VISUALIZATION-DRIVEN GOVERNANCE: 
THE VISUAL PRACTICES IN TRANSIT SYSTEM

Thesis Presented
by
Liuhuaying Yang
to
The Department of Art + Design

In Partial Fulfillment of the Requirements for the Degree of
Master of Fine Arts in Information Design and Visualization

Northeastern University
Boston, Massachusetts
May, 2018
Visualization-Driven Governance:
the Visual Practices in Transit System

by Liuhuaying Yang
to The College of Arts, Media and Design

In Partial Fulfillment of the Requirements for the Degree of
Master of Fine Arts in Information Design and Visualization

Advisor Dietmar Offenhuber
Readers Laura Riegel and Paul Kahn
Chair John Kane
Respondent Kristian Kloeckl
Member Dietmar Offenhuber

Northeastern University
Boston, Massachusetts
May, 2018
Abstract

Public transit systems are dynamic infrastructures and shared public spaces. Their management is challenged by their partial illegibility and a lack of efficient communication. My thesis follows the hypothesis that visualization can support the governance of such large complex organizational systems. Going beyond visualizing relevant transit data, it considers the totality of visual tools and practices used in their management activities. My case study is based on the visualization tools that I designed and developed at the MBTA (the Massachusetts Bay Transportation Authority).
Acknowledgements

This thesis would not have been possible without the help, support and encouragement of many people.

To Dietmar Offenhuber, my thesis advisor, for everything. I would not have the confidence to do my thesis nicely without his guides and support. And his genuine care and support for students encourage us to become better. To Nathan Felde, the professor of our program, for bringing me the perspective of the open system for my thesis. I was fortunate to benefit from his exceptional knowledge and insights. To Pedro Cruz, for spending his time helping me in thesis writing style. To Miso Kim, for opening my eyes to service design in class. This perspective not only reflected in my thesis but also affects my professional work. To Paul Kahn, one of my thesis reader, for reading and advice. To my committee members and every respondent in my defense, for compliment and feedback.

To my colleagues at the MBTA. Those projects would not have been done without their dedication, collaboration, and support To their detailed comments and materials as well as continuously contributions and support. Special thanks to Laura Riegel, who is also my thesis reader, for her thoughtful instructions, detailed feedback, and encouragement.

To my classmates and friends for critiques, support, and help. To my dear friends and family, for always accompany and support in my life.
Table of Contents

3  Abstract
4  Acknowledgements
5  Table of Contents
6  Introduction
8  Chapter 1: Theoretical Background
19  Chapter 2: Mapping in Historical Context
41  Chapter 3: Mapping a Living System
54  Chapter 4: Case Study
81  Chapter 5: Conclusion
86  Bibliography
Thomas P. Hughes identified an integrated transit system as a large technical system (2000). This large technical system not only consists of its physical artifacts but also of organizational and legislative artifacts, which indicates that the technical is socially structured (Hughes, 1987). He also clarified that a successful and well-tested system is not fully autonomous. People, businesses, organizations, as part of the system, are working to make it functional (Hughes, 1986).

Francois Caron identified a lack of proper communication between the departments as one of the challenge features of the railway system. The difficulty was rooted in the organization, not technology. Stephen Salsbury also suggested that there was a lack of proper communication between the departments of transit management. The electric telegraph had met the needs of transmitting information quickly between operators and dispatchers as well as dispatchers and trains since the 19th century. However, due to the growth of the system, in the early 20th century, the telegram was no longer enough. The railway system was in trouble again: they made more investment in technology but did not make significant changes in their management, especially on passenger services (Salsbury, 1988).

This leads me to think: how to react to the new challenges of system growth with innovative solutions. The public transit system is infrastructure-based, dynamic and real-time, invisible, public user based and space-shared, and these characteristics bring many challenges to governance. Visualization-driven governance provides a viable solution to these problems in terms of provision of timely informing for disruptive events during real-time operation, dynamic visibility for the overall system, and working efficiency, and eventually enhancement of the quality of service. The visualization-driven governance is not only about visualizing public transit relevant data. As the visualization researcher Munzner point out, “A tool that serves well for one task can be poorly suited for another, for exactly the same dataset.” (2014) My thesis is concerned with the totality of visual practices used in management activities that are intended to facilitate efficient communication for the service.
This thesis proposes a framework for visualization-driven governance. In this framework, the service system must be open and have a continuous exchange of information, what sometimes is called a “living system”, people perform their duties and exchange information in their own open subsystems to improve their work. This information includes both the physical infrastructure and the behavioral results of their own work or that of others, either from internal or external subsystems. Due to the diversity of information and roles, the communication through visualization tools is the most effective way to achieve this end, because it directly and effectively provides the legibility of information. When designing visualization tools, given the characteristics of public transportation management, the principles of accuracy, scenario-focus and temporality must be followed, and this specifically reflects on the design techniques of representation forms and interactions.

A case of the MBTA (the Massachusetts Bay Transportation Authority) is presented, including three projects that derive from the work that I developed while working at the MBTA, namely on the design and implementation of a Bus Traffic tool, a Green Line Vehicle Location Viewer, and an Origin-Destination-Transfer (ODX) Visualization tool. Each of tools addresses different issues in the overall system, including
- how on-site dispatchers use the Green Line Vehicle Location Viewer to ensure the Green Line’s reliability;
- how dispatchers use the Bus traffic tool to increase bus reliability;
- how planners can get information from the ODX tool to help them plan.

Not only do these tools show how visualization tools can be implemented in a public transit system and bring value to it, but also the totality of these implementations demonstrates how visualization-driven governance increases the capacity to communicate effectively and efficiently and thus improves the quality of service to customers.
This chapter conceptually defines visualization-driven governance. My use of the term visualization in this thesis refers to the field and practice of information visualization. By visualization-driven I mean not only visualizing public transit relevant data but the totality of visual practices used in management activities that intended to facilitate efficient communication and thus improves the quality of service. My thesis focus on the governance in a transportation agency who is offering transit services, including bus, subway, light rail train, and other kinds of public transit, excluding highways, bridges, airlines, ferry boats and other transportation modes. To further discuss the specific meaning and practical application of visualization-driven governance, this chapter also explains public transit from the perspective of service and system respectively, thus clearly defining the concepts involved in the follow-up. Finally, I will give a brief description of the fundamental characteristics of public transport services and the challenges they face, as well as these characteristics that determine that they will benefit from the visualization of governance.
What is visualization-driven governance

This thesis proposes a framework for visualization-driven governance, by which I refer to the activity of management informed by visual practices such as information visualization. Because of the widespread use of words such as visualization and information visualization, it is necessary to clarify my definition here.

First of all, information visualization here means the process of visually displaying the information provided by the data. Information visualization can also be data visualization, “a mapping between discrete data and a visual representation” (Manovich, 2011). I may cross-use these two terms later. But what I want to emphasize here is that data visualization emphasizes the process of visualizing “raw” data, and information visualization emphasizes the process of visualizing the information inferred from raw data. As a simple example, there is data of a vehicle number, the vehicle status, and the station name. Data visualization is more concerned about how to visualize this data, such as a vehicle’s icon is approaching a station icon. However, such data may contain more than one description of the information. It could be “Bus 123 is about to enter Ruggles station”, “Ruggles station is about to enter a bus with vehicle number 123”, “the bus to be approached is number 123, it will enter Ruggles”, or so on. When we focus on what information or description is provided during the design, it’s more accurate to call it information visualization.

Secondly, the totality refers to the visual context or visual environment formed through the integrated information visualization. When we state that we are using information visualization, we interact with a tool that provides visual information, namely, information visualization tool. We equip ourselves with the legibility provided by this tool when we use it. Once everyone is involved, we are in a working environment that has a visual environment that allows us to communicate visually.
This also illustrates the difference between visualization-driven and data-driven. Data-driven emphasizes the use of data as evidence to make decisions or help to form insights. The data can be raw data, structural data, analysis data, but not necessarily involves visualization. Visualization-driven is based on the premise of data, emphasizing the visual legibility, and the quality and accuracy of the data directly affect the reliability and validity of the visualization.

