technological changes. Performance metrics in government are intended to assist with challenges in citizen service. These challenges often have their own set of complications, including: differing data definitions and technologies used by agencies, a lack of effective performance metrics, and a lack of capable top management.

To assist with citizen service within government, the Intelligent Governmental Performance and Services (iGPS) model attempts to provide federal employees/analysts with data that is relevant to their current decision. Using six major datasets consisting of citizen complaints to government, iGPS uses tools such as Power BI and Hadoop to produce its data visualizations. The following model addresses many of the concerns cited regarding citizen service within government and performance metrics.

When using software such as the SQL Server or Hadoop to manage a database, different levels of government may no longer need to use different methodologies to compare performances. Uniform data standards can lead to speedy top-down initiatives and greater integration between agencies. Data visualizations are useful when a summary of a certain type of complaint is required. The visualizations produced by iGPS could be shared inside the agency or across agencies for better communication of issues and how to address them. iGPS can also help analysts to make coherent connections among differing data definitions. Common data visualizations could also be used to gauge the effectiveness of performance metrics. Additionally, visualizations are especially effective for displaying concepts to non-technical federal employees. Furthermore, the visualizations that iGPS provides can also be shared with citizens (Open Data Initiative) to help them to better understand policy-making and identify discrepancies in government data. For example, consumers and media appreciated the use of XML in Virginia’s election result reporting in 2008 [50]. However, there was a lack of data standards which led each state/reporting entity to “coordinate its efforts with media outlets independently”. Such incidents represent the need for greater data analytics to record solutions.

According to practice and the literature, government agencies are awash with performance measures. They do well when it comes to tracking a few, but the exact number is situational. Niven’s rule of thumb [24] is to have no more than **20** performance measures if possible. He finds that even the most complex organizations can attain this target. There is a need for multiple measures per objective when using leading and lagging indicators, and so he establishes a minimum number of performance measures as 1.5 times the number of objectives tracked. How did the City of New York (NYC) do according to its FY17 Mayor’s Management Report (MMR) Agency Performance Indicators? Of the 46 agencies used in this study, over two-thirds (32 to be specific) have too many performance measures according to Niven’s criteria, and only 14 had a good number of measures.

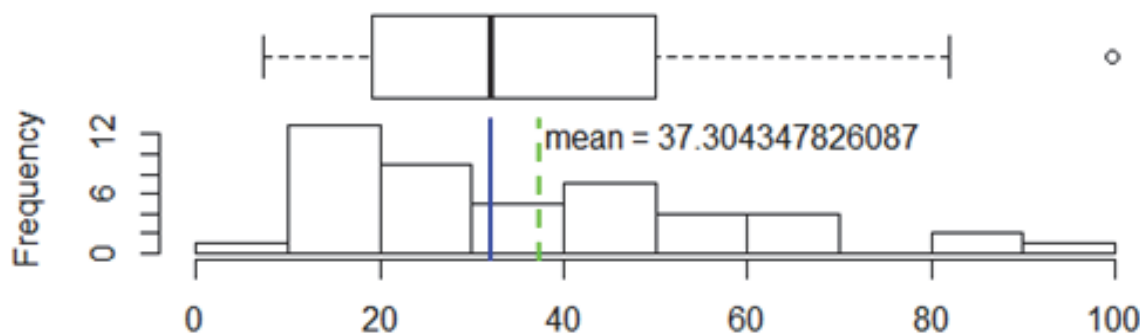


Figure 9. Histogram of Total Performance Indicators per Agency

Agencies supported anywhere between 1 and 17 MMR goals (median = 6, mean = 6.7). The total number of performance measures ranged from 7 to 100 (median = 32, mean = 37.3). Figure 9 shows how the histogram of performance measure usage among all 46 agencies is skewed to the right. On a level of two-to-one, NYC agencies under the MMR system used more performance measures than recommended by best practice. Through this analysis, an agency can reduce the number of performance measures in order to focus agency resources. It could result in increased performance. Using the iGPS model, the goals of citizen service within government can be achieved with greater speed, efficiency, and unity. This is particularly relevant when visualizations are posted on government websites [51]. Potential future steps that aim to improve iGPs and its deployment in the US government include:

