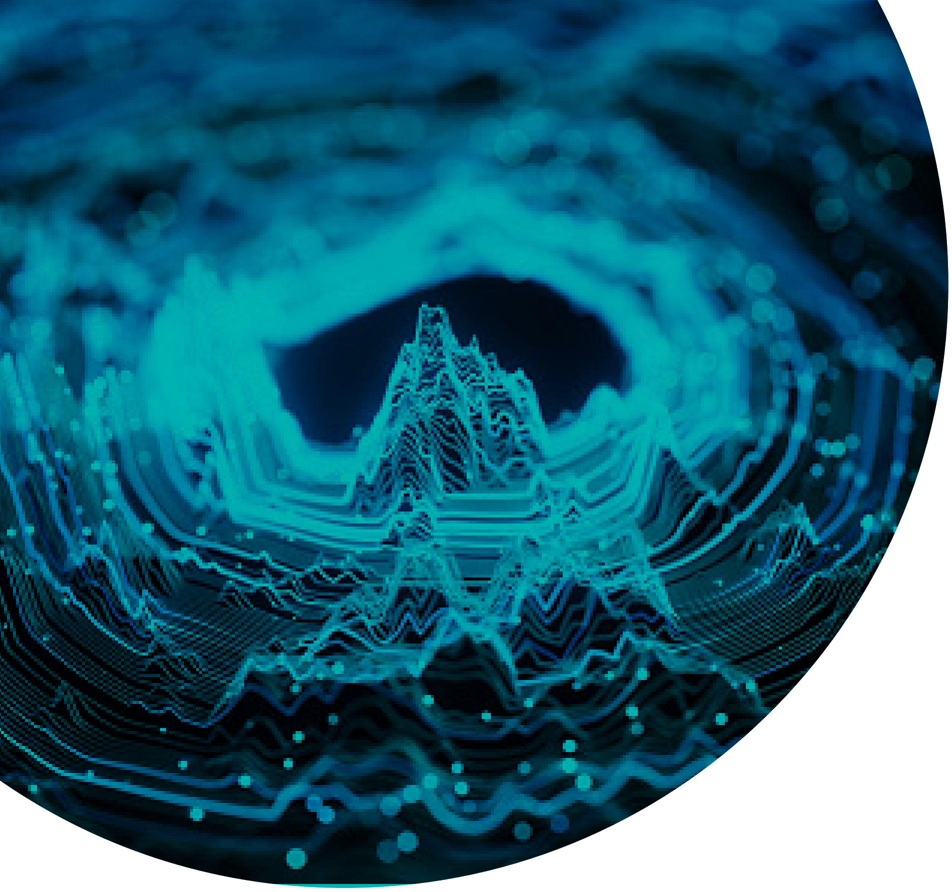


Holistic Smart Cities Data Management and Analytics to Improve the Citizen Experience

Table of Contents

Smart cities and municipalities increasingly use a combination of smart sensors, devices, and communications technology to inform decision-makers about operations, infrastructure, and services to streamline costs while improving the citizen experience. Yet, while collecting and integrating data from these different sensors and devices is intended to influence better decisions to improve operations, a persistent challenge is the perception that the sole purpose of leveraging these innovative technologies is to manage costs or increase efficiency.

According to a 2018 McKinsey Global Institute Report, "… municipal leaders are realizing that smart city strategies start with people, not technology. 'Smartness' … is about using technology and data purposefully to make better decisions and deliver a better quality of life."[1]

Fortunately, there is a growing recognition that smart cities technologies and adoption can go beyond operational efficiency and cost savings to be more broadly employed to enhance the quality of life for all citizens. According to a 2018 McKinsey Global Institute Report, "… municipal leaders are realizing that smart city strategies start with people, not technology. 'Smartness' … is about using technology and data purposefully to make better decisions and deliver a better quality of life."[1]

---

1 "Smart cities; Digital solutions for a more livable future," McKinsey Global Institute, June 5, 2018, accessed via https://www.mckinsey.com/business-functions/operations/our-insights/smart-cities-digital-solutions-for-a-more-livable-future on 2022-04-20