Finally, I want to clarify the difference between governance and management. Governance is concerned with “doing the right thing” and making strategic decisions, management is concerned with “doing things right” and following a prescribed process (Tricker, 1999). From this distinction, before knowing how to do things right, they have to know what the right thing is. On the one hand, governance is to ensure that the process, structure, and functions are put in the right place so that efficiency increases, expectations, and goals reached. On the other hand, as an urban transportation system, the right thing is not just about efficiency but also equity. Equity depends on value judgments of distribution of benefits and costs among all the stakeholders. (2010) Thus, governance is a multi-stakeholder process involving negotiations among users or customers, managers, employees, and the broader community. It, therefore, benefits from an open fluid system. Management, however, in order to do things right, needs to control day-to-day operation in a relatively closed system.
Transit services and management experience

The public transit services discussed in this thesis refer only to transit services, including bus, subway, light rail train, and other kinds of public transit, excluding highways, bridges, airlines, ferry boats and other transportation modes.

From a governance perspective, public transit is a system consisting of many subsystems (Figure 1.1). Based on the conceptual language I heard at MBTA, I divide subsystems into the physical infrastructure, the system of service structure, the system of operations, and the system of management. Of course, these subsystems may also include other aspects, such as finance, but this thesis focuses on the four subsystems listed earlier.

The infrastructure system is physical components of public transit, which includes terminals such as railway stations and bus stations, roads, railways, and vehicles. The service structure system is about how these infrastructures provide services in time and space, namely schedule. The operation system is about service execution, mainly to ensure that the scheduled service is operated as accurately as possible. The management system primarily on the planning of the service structure (which may involve the reconstruction of Infrastructure) and service execution. Governance, on the other hand, makes each of these subsystems behave efficiently and equitably to ensure public transit systems work.
Figure 1.1
Layers of transit subsystems
Another concept that needs to be clarified is service. Public transit service is to manage the public travel needs by planning its structure and infrastructure. If all the served people are called the public, and the people who use the service are called passengers, then from the service design point of view, the passenger experience includes three stages: pre-experience, during-experience, and post-experience. Pre-experience refers to the stage from when the passenger decides to travel until when he or she arrives at the station or stop. During-experience refers to the stage from when the passenger arrives at their origin station or stop until when they depart their destination station or stop. Post-experience is the stage after the passenger left the station to complete the journey. This “passenger experience” also called “passenger journey” under User Experience design discipline. But considering that this thesis is about transportation, transportation calls during-experience stage a “journey,” so here I phrase the passenger journey as “passenger experience.”

From a management point of view, the public transit service includes the stages of planning, release, operation, and evaluation (Figure 1.2). Planning, as mentioned earlier, includes the management of the structure and infrastructure, which determines whether the public will transform into a passenger, but also informs the passenger when and where to take what vehicles to complete their journey. However, passengers are able to plan for themselves because they have received the schedule information, which is achieved by a service release stage. Passenger’s during-experience also needs to be supported by an operation process to ensure the service is up and running. The running of a vehicle from the origin to the destination is called a trip. The operation is to ensure the trip is operated exactly as planned. An evaluation of the quality of the service is needed after the service is over, and the result of the assessment will affect the next planning phrase. This continuous process, corresponding to the passenger experience, is “management experience” (Figure 1.3).

The goal of governance is to improve the passenger experience by enhancing the management experience. Visualization-driven governance is a way of managing the visual communication context by visualizing the public transit system and taking advantage of it.
Figure 1.2
Map of the management experience and passenger experience

Figure 1.3
Fluid stages of the management experience
This thesis proposes a new approach for the governance of public transit services, namely visualization-driven governance. This approach enhances the quality of service by providing visualizations of service systems for increasing the likelihood and efficiency of communication. Before discussing this approach, I will first introduce the characteristics of public transit services that explain why it will benefit from the visualization-driven governance.

First, public transit is based on urban transportation infrastructure. It is different from IT infrastructure or software infrastructure. Software infrastructure provides support for service delivery by providing power, storage, networking, and more. Although big data collected and stored during this support can support precise service delivery and improve service quality, and it may also be a service product for trading directly, the infrastructure is not the product service itself. However, public transit serves through direct interaction with people through the infrastructure as vehicles and stations. This direct interaction is not only based on the deployment of the physical facilities but also the operations. It also creates customer’s experiences and meets their needs.

This infrastructure-based interaction is dynamic and real-time. Not only does it need to be designed and deployed like a park or landscape, but it also requires constant involvement in service. After the deployment of the public transit system, it requires continuous real-time service support during each service lifecycle, such as ensuring punctual departure and arrival, to ensure the completion of services and the quality of service. However, when people are playing in a park, no park facility provider continues to operate the spot and people are already able to enjoy, and the park is sufficient to realize its service value.
Furthermore, the whole picture of the public transit system is invisible. On the one hand, public transit systems include not only terminals, vehicles and tracks that have visible physical characteristics but also transit routes and time schedules that are not visible in time and space. On the whole, it is a network that extends to the entire city or region, making it partially invisible. Even though agency’s currently available services cover some specific areas only, those areas where no service is available are also under the agency’s planning. For example, MBTA is needed to meet the needs of the Massachusetts Bay region, and even if the public transport service is not available in the area of Dover or Medfield, these areas still need to be monitored to consider the potential expansion of service to meet the future needs of those areas.

Besides, the public nature of public transportation service does not have a specific service user base but requires servicing all city dwellers, such as workers, students, or retirees. It also provides specialized services for populations who require specialized care such as wheelchair access; the served groups need to be as comprehensive as possible.

Finally, it involves more than just the transport service itself, after all, it uses urban space in conjunction with other vehicles and facilities. Because of this, its planning is indeed not arbitrary. It not only needs to consider the coverage within its services but also the service feasibility under the objective conditions. For instance, whether the road permits the construction of new tracks, whether the land allows the new stations, and how to cooperate with other facilities as parking lots, new business districts, land owners and so on. This is also true for operations, for instance, by considering how to ensure service reliability during peak traffic times and how to deal with the impact of traffic conditions and congestion caused by other vehicles.
Challenges of governance

The characteristics of public service transportation bring many challenges to governance. Visualization-driven governance provides a new perspective to these problems by provision of timely informing of disruptive events during real-time operation, dynamic visibility for the overall system, and working efficiency.

Real-time service requires managers to handle the risk of many disruptions in a timely manner, which poses challenges in providing timely information. The disruption could be an unplanned or unanticipated event with the potential for impacting service performance. In the risk management process, the manager not only needs to identify the potential transportation disruption and evaluate the impact, but also to come up with a strategy, apply it, and review the result (Coyle, Novack, Gibson, & Bardi, 2006). For example, an on-site dispatcher needs to ensure that the vehicle will be on schedule at this station, for which he has to determine if the next vehicle will arrive on time and adjust for possible delays. He also needs to know the information of other vehicles on the current line to avoid unnecessary conflicts when he quickly develops a strategy for adjustment. After implementing the adjustment strategy, he needs to see if the modification is effective. The more quickly he can receive all relevant vehicle information in real time, the more quickly he can implement reliable management actions.

As all the service lines are running at the same time, the entire service system is a dynamic overall. The invisible part of the system is also essential, and we need a way to express it as a whole. Besides, because of the characteristics of the public user base and shared space, access to information outside the internal transit data is equally important. On the one hand, this external information affects the service’s needs, such as total population, population distribution, geography, income level, and government policy. On the other hand, the external information may affect the quality of service, such as road congestion (Victor & Ponnuswamy, 2012). This additional information adds complexity to the presentation of information. We have the relevant information and data is the foundation, because it provides the possibility
of system readability, but it is not enough to solve the problem of the expression of complexity, namely legibility. Visualization offers this legibility. “Visualization, embedded in the urban context, can make infrastructures more legible by exposing the structure and activity of the system. The paradigm of the invisible, seamless infrastructure, however convenient it might be, is no longer applicable when infrastructures require conscious decisions from their users.” (Offenhuber & Schechtner, 2012)

Traditionally, these agencies do the same queries or reports whenever they need information due to the dynamic changes of data during internal information exchange. The on-site dispatchers repeatedly inquire about the vehicle status through the radio in real-time operation at the MBTA. Because the diversified information is usually held by different departments, the accuracy and uniformity of the information delivered can hardly be supervised. They need an automated tool that provides readability to reduce unnecessary duplication of effort to improve accuracy and efficiency. In this case, I propose a visualization tool, and the use of it is the visualization-driven governance.