- Developing a survey to gauge federal employees' usability and iGPS acceptance rates.
- Adding more aspects to the context of data visualization developments.
- Deploying the tool in multiple federal agencies that constantly deal with citizen service.
- Adding more tools to the system – currently, iGPS supports Power BI and Tableau, but the goal is to apply Unfolding Maps (a java-based open-source mapping tool).
- Deploying more advanced data-mining algorithms, such as K-NN for clustering citizen complaints, and testing the iGPS system with more datasets.

The datasets used in this paper, the algorithms built in Spark and Hadoop, the DVE software code, information on the contexts tested, and different components of the iGPS system are available to the reader. Please contact the authors for further information and for links to download the files. Ultimately, the iGPS model also offers a solution to the overabundance of data within government. Using tools such as Spark, mountains of data can be analyzed for federal agents. As a result, cases can be solved with greater efficiency and precision. iGPS may even help with the issue of retaining technically qualified managerial staff (the tools used for Big Data management in iGPS are cutting-edge). iGPS acts as a catalyst for what government aims to achieve: customer satisfaction, transparency, and quality service.

Bibliography

- [1] Ellis, R. K., What will bad customer service cost government?, *Public Manager*, vol. 44-2, pp. 35–37, 2015, Retrieved from: <https://search.proquest.com/docview/1688659682?accountid=14541>.
- [2] Thompson, F. J., Revitalizing state and local public service: Strengthening performance, accountability, and citizen confidence, San Francisco: Jossey-Bass, 1993.
- [3] Civil Service Reform Act of 1978, U.S. Equal Employment Opportunity Commission. Retrieved from: https://www.eeoc.gov/eeoc/history/50th/thelaw/civil_service_reform-1978.cfm.
- [4] Waldman, R. L., What the Government Does with Americans' Data, *Brennan Center for Justice*, pp. 1–80, November 12, 2017, Retrieved from: <https://www.brennancenter.org/publication/what-government-does-americans-data>.
- [5] Fadairo, S. A., Williams, R., D.B.A., C.P.A., & Maggio, E., J.D., Using data analytics for oversight and efficiency, *The Journal of Government Financial Management*, 64 (2), pp. 18–23, 2015, Retrieved from: <https://search-proquest-com.mutex.gmu.edu/docview/1711620157?accountid=14541>.
- [6] Congress, U. S., Government Performance and Results Act of 1993, In 103rd Congress. Congressional Record, 1993.
- [7] Bell, P., Public Trust in Government: 1958–2017. Retrieved October 27, 2017, Retrieved from <http://www.people-press.org/2017/05/03/public-trust-in-government-1958-2017/#>.