Siloed smart city applications (such as sensors, attached to street light poles intended to improve automobile traffic flow or for monitoring a building's energy consumption to lower utility costs) contribute to operational efficiency. At the same time, smart city technologies are increasingly being used to achieve a variety of outcomes positively impacting the quality of life across a variety of constituencies. Some examples include:

- Utilizing the same light-pole sensors to enable equitable mobility safety for everyone, including pedestrians, cyclists, and those in wheelchairs.
- Expanding public transportation, ride-hailing, and micro transit options to lower-income neighborhoods to support job access or improve food security.
- Monitoring water flow rates, oxygenation, and quality across a wide swath of stormwater outfalls to identify potential water contamination.
- Mapping infrastructure investments to ensure that they are deployed equitably across underserved communities.
- Utilizing thermal sensors to monitor individual temperatures at health provider locations to identify emerging flu outbreaks.

Emerging legislation (including the 2021 Infrastructure Investment and Jobs Act) and accompanying advocacy and grant programs (such as the Vision Zero Network) help to raise awareness of the possibilities and contribute to the evolution of a truly smart municipal environment architecture. However, to be fully "smart," that environment must be more than just a portfolio of individual point solutions.

In this paper, we consider some of the principles influencing the success of a smart city strategy, especially concerning a holistic approach to collecting, managing, preparing, and analyzing data. The paper then considers some examples of coordinated smart city projects and their dependence on accumulated data. Next, the paper considers some challenges that can impede the success of a smart city program, followed by a high-level technology blueprint facilitating smart city analytics. The paper concludes with considerations and suggestions for drafting a systems architecture with an eye for analytics that support smart cities and quality of life of citizens, living in connected communities.

Principles Influencing Smart City Success

A popular perception of what makes cities "smart" is rooted in the concept of continuous collection of machine-generated data streaming from a multitude of sensors, devices, meters, mobile devices, and cameras (the "things" in the Internet of Things, or IoT) distributed across a municipality's geography. For example, the United States International Trade Administration's overview of their support for the IoT and Smart Cities industry notes that "Smart Cities development is largely driven by technologies such as the Internet of Things (IoT), which is broadly defined as an internet-connected sensor or actuator that can collect and share data."[2]

But despite the enthusiasm for introducing sensors and actuators across the city, these devices are not the only source of data that contribute to a smart city strategy. Generally speaking, eGovernment refers to the use of information and communication technology to facilitate the flow of information across government processes, support the efficient provision of government services to citizens and businesses, and increases transparency in providing public access to government data. Some examples include systems supporting permitting and licensing, waste management, emergency response, or citizen service requests ("311").

A forward-looking approach to smart cities does not measure success in terms of the number of deployed sensors or the installation of point solutions. A holistic alternative considers key performance measures tied to the degree to which integrated information and communication technologies simultaneously improve jurisdictional operations while contributing to enhanced quality of life and equitable treatment for all citizens. This suggests some fundamental principles that may influence the success of smart cities:

2 "Overview of the International Trade Administration Support for the IoT and Smart Cities Industry," accessed 2022-04-25 from https://www.trade.gov/iot-and-smart-cities

- **Point products must contribute to a holistic solution**: Smart city "products" that are only core technology componentry (devices, sensors, and communication technologies) or point solution applications focusing on one facet of data collection and use only form a part of an objectively universal smart city program.
- **Data from traditional eGovernment systems must be included**: IoT technologies are not the sole source of data for smart city applications. Existing eGovernment systems comprising computers and communication networks supporting government operations (such as zoning, permitting, food inspection, transportation, sanitation management, event management, parks and recreation, construction, or emergency services) are primary sources of data that must be included as part of the reporting, analysis, and business intelligence supporting decision-making.
- **The necessary information management infrastructure must be established**: Strategic decisions are informed by and consequently affect different areas of municipal management. A smart city's analyses will rely on data collected, curated, and analyzed horizontally across operational domains, and must be supported by the appropriate data management infrastructure.
- **A wider net must be cast for valuable data**: To best leverage analytics, a successful smart city initiative must not be limited to the data streaming from IoT devices and the data produced as a byproduct of operational systems. Truly smart cities must cast a wider net to include and archive data sourced from external sources such as other public data sources, weather, news, third-party organizations such as ride-share and bike-share companies, non-governmental organizations, and other data aggregators/publishers.
- **Integrate analytics into operational processes**: Data collection is just the beginning of the processes that make cities and communities smart. The ability to inspire continuous operational improvement is powered through integrated descriptive, predictive, and prescriptive analytics that engage stakeholders and influence behaviors.