Finally, of course, not all transportation agencies will need to implement visualization-driven governance. For small businesses or small teams that operate only a few lines, the cost of developing and maintaining visualization-driven governance can be far greater than the traditional approach. But the value of visual governance itself still exists. For instance, the standardization of schedule data, like GTFS (the General Transit Feed Specification), allows larger transit agencies to develop tools and share them with smaller agencies as long as they are using the same data standard.
Andrew Vande Moere and Helen Purchase propose a model of three potential roles for design in information visualization: visualization studies, visualization practice, and visualization exploration. The three roles reflect the three worlds and the stakeholders’ activities. In the academic research world, the researchers use visualization to understand, communicate and predict the explicit academic questions and knowledges. In business world, the commercial enterprises do visualization for profit, which is “Visualization practice”. In design world, the designers and artists are focused on exploration of design solutions. (2011)

The totality of visual practice in a transit agency may involve these three in practice. Experts working inside need to obtain insights, analysis or planning through data visualization. As a service provider, delivering service information to consumers is also part of their service. As a public service department, it also involves responsibility for popularizing knowledge to the general public. Although there may not be direct design exploration, designers and artists outside of the agency have had a profound impact on their visualization technology. Based on this, I divided the totality of visual practice from the roles in legibility into three: legibility for experts, for the customer, and for the public.
Mapping for management

Etienne-Jules Marey visualized train schedule for Paris to Lyon in 1880s (Figure 2.1). In this visualization, the x-axis is time and the y-axis is location or stop in this context. Each line across the grid represents an entire trip of the train. The slope of the lines represents the speed and the direction of the train. The intersecting points represent where the trains will meet or pass. It shows the potential conflicts in scheduling. This visualization is also called “the time-distance diagram” or “the string line diagram”.

“The time-distance diagram” is applied in modern transit management projects for tracking and planning. The U.S. Department of Transportation (DOT) released a report that introduced a tool for planning and monitoring railroad traffic (Lee & Multer, 2008) (Figure 2.2). The tool retrieves the real-time data to track the vehicle’s performance based on GTFS (The General Transit Feed Specification) data. The real-time data are collected from GPS sensors which are installed on the vehicles.

Stewart et al. (2015) demonstrated the potential of transit data visualization by generating and visualizing the performance measures at four different scales: system level, neighborhood level, route level, and stop level. They used heat maps extensively to depict the data on a system level. In the three visualizations at the system level (Figure 2.3), the white dots represent the stops, and the white lines represent the routes. They all use the heat map as the analysis technique and color code the geographic contours. “Transit stops of the overlapping coverage” presents the overlapping coverage of transit stops. The color from dark green to red indicates the lower coverage to the higher one. “The number of buses in the AM peak” presents the number of buses in the AM peak. The color from dark green to red indicates the fewer buses to more buses. “The average bus-stop waiting time” presents the average waiting time at the bus stop. The color from dark green to red indicates the shorter waiting time to longer time.
Figure 2.1
Etienne-Jules Marey
Paris-Lyon Train Schedule
1885

Figure 2.2
Mary T. Lee and Jordan Multer
Visualizing Railroad Operations: A Tool for Planning and Monitoring Railroad Traffic
2008
Figure 2.3
Colin Stewart, Ehab Diab, Robert Bertini, and Ahmed El-Geneidy
Visualizing Performance Measures At System Level: (a) transit stops of the overlapping coverage; (b) the number of buses in the AM peak; (c) the average bus-stop waiting time
2015
3D visualizations have also been explored in transportation management in two ways. One is to build a completely virtual 3D space. Matti Pouke et al. (2016) explored whether the utilization of 3D objects brings added value in public transportation visualization systems. In this project, building, forest, and bus are all 3D objects, and the users are allowed to interact with these objects (Figure 2.4). The other is to create 3D data visualizations. Somdulyawat et al. (2015) built a visualization tool for the mobility and ridership of the city. It uses the geometric shape to represent the different area of the city. The color and the height of that shape indicates the density of the riders. Stewart et al. (2015) also use 3D to present the performance in Route and Segment Level, where the height of the cube represents the percentage of seats used at each stop, and the color of the cube from green to red indicates the higher seating capacity to lower one.
However, after-the-fact analysis can only meet part of the demand, and it still lacks support for time-sensitive operations. Moreover, the statistical data visualization is reduced, and a lot of information is selectively missing, and it is difficult to provide the comprehensive legibility of the system.

In the market, professional transit analysis software is also available. For instance, “Remix” is a scenario planning tool targeting transit agencies worldwide. It provides the information about the accessibility, demographics, and cost. By dragging “Jane” (a virtual figure on the map who represents the isopleth) onto the map, users can see how far Jane can go right away on an isochronic map (Figure 2.5). By integrating demographic data layers such as population, jobs, minority, senior, it enables the users to see who will be affected by the alternative plan through a choropleth map (Figure 2.6). These two maps are both color-coded by the group. The shades of purple indicate the different levels of poverty.
Figure 2.5
Remix co.
A Planning Platform for Public Transit: Jane and Isochronic Map
2015

Figure 2.6
Remix co.
A Planning Platform For Public Transit: Choropleth Map
2015
But in my experience, such tool do not quite fit the needs of a large transportation agency. One significant reason is the complexity of services. I had a conversation with a manager who leads many visualization projects at MBTA about the software as “Remix.” She pointed out that there are many extreme cases in a large system which requires customized analysis. “Remix” has to simplify the analysis to handle general needs as their business strategy. This is why the MBTA prefers building their own tools over using standard analysis software.

Nowadays, the transportation agency also began to open up its data to seek exchanges and cooperation with outside scholars and students. For instance, MTA (Metropolitan Transportation Authority in NYC) with Taxi and Limousine Commission, NYNJ Port Authority, Waze, and MoDA held an “Open Data L-Train Innovation Challenge” in March 2018. It joined the public and facilitated collaborative teams to tackle one of several challenges facing the city as part of the L-train shutdown scheduled for April 2019. In March of the same year, MBTA launched “MBTA Systemwide Passenger Survey Data Challenge” to see how students and researchers can creatively use their passenger survey data and participate in answering research questions.
The legibility for passengers begins with the need for travel information. Francis Galton created “Isochronic Passage Chart for Travellers” in 1881 (Figure 2.7). It shows the travel times from London to different parts of the world in days. Metropolitan town planning commission created “Map of Melbourne and environs minimum railway or tramway time zones” in 1926 (Figure 2.8). These two isochronic maps both reveal the time contours of travel.
Figure 2.7
Francis Galton
Proceedings of the Royal Geographical Society 1881
Figure 2.8
Metropolitan town planning commission
Map Of Melbourne And Environs Minimum Railway or Tramway Time Zones
1926
Inspired by an “electrical schematic” diagram, Harry Beck redesigned Travel for London metro map, which is called “Tube map” in 1931 (Figure 2.9, 2.10). Beck described his conception thus:

“Looking at the old map of the Underground railways, it occurred to me that it might be possible to tidy it up by straightening the lines, experimenting with diagonals and evening out the distance between stations. The more I thought about it, the more convinced I became that the idea was worth trying, so selecting the Central London Railway as my horizontal base line I made a rough sketch. I tried to imagine that I was using a convex lens or mirror, so as to present the central area on a larger scale. This, I thought, would give a needed clarity to interchange information.”

Figure 2.9
Harry Beck
Original sketch for Underground diagram
1931

Figure 2.10
Harry Beck
Underground Diagram
1931
Navigation software now provides passengers with service information by providing visual maps. Google Maps is a web mapping service that provides users with a public transportation travel plans. Not only that, but Google has also developed the “Transit Partner Program” to attract transportation agencies to add their transit data to Google Maps. Currently, over 18,000 cities worldwide make their information available in Google Maps. Just as Google’s website quotes:

“Google Maps for Transit is a truly innovative marriage of information and infrastructure. It is a perfect example of how the public and private sectors can partner together to benefit us all -- and it didn’t cost New York taxpayers a penny. I applaud my colleagues at the MTA and Port Authority for making this a priority, and our friends at Google for continuing to make the world an easier place to navigate.”

— New York Governor David A. Paterson

“Transit” has become a favorite navigation application for transit passengers in many cities. It provides transit plans for passengers through data opened by the transit agency. The current service covers cities in 12 countries including United States, Canada, France, United Kingdom and manage 350 transit feeds in 125 cities. In the Boston area, it not only partnered with MBTA and regional transit systems (such as the Worcester Regional Transit Authority) but also with Hubway, Uber and other sharing systems. In the App’s design, they not only visualized the travel plan in the time dimension (Figure 2.11) but also spent a lot of effort on the transit map. In general, maps automatically generated from GTFS data have a problem. Since the path of a route is generated from the actual trajectory data of the vehicle, even if the two roads cross, it does not mean that there are stations available for transfer. In 2016, they solved the intersection problem of automatically-generated transit maps through algorithms, which visually improved the accuracy of information representation (Figure 2.12).