- [8] S. Cavalluzzo, Ken & D. Ittner, Christopher, Implementing Performance Measurement Innovations: Evidence from Government, *Accounting, Organizations and Society*, section 29, pp. 243–267. DOI: 10.1016/S0361-3682, 2003.
- [9] Kravchuk, R. & Schack, R., Designing Effective Performance-Measurement Systems under the Government Performance and Results Act of 1993, *Public Administration Review*, section 56(4), pp. 348–358. DOI: 10.2307/976376, 1996.
- [10] How the Government Measures Unemployment. Retrieved October 27, 2017, from https://www.bls.gov/cps/cps_htgm.htm#unemployed.
- [11] Fountain, J. R., Service Efforts and Accomplishments Reporting, *Public Productivity and Management Review*, vol. 15 (2), 1991.
- [12] Jones, L. R. & McCaffery, J. L., Implementing the Chief Financial Officers Act and the Government Performance and Results Act in the Federal Government, *Public Budgeting & Finance*, section 17, pp. 35–55, DOI:10.1111/1540-5850.01091, 1997.
- [13] Peled, A., Traversing digital Babel information, e-government and exchange, Cambridge, MA: The MIT Press, 2014.
- [14] Program Performance Report, Smart Policing Initiative Grant Program, Retrieved from: https://www.bja.gov/Publications/SPI_PPR_Jan-Jun2013.pdf.
- [15] Office of Management and Budget (OMB), Government performance results act of 1993, Washington, DC: OMB. Retrieved from: <https://obamawhitehouse.archives.gov/omb/mgmt-gpra/gplaw2m>.
- [16] Mokyr, J., The second industrial revolution, 1870-1914, 1998, Retrieved from: <https://pdfs.semanticscholar.org/769c/a06c2ea1ab122e0e2a37099be00e3c11dd52.pdf>.
- [17] Fisher, R. A., The design of experiments, Edinburgh: Oliver and Boyd, 1937.
- [18] Shewhart, W. A., Statistical method: From the viewpoint of quality control. W. E. Deming (Ed.), Washington, DC: The Department of Agriculture, 1939.
- [19] Tukey, J. W., Exploratory Data Analysis. Reading, PA: Addison-Wesley, 1997.
- [20] Martínez-Lorente, A. R., Dewhurst, F., & Dale, B. G., Total quality management: Origins and evolution of the term. The TQM Magazine, 10(5), 378-386, DOI: 10.1108/09544789810231261, 1998.
- [21] National Institute of Standards and Technology (NIST), How Baldrige works. United States Department of Commerce, 1996, Retrieved from: <https://www.nist.gov/baldrige/how-baldrige-works>.
- [22] National Institute of Standards and Technology (NIST), Baldrige program name change emphasizes performance excellence [Web log], 2010, Retrieved from: <https://www.nist.gov/news-events/news/2010/10/baldrige-program-name-change-emphasizes-performance-excellence>.
- [23] Kaplan, R. S. & Norton, D. P., The balanced scorecard: Translating strategy into action, Boston, MA: Harvard Business Review Press, 1996.
- [24] Niven, P. R., Balanced scorecard: Step-by-step for government and non-profit agencies, Hoboken, NJ: John Wiley & Sons, 2003.
- [25] Hey, T., Tansley, S. & Tolle, K. (Eds.), The fourth paradigm: Data-intensive scientific discovery. Redmond, WA: Microsoft Research, 2009. Retrieved from: https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf.
- [26] Orszag, P. R., Memorandum: Open Government Directive, December 2009, The White House, *Washington, DC Weekly Presidential Address*. Retrieved from: <https://obamawhitehouse.archives.gov/open/documents/open-government-directive>.
- [27] Obama, B., Weekly Address: President Obama Discusses Efforts to Reform Spending, Government Waste; Names Chief Performance Officer and Chief Technology Officer, April 2009, The White House, Washington, DC Weekly Presidential Address, Retrieved from: <https://obamawhitehouse.archives.gov/the-press-office/weekly-address-president-obama-discusses-efforts-reform-spending-government-waste-n>.
- [28] Davenport, T. H. & Patil, D. J., Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, Edition 90 (10), pp. 70–77, 2011.
- [29] Provost, F. & Fawcett, T., Data science for business: What you need to know about data mining and data-analytic thinking. Sebastopol, CA: O'Reilly Media, Inc, 2013.
- [30] National Science Foundation (NSF), Big data regional innovation hubs: Establishing spokes to advance big data applications (BD spokes), 2017, Retrieved from: https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505264.