The above principles reinforce what should have been apparent all along: point products for IoT-based applications are just one facet of what comprises a holistic smart cities solution. Despite the existence of these point solutions, data produced by all applications serving the municipality must be collected, archived, aggregated, and potentially enhanced to provide the information and actionable knowledge that improve operations, inform strategic planning, and expose new opportunities for improving the citizen experience.

Examples: Analytics for Data-Driven Decision-Making

The best way to illustrate these principles is by example. Each of the following types of projects, most of which are projects that are eligible to be funded under the Bipartisan Infrastructure Law (Infrastructure Investment and Jobs Act)[3] relies on data sourced from a variety of eGovernment systems using both traditional and emerging information technologies. More importantly, the desired outcomes of these types of projects depend on a holistic information infrastructure integrating data from multiple sources and facilitating analytics to drive informed business decisions:

- **Improving public transportation**: Federal resources are made available by the Urbanized Area Formula Funding program for transit capital and transportation-related planning[4] and the Accelerating Innovative Mobility program to support innovation in the transit industry[5], both of which can increase travel alternatives for individuals without access to a car. In addition to real-time traffic camera streams and streetlight data, these programs benefit from the collection and analysis of transit-related data, data published by bike-share program operators, rideshare data, parking data, street sensors, and pedestrian data to improve multimodal travel alternatives, improve public transportation, and even support first-mile/last-mile options for underserved populations.
- **Water management**: One significant challenge of water quality management is that water does not respect municipal boundaries. Rainwater falling in one

3 The text of the law can be viewed at https://www.congress.gov/bill/117th-congress/house-bill/3684/text

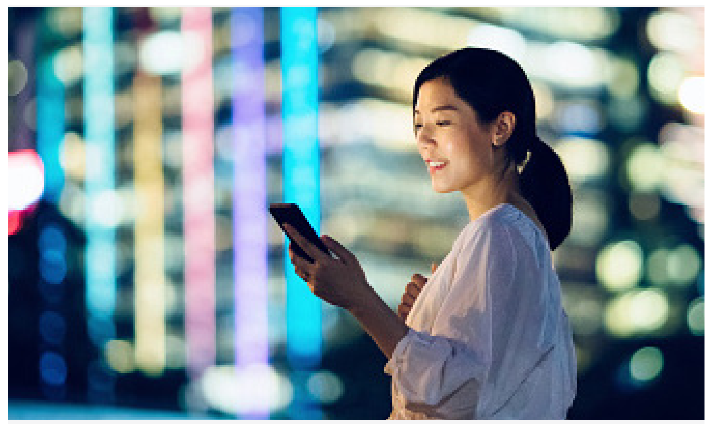
4 See https://www.transit.dot.gov/funding/grants/urbanized-area-formula-grants-5307

5 See https://www.transit.dot.gov/AIM

town will eventually stream through several different jurisdictions, requiring greater inter-jurisdictional cooperation. Of course, in many areas, there are water utilities there are continually monitoring water quality. Yet these static measurements could be improved significantly when integrated with data generated by IoT sensors embedded in water outfalls, road data, historical weather data, as well as residential and commercial water usage. These same data sets can be combined to solve problems with sewer overflow and roadway flooding.