“Our diagrammatic transit maps provide more than just the basic information about transit systems. They're emblematic of cities themselves: important pieces of functional art that connect people to their environments. We want to help build that connection, and we believe that our new Transit Maps do just that.”

— From their Medium post “A Technical Follow-Up: How We Built the World’s Prettiest Auto-Generated Transit Maps” by Anton Dubrau (2016)
Figure 2.11
Transit screenshot:
It visually shows time-consumption for each transit plan.

Figure 2.12
Transit screenshot:
Visually, the Orange Line is parallel to the Commuter Rail (purple line), not overlapping, and they have a transfer stop (a white rounded square). The Green Line crosses the Commuter Rail but does not have transfer points.
“Waze” is a community-based traffic and navigation app for drivers. Users can visually see the real-time traffic and road information that shared by other drivers in their area (Figure 2.13). The App is both their service terminal and data collection source. In addition to the user can report traffic information on their own, when the user opens the App to query information, the App also collects their driving status, which in turn becomes the basis of the service. Besides, Waze also provides transport SDK to other service areas. I used their private API in MBTA project.

In addition to navigation, visualization has also become part of the service experience. Uber is a pioneer among ridesharing companies with operations in 633 cities worldwide. In design, Uber visualizes availability and needs. Passengers can see on the Rider App if there is an empty car near them. The driver can see on the Driver App through a heat map which area in the city has a higher demand for cars (Figure 2.14). As a result, passenger groups and driver groups realize visual communication between supply and demand. And after demand matching, passengers and drivers can see the location of each other on the map in real time, which effectively solves the uncertainty in the peer-to-peer service by providing visual information.
Figure 2.13

**Waze screenshot:**
It visually shows where other drivers (cartoon character icons) are in your area, and construction alerts (prohibited line icons).

Figure 2.14

**Uber driver screenshot:**
the heat map showing customer demand

*Photo from uberpeople.net*
Visualization for the general public can be called casual visualization. Casual visualization supports a larger user population in both everyday work and non-work situations (Pousman et al., 2007).

Casual visualization accumulates knowledge for the general public. For instance, “Touching Bus Rides” (MIT Senseable City Labs 2012) provides two types of visualization to represent the volume of the passengers at that particular station (Figure 2.15). One is a map layout that emphasizes the location-based comparison. The location of the circles indicates the location of the station. Another is a grid layout and emphasizes the time-based comparison. The y-axis is time, and the x-axis is the stops of the selected bus by order. In both modes, the size of the circle indicates the number of the rides related to that stop. The green circle represents how many people board at this stop during that time period, and the yellow circle is how many people get off at that stop during the same time period.

Although the knowledge provided by this project is accurate, it is not sufficient for professional management. For example, in the professional management analysis, the station is hierarchical, and some stations are child stations of a parent station. The difference between the analysis of parent station and the non-parent station is significant. For another example, real planning scenarios heavily rely on zone analysis and other related urban contexts.
Figure 2.15
MIT Senseable City Labs
Touching Bus Rides:
(a) map layout; (b) grid layout
2012
However, casual visualization creates workable solutions for design. “Traffic in Lisbon” (Cruz, Machado & Bicker, 2010) and “Lisbon’s Blood Vessels” (Cruz & Cruz, 2013) explore the metaphors of blood vessels. As Cruz explains, “a living organism with circulatory problems.” In “Traffic in Lisbon: Arteries” (Figure 2.16), the color of arteries indicates the speed of the vehicles with cool colors for rapid transit arteries, hotter colors for sluggish ones. The thickness of the arteries represents the traffic intensity, and the white dots represent the vehicles. “Traffic in Lisbon: Clots” depicts the slow traffic, in where the red clots appear when traffic is slow. In “Lisbon’s Blood Vessels” (Figure 2.17), the color of vessels indicates the speed of the traffic, brighter red represents higher speed, and darker red represents lower one. The thickness of the vessels represents the volume of vehicles, and lengths are regenerated according to their velocities based on an assumption that slow traffic perceptually implies longer distances and vice versa. It also applied animation to distort of the representation, which he called “spring model” (Cruz et al., 2015), to emphasize the emotional provocation.
Figure 2.16
Pedro Miguel Cruz, Penousal Machado and João Bicker
Traffic in Lisbon: Arteries
2010

Figure 2.17
Pedro Miguel Cruz, António Cruz
Lisbon’s Blood Vessels
2013
Another example is “Cascade on Wheels” (Thirion, 2007). It is a sonification and visualization that presents the vehicle volume by 3D vertical columns emerging from the streets map, like walls (Figure 2.18).

In conclusion, visualization-driven governance is inseparable from the visualization technology, but it is not just the visualization of transit data. Casual visualization also visualizes traffic data, but it is distinctly different from the visualization tools used for management in the task focus and accurate representation. Moreover, whether it is a statistical analysis chart or a network map, it is just one corner of the practice. Visualization-driven governance is not just the use of visualization in one place but the entire system.
This chapter focuses on the design framework for the visualization-driven governance. As previously mentioned, visualization-driven governance provides timely information for disruptive events during real-time operation, dynamic visibility for the overall system, and working efficiency. To do this, such service systems must be open to allow the continual exchange of information, namely a living system. In this life system, people perform their duties and exchange information in their open subsystems to improve their work. This information includes both the physical infrastructure and the behavioral results of their own work or that of others, either from internal or external subsystems. Due to the diversity of information and roles, the communication of visual language through visualization tools is the most efficient way to achieve this end, because it directly and efficiently improves the legibility of information. When designing visualization tools, given the characteristics of public transit management, the principles of accuracy, task focus and temporality must be followed, and this reflects explicitly on the design techniques of representation forms and interactions.
Living system

The concept of the living system comes from the theory put forward by Miller (1978). He defined it as a system that exists in space and is made of matter and energy organized by information. Meanwhile, he identified the essential concepts of space, time, matter, energy, and information in the living system theory. I'm not going to analyze these concepts that correspond to the public transportation system. Although it is possible to take the time to do it, it has little to do with what I want to discuss but adds complexity and creates confusion. By saying transit service systems must be a living system, I just want to emphasize its properties of openness (permeability boundaries) and capability for conversation.

The living system is open to conversations and provides the basis for different stages of the process. From a management experience perspective, public transportation services are a process of planning, delivery, operation, and evaluation. In this process, different roles perform their duties. Service planners design the structure and processes of the service and then deliver it to the public. The dispatchers and operators implement the service plan to ensure that passengers have reliable and on-time travel services. The management personnel and service planners then evaluate the service plan according to the operation status and make improvements to the subsequent service. Whether administrators and planners need to improve services depends on both external market pressures and the actual operational data of the service. Whether or not a service plan is executed accurately is a prerequisite for using operational data for analyzing improvements to the plan.

The boundary of the entire service system is permeable, and the essence of the penetration action is the interactive and evolving conversation. For instance, the criterion for assessing quality is a simplification (Offenhuber, 2017) of service standards and measurement methods, both of which can be revised. The same service output will lead to different conclusions under different standards. Even under the same rule, using different measurement methods will lead to different outcomes. So essentially, the definition of
these standards and measurements is the result of the conversation between passengers and the agency. Let me make a cat-mouse analogy (Dubberly Design Office, 2002) here (Figure 3.1). The agency is not a non-profit organization and still needs to consider the ride rate. In order to send a car to pick up passengers, the agency needs to observe people’s travel behavior and gradually adjust the service plan accordingly. As a passenger, he is also planning and modifying his travel strategy according to the service plan.

Figure 3.1
Cat-mouse analogy of transit services
The service operation is also based on a series of conversations on the exchange of information. The dispatcher at the platform waits for the vehicle on the line to enter the station and sends the vehicle out on time. Each arrival and departure require a dialogue between the dispatcher and the system. However, the delay is overdetermined. Delays in starting a train at the previous stop, stoppages in a station, delays or breakdowns in previous orders, road congestion, etc. may all be the causes. In this dialogue mode, the dispatcher not only needs to ask about the current location of the vehicle that is about to arrive but also needs to care about other relevant vehicles and road information.