- [31] Exec. Order No. 13642, 3 C.F.R., 2013. Retrieved from: <https://www.gpo.gov/fdsys/pkg/CFR-2014-title3-vol1/pdf/CFR-2014-title3-vol1-eo13642.pdf>.
- [32] Patil, DJ., Memorandum: Unleashing the Power of Data to Serve the American People, 2015, *The White House, Washington, DC Address*. Retrieved from: <https://obamawhitehouse.archives.gov/blog/2015/02/19/memo-american-people-us-chief-data-scientist-dr-dj-patil>.
- [33] Gendron, G. R., Mortimer, S., Crane, T., & Haynes, C., Transforming data science teams in the future. In F. A. Batarseh & R. Yang (Eds.), *Federal data science: Transforming government and agricultural policy using artificial intelligence*. Elsevier Academic Press. ISBN: 9780128124444, 2017.
- [34] Batarseh, F. & Yang, R., *Federal data science: Transforming government and agricultural policy using artificial intelligence*. Elsevier Academic Press. Chapter 13, 2017, Retrieved from: <https://www.elsevier.com/books/federal-data-science/batarseh/978-0-12-812443-7>. ISBN: 9780128124444.
- [35] Sewadeh, M. & Sisson, J., Disseminating government data effectively in the age of open data. In F. A. Batarseh & R. Yang (Eds.), *Federal data science: Transforming government and agricultural policy using artificial intelligence*. Elsevier Academic Press. ISBN: 9780128124444, 2017.
- [36] Batarseh, F. A. & Yang, R., Making the case for artificial intelligence at government: Guidelines to transforming federal software systems. In F. A. Batarseh & R. Yang (Eds.), *Federal data science: Transforming government and agricultural policy using artificial intelligence*. Elsevier Academic Press. ISBN: 9780128124444, 2017.
- [37] Microsoft Power BI, What is Power BI? <https://powerbi.microsoft.com/en-us/what-is-power-bi/>, 2017 (accessed 20 September 2018).
- [38] Apache Hadoop, HortonWorks <https://hortonworks.com/apache/hadoop/>, 2017 (accessed 20 September 2018).
- [39] Hadoop, What is it and why does it matter? https://www.sas.com/en_us/insights/big-data/hadoop.html, 2017 (accessed 20 September 2018).
- [40] Apache Spark, HortonWorks <https://hortonworks.com/apache/spark/>, 2017 (accessed 20 September 2018).
- [41] McCarthy, J., Notes on Formalizing Context, *International Joint Conference on Artificial Intelligence*, vol. 1., pp. 555–560, 1993.
- [42] Bazire, M. & Brezillon, P., Understanding Context Before Using It, *Lecture Notes in Computer Science*, Proceedings of the 5th International and Interdisciplinary Conference on Context, pp. 29–40, DOI: 10.1007/115083733, 2005.
- [43] Bureau of Consumer Financial Protection, Consumer Complaint Database, September 26, 2015. <https://catalog.data.gov/dataset/consumer-complaint-database>.
- [44] State of Oregon, Oregon Consumer Complaint Database, November 1, 2017. <https://catalog.data.gov/dataset/oregon-consumer-complaints-7f511>.
- [45] City of New York, Consumer Services Mediated Complaints, November 7, 2017, <https://catalog.data.gov/dataset/consumer-services-mediated-complaints>.
- [46] State of Washington, Attorney General Consumer Complaints, November 7, 2017, <https://catalog.data.gov/dataset/attorney-general-consumer-complaints>.
- [47] (dataset) Federal Communications Commission, CGB- Consumer Complaints Data, November 7, 2017, <https://catalog.data.gov/dataset/cgb-consumer-complaints-data>.
- [48] (dataset) City of New York, FY16 MMR Agency Performance Indicators, June 29, 2017, <https://catalog.data.gov/dataset/fy16-mmrm-performance-indicators-data-for-open-data>.
- [49] (dataset) City of New York, FY16 MMR Agency Performance Indicators, October 31, 2017, <https://catalog.data.gov/dataset/fy17-mmrm-agency-performance-indicators>.
- [50] Stenbjorn, P., The Use of XML in Election Results Reporting, A historical review of Virginia’s use of XML in 2008. Retrieved from: <https://www.nist.gov/sites/default/files/documents/itl/vote/paper-stenbjorn-VA-use-xml.pdf>.
- [51] Graves, A. & Hendler, J., A study on the use of visualizations for Open Government Data, *Information Policy*, section 19, pp. 73–91, DOI: 10.3233/IP-140333, 2014.