- **Public safety**: Integration of traffic, transit, bicycle, and pedestrian data can inform road traffic safety programs such as those employing the guidelines of Vision Zero advocating for a smart system approach to reducing traffic fatalities. Integration of real-time traffic data, hospital/emergency room (ER) utilization, construction and accurate geolocation data, coupled with traffic signal preemption, can clear the roadways for first responder vehicles and speed emergency services arriving at the right locations.
- **Asset management**: Integrated data can inform the management of municipal public goods such as parks and recreation facilities. While sensors can capture the status of equipment and pathways within a park, this data can be combined with historical maintenance data to determine if (and when) assets could be scheduled for replacement.
- **Commerce and logistics**: The Covid-19 pandemic has put a magnifying glass on the challenges of efficient and productive supply chains. The corresponding increased reliance on eCommerce has created new pressures in dense urban areas; as the number and volume of parcel deliveries increase, so do urban congestion as well as greenhouse gas emissions. This poses an interesting, smart city opportunity in a public-private partnership, in which municipal traffic data, energy consumption data, and environmental measures can be integrated with logistics/supply chain data shared by private organizations to help with siting urban consolidation centers, managing parcel lockers, defining sliding time frames for deliveries, as well as ensure that underserved areas are amply provided with "last-mile" delivery. Fortunately, the Infrastructure bill's Strengthening Mobility and Revolutionizing Transportation (SMART) Grants can be used to fund projects demonstrating how smart city techniques and data support commerce delivery and logistics[6].

To more fully realize our stated end goal of the smart city, improving citizens' lives, in each of the project types above, holistic integrated end-point and municipal data sources and systems must come together.



The complexity of designing comprehensive smart city applications to improve citizen lives is magnified when smart city application designers adopt a collection of point solutions, each of which addressing a single set of issues.

Challenges

Each of these examples share one common theme: they require a data lifecycle process that employs data shared from multiple sources and feeds intelligent analytics intended to inform decision-makers grappling with challenging problems. But that means that a more sophisticated approach to data management and analytics needs to be taken to empower the full potential of a smart city initiative. That requires recognizing,

6 "Building a Better America," Bipartisan Infrastructure Law Guidebook, accessed 2022-05-06 via https://www.whitehouse.gov/wp-content/uploads/2022/01/BUILDING-A-BETTER-AMERICA_FINAL.pdf

acknowledging, and overcoming some common challenges that smart city architects may face. Contrary to expectations, these challenges are magnified when smart city application designers adopt a collection of point solutions, especially when these products only address a single facet of a more comprehensive set of issues.

It is critical to acknowledge these challenges and understand their implications when designing applications intended to inform municipal stakeholders. The presumption that a point solution can automate decision-making is flawed, especially since good strategic decisions are made by people who interpret analytical results of data sets that are accumulated and integrated from a wide variety of sources within the contexts of their own knowledge and experience. Empowering these stakeholders and decision-makers requires awareness of available data sources and accessibility and visibility to those data sets. A holistic smart cities information environment benefits from considering ways to overcome the following common issues, such as:

- **Establishing Data Awareness and Data Literacy**: While different groups, divisions, or agencies within a municipality may have engaged partners and vendors to install IoT sensors and devices, the individuals may not be aware of how the devices generate data. Moreover how that data is stored, shared, or managed, or even how that data can be shared with other agencies or units. Data awareness is a framework for publishing the details about available data sets. Purposeful smart cities that improve citizens' lives must be driven by data literacy. Data literacy describes the establishment of the capabilities to collect, curate, catalog, and manage growing volumes of machine-generated sensor/device data, semi-structured and unstructured human-generated data operational data from existing government systems.
- **Instituting an infrastructure for data integration**: As suggested, the benefits of a holistic smart city deployment are amplified when data from multiple sources (including system data, device data, and human-generated data) can be accumulated, combined, and organized for downstream data consumption. Consider designing and implementing a data platform for data organization and management along with the right set of data ingestion, integration, and preparation tools that enable the integration of data sets together to facilitate reporting, business intelligence, and more advanced analytics.
- **Improving analytics competency**: With the growth of reporting, visualization, and more advanced machine learning and artificial intelligence tools, it has never been easier for a wide range of data consumers and decision-makers to employ analytics to improve ongoing operations (such as real-time traffic light management) and better inform overall municipal strategy (in terms of transit investment, ensuring equitable treatment, or improving the power grid). Over time smart city planners will explore ways of establishing a data analytics lifecycle to create opportunities for improving the citizen experience.
- **Increasing computational efficiency**: As the number of sensors, devices, cameras, and other input devices increases, there is a predictable explosion in accumulated data. This data will be of minimal value if there is no plan for a technical architecture for ingesting real-time data streams, managing and analyzing massive data volumes, and integrating automated decision models into operational systems. Develop a computing framework that is robust enough to automatically scale according to the business data consumption demands.