For the dispatcher, the process of inputting from his own operation to the output of the result is a black box: my on-time departure does not guarantee that the car will arrive on time; my on-time departure may cause the next stop to be more congested; my delayed departure may ease the congestion at the next stop. Thus, he needs to get feedback on the results of the operation through dialogues so that he can learn and improve scheduling skills in so-called constant experiments.

I don’t entirely enumerate such evidence, but it is enough to see that the living system is the essential environmental condition for public transit service governance. Without this condition, managers cannot establish a dialogue, and it is difficult to achieve information exchange and obtain feedback.
Information exchange is the meaning of dialogue, but it is unrealistic or costly to acquire and transmit information through the traditional dialogue between individuals. The public transit service system is complex, and each person only participates in part. The planner is responsible for service planning; the dispatcher is responsible for operational scheduling; the analyst is responsible for summarizing the assessment, and so on. If the dispatcher, for instance, wants to know the cause and effect of his action, he needs to ask the driver of each relevant vehicle one by one, and then organize the information on demand.

Moreover, data and technical results are readily available and should be taken advantages of for information exchange. With the technology of the Internet of Things, companies can acquire big data by installing sensors, GPS, and so on. There are mature protocols and specifications in the industry to implement system-to-data conversions, such as GTFS. Data mining and analysis provide a variety of traffic data models, such as Origin-Destination-Transfer (ODX) data model that MBTA uses to solve passenger boarding, getting off, and transferring, and performance measurement metric models.

A real data application evidence here is Google map. By entering the origin and destination on the Google map, the user can get specific local public transportation options. These options not only provide the location of boarding, transfer, and destination but also relevant real-time vehicle arrival times, such as “in 18 min”. The calculation of these plans and the arrival time of real-time vehicles is based on the GTFS data published by the particular agency. Since the public benefits from these data, then the agencies who as data producers, are supposed to benefit as well and even more.
However, even with data technology, people are faced with the challenge of dealing with vast and complex amounts of information in such complex systems. Information visualization is an efficient solution because it hierarchically exposes the system with images and people do not orderly read of the system by text. Colin Ware (2013) suggests visualization has the attribute of sensory which has the expressive power for using the perceptual processing. Its representations have the properties of understanding without training, resistance to alternative denotation, sensory immediacy, and cross-cultural validity. While the text has no perceptual basis and thus its representations are hard to learn, easy to forget, and embedded in culture and applications.

Looking again at the Google map example, when we select the list to see the results of the program, we have to read in a specific order. By default, Google map gives priority to displaying the fastest travel time, although it also offers other preferences such as fewer transfers and fewer walks. When we choose to read on a visual map, not only can we see and compare each option at the same time, we can also intuitively see the driving route of each choice on the map. As a result, we obtain more information about the way we would pass through, such as whether a nearby shop passes by, or if a transfer requires crossing several intersections.

Information visualization always has a specific audience. The audience is the particular individuals, to be precise, is a group of people with the same need. From the definition of product design, the used Information visualization is called the Information visualization tool, and the person who uses it is called user. For public transit management, different visualization tools face the different management groups, such as operators, dispatchers, and planners, with the same functions, responsibilities, and specific work matters. An information visualization tool should answer questions raised by the particular management group to solve problems encountered in handling the particular task. For example, the question for bus dispatchers could be whether the buses are suffering or are about to suffer from the traffic jam.
However, an information visualization tool can also be a set of information visualization tools. For example, the dispatchers in the control room face a stack of screens, each with different information. At this point, each screen is an information visualization tool, and the entire control room includes multiple tools. But as a whole, the control room itself is an information visualization tool for dispatchers to complete daily monitoring and scheduling work. This tool of tools also needs to be designed and follows the same design principles as the single tool.

Finally, I would like to emphasize that the significance of the visualization-driven governance lies in the totality of integrated visualization tools. Every conversation can produce a need for a tool or even multiple tools. Using visual tools in one place only, it’s as if Bondy stops bleeding. A tool can solve this specific problem at this moment, but neither tool alone can handle all the challenges in the entire system as mentioned previously. Besides, each tool targets its particular end user and satisfies user requirements in different scenarios. Each user group as persona requires a separate and unique interface (Cooper, 1999). Of course, we can indeed have a so-called super tool or a platform for multiple tools to improve the implementation efficiency through resource integration and allocation. The tools are a pre-requisite for a platform, but a platform is not needed for visualization-driven governance.
As mentioned in Chapter one, visualization-driven governance is not only visualizing public transit relevant data but the totality of visual practices used in management activities. This section will discuss the practical implementation of visualization-driven from the elements of system legibility and design principles.

Elements include information both within and outside the system. Victor and Ponnuswamy point out that the transportation needs are affected by population and its distribution, geography, income, and government policy. They also identify five components that can be visualized in urban system: moving objects (people, goods, vehicles, services), activities (residences, jobs, production of goods and services, movement), infrastructure (buildings, roads, railways, power plants), land (land under varied uses), and policy (goals, decisions, plans). (2012) It shows that taking transit systems as an urban transportation system, the relevant information from outside the system should also be considered.

The information within the system is from elements of subsystems (infrastructure, structure, operations, management). Offenhuber (2017) identified “legibility from below” and “legibility from above.” Legibility from below is physically descriptive, legibility from above is for large-scale prescriptive. This point shows again that infrastructure information alone is not enough for an infrastructure-based transit system. Because it’s socially constructed, information based on management consequences and demands is more critical for governance.
The relevant information outside the system is called context. The advantage of such a classification is that it defines the fundamental boundary of the enterprise from the information resources. In practice, because of the limited resources (human resources, time, and budget), some small and medium-sized enterprises may not have the data capabilities within their borders but have to cooperate with other qualified companies who can provide the data. In other words, beyond the borders, companies rely on sharing, cooperation, mergers, and acquisitions to obtain the access to the data. Of course, this boundary is not fixed, for example, a company may acquire or merge resource companies outside the boundary, thereby expanding the border (Porter, 2014).

Based on this, I divided the elements into five categories:

(a) **Infrastructure** refers to terminals, stops, stations, etc. These types of data are objects of interest (Ware, 2013), including their geographical location and functional space, such as whether they have a parking lot.

(b) **Service structure** refers to elements related to schedules, including routes, shapes (various trajectories of routes), mode (the type of transit, such as bus, rail, train), and other information related to trips, such as the destination, direction of a trip and so on. This type of the data does not just have attributes of the entity, for instance, a route is an entity with an identified number, name, color, service type and other attributes. More importantly, it represents relationships that form the structure of related entities (Ware, 2013). For example, a route is the relationship of the infrastructure elements at the service level, which is about which stations constitute a route in what order.

(c) **Operation** consists of two parts, the behavioral consequences of the manager and the behavior of the passengers. The behavioral consequences of managers, in fact, constructing a realistic version of the structure, including the data as current location, status, and bearing of the vehicle. The behavior of passengers is about tapping cards on the fare machine or boarding a bus, including not only the entity, such as the type and time of the card but also the relationship to the behavior of the manager, such as when and where the passenger enters the subway station.
(d) **Management** means elements related to performance measurement metrics from service delivery policy, such as average daily boarding, the reliability of an entire route, stages of the journey, etc. This category is more related to operations and uncertainty (Ware, 2013).

(e) **Context** mainly refers to information that is indirectly related to the system. The Federal Highway Administration (FHWA) proposes a concept of context sensitive solutions (CSS) to indicate a collaborative, interdisciplinary, and holistic approach to the development of transportation projects. Referring to FHWA as indicated by what that includes in the context, context elements may include but are not limited to: the area’s natural environment, such as a river, open space; the area’s social or cultural environment, such as area’s demographics, elderly, low-income, or minority communities, aspects or significant features of the community; the function of the transportation facility and transportation behavior, such as types of users need to accommodate, modes; the area’s economic environment.

In practice, the principle of legibility includes accuracy, scenario-focus, and temporality.

(a) **Accuracy**: The quality and accuracy of the data directly affect the reliability and validity of the visualization. So accuracy has two meanings here. The first is the accuracy of data quality, including the correctness of the data itself and the validity of the data model. Although this part is not controllable by the visualization process, visualization practices often reveal the precise problems of individual data or models. Munzner identifies that visualizing data structure in detail help people in the situations of exploring the data to find patterns and assessing the validity of a statistical model. (2014)

The other is the level of accuracy at the perceptual level. Munzner suggests that “any depiction of data is an abstraction where choices are made about which aspects to emphasize.”(2014) It’s a data abstraction when a designer considers what can be visualized. It’s a task abstraction when considering why a tool is being used.
The simplest situation is whether or not to use an existing canonical example to make a visual image that represents a class of things (Ware, 2013), such as visualizing a bus using a car icon instead of a person icon. In the representation level, the situation could be whether the exploratory diagram is evident to the operations of the data model.