Data streamed from a myriad of sensors, cameras, and other devices will be of minimal value if there is no plan for a technical architecture for ingesting real-time data streams, managing and analyzing massive data volumes, and integrating automated decision models into operational systems.

A Technology Blueprint Facilitating Smart City Analytics

Smart city application designers should develop a technology blueprint that accommodates the ability to accumulate data, integrate that data, and provide data accessibility to the downstream consumers. Figure 1 shows that the different point solutions are part of a system-wide network that streams data from smart devices to a centralized computing environment and, in turn, can communicate with all of the solutions.

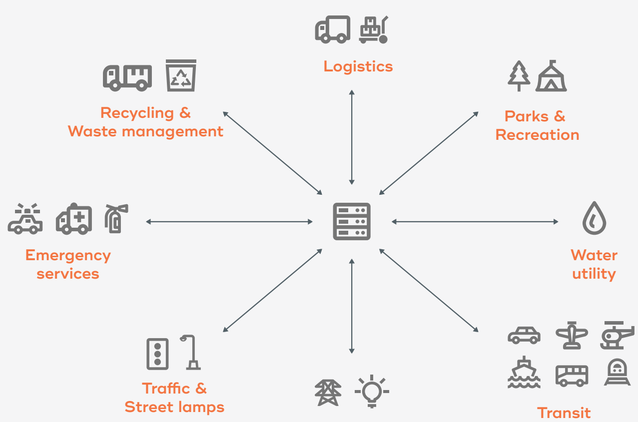


Figure 1. Logical interconnectivity among all smart applications/ solutions and a data management environment.

As an example, recall our earlier discussion of a smart cities project to improve commerce and logistics. This application must have traffic, transit, and logistic data shared among the different point solutions that can inform (and optimize) commerce delivery scheduling. That requires adopting some key components of a data technology platform addressing these challenges and thereby powering a smart cities initiative.

Figure 2 shows a high-level architecture diagram enumerating components for the data analytics lifecycle that supports the utilization of data sourced from across the existing governmental systems along with the incorporation of data streaming from IoT devices, end-point solutions, and edge computing systems. This comprehensive architecture fills in the gaps, to create a unified mesh of the mess left behind by point IoT solutions. This mesh is accomplished by enabling the end-to-end processes of the data analytics lifecycle. High-level steps in that lifecycle include tasks such as:

- **Data ingestion**, facilitating the acquisition and onboarding of data from multiple disparate systems.
- **Metadata assessment and management**, capturing and documenting the structural and lineage information about acquired data sets.
- **Data curation**, using the metadata to supplement the data catalog and raise data awareness.
- **Data organization**, providing models for organizing and managing integrated data sets as a prelude to reporting and analysis.
- **Data engineering**, enabling developers to apply standardizations and transformations to better facilitate data utility.
- **Data access**, allowing for data consumers to use self-service access methods to bypass the information technology bottleneck for data availability.
- **Analytics**, including basic descriptive analytics, as well as more sophisticated libraries and utilities for predictive and prescriptive modeling.

This data platform ingests and curates the data (including captured metadata) and flows managed data to a variety of analytics capabilities and services such as descriptive analytics, data visualizations, location intelligence/geospatial analysis, as well as more sophisticated machine learning and AI services.

Considerations

Cities that use point solutions providing advanced connectivity and communication comprising of emerging state-of-the-art IoT sensors and devices can certainly add value. However, while point smart city solutions will show results, if you don't have an overall plan for integrating information from those point solutions, you won't be able to benefit holistically to make impactful decisions using the outcomes for tactical and strategic planning. As such smart city application designers will miss out on the ability to

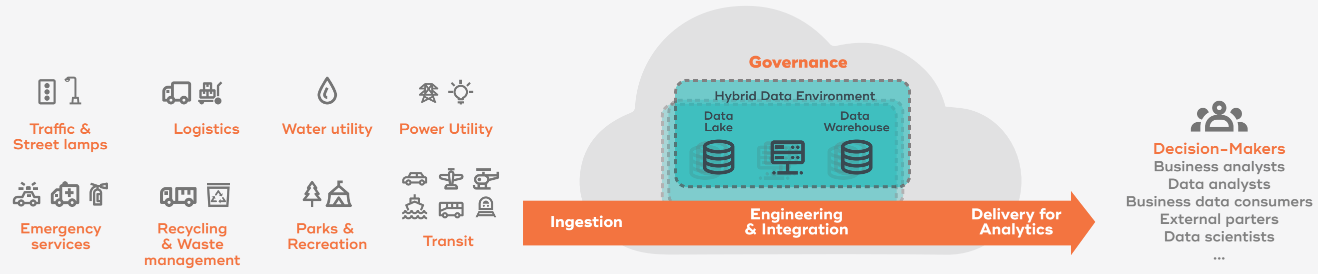


Figure 2. High-level architecture for smart city data integration and analytics enablement.

drive greater improvement in citizens' lives. It is critical to distinguish between point solutions and a horizontal platform strategy that integrates point solutions with various inputs to provide analytical insights that drive data-driven decision-making.

The promise of smart cities to improve the quality of life for residents requires the navigation of best practices associated with how information is created through the emergent innovative physical infrastructure. At the same time, both facets must be managed within the sociotechnical context of overall improvements to the citizen experience.

At the University of Maryland, that recognition inspired the creation of a smart-cities dual master's degree program to fill the need. Students would earn both a Master's degree in Community Planning and a Master of Information Management degree[7]. It is through the dovetailing of the unique strengths of faculty in the School of Architecture, Planning, and Preservation and in the College of Information Science that students can assimilate knowledge and skills in data science, eGovernment, and the Internet of Things and blend that with knowledge of physical, social and economic community planning.

A successful smart city application designer must adopt a platform architecture supporting different use cases through the ingestion, management, and analysis of large amounts of data (from municipal as well as, point solution, and external sources) to create embedded analytical models. Together this holistic data environment can allow for instantaneous reactions within operational environments (e.g., real-time traffic applications or pedestrian safety) and support strategic decision-making for long-term improvements to the citizen experience.

Therefore, when designing solutions to empower the smart city, consider the holistic need for information integration in a way that incorporates data from a variety of traditional government lines of business along with innovative data streams, and third-party sources. Plan to incorporate a data warehousing and analytics environment that supports the full data analytics lifecycle. Make sure that your reporting, business intelligence, and analytics architecture embraces the integration of data from among a broad variety of sources ranging from traditional municipal and eGovernment systems, smart city systems that employ emerging IoT technologies, and third-party data. Finally, integrate descriptive, predictive, and prescriptive analytics to inform decision-making and influence positive activities to improve the lives of citizens.

7 https://www.umdsmartgrowth.org/city/smart-cities-dual-degree-program/

David Loshin, president of Knowledge Integrity, Inc., (www.knowledge-integrity.com), is a recognized thought leader and expert consultant in the areas of data management and business intelligence. David is a prolific author regarding information management best practices as the author of numerous books and papers, including Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph and The Practitioner's Guide to Data Quality Improvement. David is a frequent invited speaker at conferences, web seminars, and sponsored websites and channels. David is also the Program Director for the Master of Information Management program at the University of Maryland's College of Information Studies.

David can be reached through LinkedIn, or email via loshin@knowledge-integrity.com.