(b) **Scenario-focus**: The scenario mainly refers to the business scenario. Management experience can be used as a big scenario frame (Figure 3.2). Each large scenario has different subdivided user scenarios, namely business scenarios, composed of users, tasks, and environments. The principle of scenario-focus shows that the design of the visualization tool should apply User-Centered Design principles and processes. The reason why the term user-centered is not used here is that user-center emphasizes solving the problem of usability of a single tool, while legibility emphasizes the problem of communication from the perspective of the tool as a whole, namely the problems of the message, signal, and noise (Shannon, 1964).
The scenario determines the legibility of elements. When visualizing the stations, for instance, for the dispatcher who operates a single subway line, the geographical distribution of the station is not crucial information, but rather the relative location. But for planners, geographical distribution needs to be visualized because of the need to refer to data from other lines as well as other external data.

The scenario also determines the direction of legibility or the architecture of the information. In the horizontal direction, Ware proposes strong pre-attentive cues for the ease of information search. In the vertical direction, it needs to solve the focus-context problem which is a problem of finding detail in a large context. So which information should use strong or weak cues and which information is detail or context depends on the management scenario.

Therefore, selection of relevant elements and fitting them into proper architecture is the meaning of the scenario-focus principle. Natalia Andrienko, Gennady Andrienko, etc. proposes that visual analytics of spatial and temporal data has four different focuses on movers, spatial events, space or time. (2006, 2013) The representation of focus requires the different view. Manuel Lima introduces three fundamental views: macro view, relationship view, and micro view. (2011) Macro view provides a bird’s eye view and highlights certain patterns. Andrieko et al. described it as “see the whole (overall behavior),” “simplify and abstract (display essential features of the behavior),” “divide and group,” “look for recognizable.” Relationship view is concerned with connections between representations. Andrieko et al. described as “see in relation,” “establish linkages,” “establish structure.” Micro view detailed information on a single representation. Andrieko et al. described as “attend to particulars.”
To manage the intricacy of the three views, Lima proposes three concepts: adaptive zooming, overview and detail, and focus and context. These concepts require “interactive environments” (Fry, 1997) where users can zoom, filter, extend detail or rotate. Interactive visualization is a method for supporting scenario decision-making process in transit planning. As the success depends on the possibilities of communication between participants, and the effectiveness of the decision-making process depends on the high awareness of their choices. The key of information visualization tool is providing support for the understanding, exploration, and communication of information. (Masala & Pensa, 2014) A high level of interaction enables the cognitive process of big data and their complex relationships, especially for expert-level users. It facilitates two general information processing approaches: one is interactive with custom filters, and another one is scenario comparison from static performance outputs (Miranda & Kurkcu, 2017).

(c) Temporality: Public transport services are not only spatially deployed but also time-based. This principle determines the time dimension of the visualization. For example, metric evaluation requires the use of archive data over a period of time, while real-time data may be more needed for those focusing on current operating status. However, there are also some scenarios using both data at the same time. For example, the dispatchers in the control room not only care about the current moment of operation or where the buses are but also need to pay attention to the punctuality of each station and the volume of waited passengers to decide whether or not to require temporary re-scheduling. These metric data needs to be compared and calculated through archive data.

In summary, the principle of legibility determines the specific design decisions from representation and interaction. Specific examples are found in the case study.
This chapter describes the application of visualization-driven governance approach to the MBTA (the Massachusetts Bay Transportation Authority). Three projects presented and described in this chapter derives from the work that I developed while working at the MBTA, namely on the design and implementation of a Bus Traffic tool, a Green Line Vehicle Location Viewer, and an Origin-Destination-Transfer (ODX) Visualization tool. Each of tools addresses different issues in the overall system: how on-site dispatchers use the Green Line Vehicle Location Viewer to ensure the green line’s reliability, how dispatchers use the Bus Traffic tool to increase bus reliability, and how planners can get information from the ODX tool to help them plan. These tools are currently used by them for their daily work. They not only show how visualization tools can be implemented in a public transportation system and bring values to it, but also the totality of these implementations demonstrates how visualization-driven governance increases the communication possibility and efficiency and thus improves the quality of service for customers.
About MBTA

MBTA is the public agency responsible for operating most public transportation services in Greater Boston, Massachusetts. By the end of 2017, MBTA has 8,362 stops and stations, 8 subway lines, 192 bus routes as well as ferry and other special services (Figure 4.1). In Aug 2017, around at least 41,666 people are using the services for daily life.

The purpose of MBTA is to provide high-quality public transit services. In the “MBTA Service Delivery Policy,” MBTA stated their service objectives, including service availability, accessibility, reliability, comfort, communication, safety and security, rider satisfaction, and environmental benefits, and established quantifiable evaluation criteria.

The department where I worked is OPMI (Office of Performance Management and Innovation). It is a shared service between the MBTA and the MassDOT (the Massachusetts Department of Transportation) with a focus on using data to improve performance across all of the operating divisions. Simply put, OPMI is equivalent to an inter-departmental internal consulting group that explores possible solutions to improve the performance of the system. To this end, OPMI works with staff from the MBTA, the MassDOT, and the Central Transportation Planning Staff (CTPS), along with members of academia, and various planning and advocacy groups.
The following three visualization tools are from the projects that I participated in OPMI. Because these projects are exploratory, users are substantially involved in the requirements definitions and development process. From the beginning of the project, the user puts forward a simple idea that what answer they want to get through this visualization tool. Then, after we have developed the simplest version of the function, we continue to experiment and communicate with users and continue to add new features, while optimizing the data and improving the design, until the tool can meet the needs of users maximally.

The benefits of this continuous feedback process for the thesis are not limited to:
• being able to know how users use these tools in their specific work scenarios;
• finding an answer involves mining other relevant and useful information. For example, the information could be about other subsystems or context as well as reflections on the relationships between subsystems;
• understanding how the user interprets the information provided by the visualization, including whether the representation of visualization meets the definition of the data in the perception level, which embodies the importance of the principle of accuracy;
• discovery of extreme cases. The need to deal with these extreme cases reflects the user’s requirements for the principle of accuracy. And the process of dealing with extreme cases lead to new dialogues and inspired a mutual learning process. The deeper the users understand the construction principles and limitations of the data model, the more practical situations the builders gain to improve the data model.
• discovery of limitations. “Serving as a liaison between the developers and the users, I saw firsthand how challenging it was at times for users to understand limitations imposed by the data structure.” one of service planner working with us on ODX project says, “most of my attention has been focused on understanding how ODX handles transfers, since it needs to properly capture transfers in order to meet that need. And by ‘properly capture’, I mean ODX’s definition of transfer should be compatible with Service Planning’s definition of transfer. As we dig in to the weeds, we realize that it is not always the case that we approach the same fuzzy concepts in the same way.”
Project: Bus Traffic tool

Background

Bus Control Center is responsible for the operation and dispatch of buses. In day-to-day scheduling, dispatchers sit in the control room, surrounded by a large screen of different information, enter operations and record information in their own operating systems. Each dispatcher is responsible for several fixed lines and communicates with the drivers on duty for delays that may be encountered. Because the current operating system does not have information of traffic jams, whenever the traffic jam happened, only the driver can tell the dispatcher from the radio, but the dispatcher cannot get a better prediction nor knows more about the real current situation. Bus Traffic tool is a web-based tool for bus dispatchers to know system-wide traffic jams using Waze crowd-sourced traffic jam data and bus vehicle location data (Figure 4.2). Through the tool, dispatchers are able to see whether the buses are or are about to be suffering from traffic jams on screen in real time. By grasping this information, they can take actions as issuing an alert for the delay to the public or notifying the driver about a detour.

I designed and developed this tool based on a simple description of requirements and follow-up feedbacks. I started with visualizing the location of vehicles and traffic jams and appended more details such as the label and direction of the vehicles, the speed of traffic jams, and alerts. Finally, I developed more advanced features such as a search for a specific vehicle or a filter of the current display by the given routes. I developed by D3.js and Leaflet.js. D3.js is used to visualize the vehicles, jams, alerts, stops, and routes. Leaflet.js is used to create a base map.
Legibility

In general, the tool visualizes infrastructure system, service structure system, operation system, and context of traffic jams and alerts. To this end, it integrates Waze API, MBTA real-time API and GTFS (General Transit Feed Specification). GTFS contains infrastructure information of stops, and structure information of routes and their shapes (trajectories), trip schedules, and their relations. MBTA real-time API provides its real-time operation information, namely, current vehicle position by vehicle identification. Waze API provides two types of real-time information as context: traffic jams and alerts. The data is pre-filtered by geographic area query (a polygon that represents Massachusetts).

During the implementation, the tool also applied the principles of accuracy, scenario-focus, and temporality. The following examples will be discussed through the specific features and design of the tool.

In the management experience, this tool is used for operation stage. As mentioned earlier, the dispatcher sits in front of the computer and monitors the real-time operation of the bus, identifies potential delay events and takes corresponding scheduling actions. In information priority, they need to first see the actual movement of the vehicle in the city and then combined with traffic jam information to estimate whether a bus is about to enter a congested road segment or whether it is already in congestion.

For congestion information, they want to know the degree of congestion, such as the speed of congested roads. For the bus approaching a congestion, in order to know if this bus will pass through congested roads, they need to look at the trajectory of the vehicle. Although they are already familiar with the routes for long-term working experience, they occasionally want to see it more directly on the map. For this reason, I only added hover interactions so that they can hover a bus and then the bus’s route will be highlighted. Besides, they not only want to be able to see all the buses but also to focus on some buses of some routes or one particular bus on the map. This is why I later added features of the search for a specific vehicle and the filter for current display by some certain routes.
Figure 4.2

Bus traffic tool screenshot
The tool primarily embodies the principle of accuracy in color coding, vehicle information, and map tiles. First of all, in color coding, congestion information is displayed on the map as a polyline. In order to visualize the congestion level, the color of the polyline is from red to green, to indicate extreme congestion to less congestion. The rationale of color coding comes from people’s perception of the color of the traffic, such as the red light suggests to stop and green light to go, the color code Google uses when displaying congestion for drivers.

Second, it is the vehicle code used to identify the vehicle. In the initial version, I used the route name to indicate which route the bus was. Later, users responded that they are more concerned with the identification of a particular bus, which is, the bus’s identified label. This also reflects their requirements for the accuracy of information.

Finally, in order to combine information of congestion and the current location of the buses, a geographical map was used in the design. Initially, I just used a generic map tile, simply showing street names and landmarks. Later, after feedback, they hoped to see more detailed traffic information, such as whether there are traffic lights at the intersection and whether the intersection can turn left. So I changed the tile so that they could tell the driver more accurately whether to turn left in front of the congested road section based on traffic information without affecting the passengers boarding and dropping at the next stop.

In principle of temporality, due to the extreme demand for timeliness of operations, the tool uses real-time data updates. The data is requested every ten seconds for a real-time update. The significant reason is that MBTA real-time API takes ten seconds for the data to be synchronized on multiple servers. If the interval of the request is less than ten seconds, some vehicles might visually move back and forth on the map when data updates.
Project: Green Line Vehicle Location Viewer

Background

Green Line Vehicle Location Viewer is a web-based tool for Green Line on-site dispatchers to track the trains on tablet and smartphones. Through the tool, dispatchers can see where the trains are for each Green Line branch. By getting this information in a timely manner, they can send the train more accurately and ensure the reliability of services.

This is a pilot project to verify that visualization tools are beneficial to the on-site dispatchers of the Green Line. The Green Line is a rapid transit line of MBTA, its operation ensuring the reliability of the Green Line is very complicated. The Green Line has four branches (B, C, D, E), the four branches converge at Copley and change direction at different terminals, and a few of them are traveling on the ground with other vehicles (Figure 4.3). On-site dispatchers are staff distributed on the platform, and their primary responsibility is to ensure that each train operates on time (Figure 4.4).
Temporary changes in the schedule, such as a train will be delayed into the station or change the branch, is required the on-site dispatcher to inform the control center. And the dispatcher in control center enters new information into the computer system to ensure the normal operation of AVI (Automatic Vehicle Identification). AVI not only automatically changes track for different branch vehicle, but also collects the correct data. Since the trains arrive and depart at the same track, the on-site dispatchers not only need to know the arriving train but also the departing train. Before, they were able to ask colleagues at other stations via radio or by phone about other trains, and only a few terminal platforms, such as Riverside, have dispatchers.
The development of this tool includes two iterations. In phase one, it provides the legibility of D branch trains for Riverside dispatchers, and I’m responsible for development utilizing D3.js. In the second stage, all the dispatchers were supposed to use this tool to filter the data of a particular branch to get information (Figure 4.5), and I did both design and development. According to user feedback, they do spend more time focusing on their own branch train, but they also look at the trains on other branches on the mixed segment, where all trains pass from Copley to the North Station. Because they want to know whether the trains from other branches would affect their trains for the merge.

**Legibility**

This project, like Bus Traffic tool, visualizes the infrastructure system, service structure system, and operation system. However, this tool only uses the MBTA real-time API. Because MBTA real-time API is based on GTFS and provides information of stops, routes, trip schedules, and also provide prediction data in addition to providing real-time operation information.
Figure 4.5
Greenline headway viewer tool screenshot
The tool is used for operations and so follows the same temporality principle as Bus Traffic tool. However, the information that the on-site dispatcher of the Green Line is concerned with is different from that of bus dispatcher. The on-site dispatchers focus on their platform-related information, that is, where the train that should have entered this time is now. Moreover, the Green Line train is always on the track. The impact of congestion on the general road is not as great as the bus, and there is no possibility of the detour. Therefore, in designing, there is no geographical map, but an abstract diagram.

This diagram reduces the complex information and shows only the relative location of the stop, the relative position, and direction of the train. And according to a more subdivided user scenario, five views for each branch and mixed segment are provided. Under each view, the on-site dispatcher can not only see the locations of his trains under that particular branch, but also select the platform he works on, and only obtain the details of the trains that his platform is about to arrive at.
Project: ODX tool

Background

ODX tool is a web-based tool for service planners to gain ODX data. ODX is a ridership model, which stands for Origin, Destination, Transfer of a passenger’s journey (Figure 4.6). Through the tool, planners can know where people go and from, and how they transfer at a certain time and with a certain fare type and can evaluate how many people will be affected by their planning decisions.

ODX is a data model based on site traffic, which assesses passengers’ travel needs by using which station a passenger taps into. However, ODX analysis is far more complicated than simply displaying data. The analysis of passenger volume on each station is only a simple review of current services. In fact, for planning, what is more important is which area the passengers come from and go to. In the regional analysis, station volume is only a source of regional data, not the ultimate goal. And this analysis cannot be separated from the actual natural environment of the city, the status of infrastructure constructions. So there is a need for this tool.

I am responsible for the design and development of this tool utilizing by D3.js and Leaflet.js. From the very beginning, this project has only one general idea, and it can now support many complex functions such as regional analysis and site analysis. In this process, I also experienced the complexity of service planning, especially in such a large service system that has been running for many years. Not only is the comprehensive view of the system critical, because a single change in service may lead to many changes, but also extreme cases are making it difficult to understand the exact travel behavior of users only through data.
Legibility

This tool visualizes the information of infrastructure system, service structure system, operation system, and management system. Same as the previous two tools, GTFS provides infrastructure information of stops, and structure information of routes and their shapes (trajectories), trip schedules, and their relations. ODX data provides ridership data based on AFC (Automated fare collection). Moreover, it integrates TAZ (Traffic Analysis Zones), clusters (a custom geographies data), as well as route characteristic tables derived from Service Delivery Policy metrics and passenger surveys, such as span and frequency of service, reliability, service usage allocation and so on.

ODX data model is based on the stop at which the passenger gets on, gets off and transfers. So users get ODX results by selecting stops. But in fact, the users have their mental model in the analysis. After the user interviews, there are mainly three scenarios: stop analysis, route analysis, and area analysis. Therefore, in addition to direct searching from more than 8,000 stops, we also provide more options depending on the scenario (Figure 4.7), including selecting stops by selecting one or more routes (Figure 4.8), or by selecting one or more zones (clusters, TAZ, or draw a custom shape, Figure 4.9). The three scenarios may also be combined when doing analysis, such as analyzing stops on a certain part of a route. In order to allow users to complete different ways of selection in one interface, I built the legibility of network first.

The result of ODX is based on the origin, destination, and transfer points of each passenger journey. Each journey also has many other attributes, including the mode, the time period, the day of the week, the routes, the fare method, and the fare type. How to organize and structure this information depends on the hierarchy of analysis questions.
Figure 4.6
Anatomy of ODX data:
“O” stands for origin, “X” for transfer, “D” for destination

Figure 4.7
Selection options of ODX tool:
select stops by stop, route, or area(cluster/TAZ)
Figure 4.8
ODX tool screenshot: selection by a route:
34 stops on route 43 have been selected
Figure 4.9

**ODX tool screenshot: selection by a cluster:**

7 stops in cluster 223 has been selected
After user interviews, in the results representation, the summary data is given first, namely all journeys that use these selected stops as starting points, endpoints or transfer points (Figure 4.10, 4.11). For users who need more in-depth information, they can click to see where the journeys starting from the selected stops are from, where the journeys getting off from selected stops are coming from, and which routes they transfer from and to at these stops (Figure 4.12). Corresponding to the previous analysis scenario, the results are divided into the analysis of cluster, TAZ, stop and route. Cluster, TAZ analysis answers which area the passengers come from and go, and the count of the journey (Figure 4.13). Stop analysis answers the number of the journey at specific stops (Figure 4.14). The Route analysis answers which route the passenger used, but this piece of data is only provided as table and has not yet been visualized.

In the real world, analysis of the area is much more useful than the stop. As the planner working with us on this project says, “In an era where we have orders of magnitude more data available than was imaginable just 20 years ago, granular data isn’t necessary for most analyses. That’s why clusters are generally more useful to me than stop-level info—there are too many stops to consider them individually.”
Figure 4.10

ODX summary result from selection:
calculate the count of passenger’s journeys using selected stops as
origin, transfer, and destination
Figure 4.11
ODX tool screenshot: summary result
Figure 4.12

**ODX in-depth result from selection:**
calculate respectively where the destinations are when they use selected stops as origins, how they transfer by routes when they transfer at selected stops, and where the origins are when they use selected stops as destinations.
Figure 4.13
ODX screenshot: in-depth result on cluster analysis
It shows which area people come from when they end their journeys at those two selected stops
Figure 4.14

**ODX screenshot: in-depth result on stop analysis**

It shows which stops people go to when they start their journeys at these selected stops.
Discussion

The tools developed in these projects either open up new channels of communication for managers, as ODX tool or consolidate the original dialogue methods as Green Line Vehicle Location Viewer. But more importantly, this kind of dialogue is in the form of data visualization and makes the service system more open.

On the one hand, it makes the dialogue between subsystems more effective. The flow of data proves the efficiency of the conversation between subsystems. The service planning data comes from the data generated by the operation and also from the metric measurement data of the service evaluation. The basis for operations comes from the initial service plan. The temporary dispatch in operation is not only a decision made on the basis of the original plan but also a response based on the specific situation of the current operations. The evaluation of services comes from the statistics and comparison of service operations and plans.

In the absence of these visualization tools, MBTA’s service planners can obtain data through database queries, but there’s a knowledge barrier that not everyone can write a query. Besides, data query and reading are time-consuming and threshold-consuming. As the planner working with us on ODX project says, “I have been playing with ODX data for over a year now, primarily by querying it using PostgreSQL. I frequently serve the role of “the data guy” in the Service Planning Department, and so I’m the only one who has enough knowledge of SQL to be able to easily query the ODX database. I have needed more planners to help me test out ODX in order to understand how it handles different issues, but since you needed SQL skills, I needed to run the queries. Now that we have ODX tool, more people are able to explore ODX on their own. This has sped up the testing process by allowing more people to investigate issues and bring them to my attention for further testing.”
Moreover, because the topographical features of the Boston area are also crucial to planning, it is difficult to draw conclusions based on raw data because of its complexity. For example, which routes are crossing the Charles River and which routes are actually circled, only very experienced planners may remember. As the planner working on analysis of ODX says, “Looking at load profiles (spreadsheet based table of ons, offs, and load by stop order by trip), one can be tricked into thinking that these are evenly spaced since each stop gets one row. Once you put it on a map though, you can start seeing how that’s not necessarily the case. I think we are really just getting started in trying to understand how to use this tool.”

On-site dispatchers from Green Line have no quick access to comprehensive information. Not only are they unable to query the database, even permitted them, but they also have no way to deal with the fast scheduling requirements through slow queries and analysis. The way to get information from other drivers or dispatchers requires not only preconditions, such as knowing whom can get the needed information from, but also how to contact them. And the information obtained in this way is not always comprehensive. For example, they may miss the critical vehicle information, or do not have time to contact the subsequent vehicles to observe their own scheduling results.

The combination with third-party data, such as Waze, further opens up the overall relationship between the service system and the outside. In this way, the dispatcher can not only ensure that there is no congestion between the vehicles he operates, but also expand his perspective outward to his own subsystem by knowing the traffic conditions of the entire city, and eventually make decisions that are more beneficial to both the entire city’s traffic and bus services.
Visualization-driven governance can bridge perspectives between the various subsystems in an innocuous manner. Because of the board collaboration, OPMI sees the gaps among different perspectives of the stakeholders. As the Director of OPMI said during my interview, “They have their own bubbles in their minds.” It’s important to coordinate these perspectives of stakeholders, expecting to make them have higher or broader perspectives to do the work better. “What we do will not break the system. The trains are still running there even if our tool is broken.” She said.

On the other hand, it also makes the entire system more competitive. Objectively speaking, public transit is a competitive relationship with other modes of transportation, passengers can choose taxis, Uber or self-drive. This competition puts open system inward pressure. At present, MBTA has almost nothing involved passengers in participation in service evaluation in visualization way. However, it’s conceivable that how visualization could bring benefits by adding it to the service evaluation stage and involving the public.

Finally, relying on existing tools alone is not enough. These projects not only confirm the possibility of success of the visualization tool in governance but also trigger the thinking and needs of the follow-up visualization-driven projects. In MBTA case, they benefit from the application of visualization to presentations in the first place. As a designer in OPMI says, she likes to use visualization for awareness and communications. People already have too many documents to read and do not tend to pay attention. The visualization catches their eyes. It’s also a good way for the employees and leaders to explain or present their work to the external parties. After they implemented some tools, they expect to build more. Not only on web-based applications, but also the traditional media, for instance, the large screen on the wall with numbers or diagram for operators working on the platform. As they said, “We don’t have doesn’t mean we don’t need.”
The practice of visualization-driven governance is not only to visualize data, but more importantly, to define information requirements from the entire management experience, and to consider information architecture and visualization from the user scenarios. In this process, I continuously accumulate transit-related industry knowledge, data knowledge, and opportunities to observe the daily work and habits of people in these systems, including what difficulties they encounter, how they address, and even their acceptance and preference of information visualization. As a result, I saw great potential for the visualization-driven approach.

In addition to providing internal employees with information visualization tools, I also had a chance to design and develop a visualization mockup for the MBTA marketing department. This mockup expresses availability of MBTA service network. It’s hard for me to imagine anything better than visualization to convey this information. In this visualization, we can filter time and weekdays to see what the service network looks like at different times. Although we know we can expect to see different lines interweaving together during the interaction, when we actually see it, we will still be surprised by the results of exploration. After we saw the commuter rail appeared on the map after 3:30 in the morning and the Green line appeared after 4:30 in the morning, although we have faced this data for a year or even several years, we still feel that we know more about our service and city than before.

Another story takes place in the ODX project. In order to optimize the ODX model, team members hope to find a critical value related to transfers as a new reference variable value. On the team’s proposal, I spent a few hours making a visualization model for them to explore. In this model, they can see how the data is distributed on the map, combining each transfer case with the actual geographic features and the available services, and filtering the map through the critical values. After 2 hours they got the desired result. If they just look at more than 5,000 lines of raw data in a database or Excel table, I can hardly imagine how much time they will spend to produce results. As the planner working with us says, “a quick, disposal visualization may not be used for very long, but help explore an issue in depth more efficiently.”
During this thesis, I also participated in some information visualization lectures. Siqi Zhu introduced the project on transportation in his speech “Visualizing the City: Now in Progress” in 2018. His team used maps to visualize passenger flow and travel time to help planners locate specific areas where the service needs improvement. They also used similar data to create different visual maps for the general public to get information and help them decide on travel plans. This led me to think: where are the differences in visualization when it comes to mass, market, and management?

From an empirical point of view, indeed, as Siqi said, the public does not have as much energy, and time to read the information, we need to quickly grasp their attention and interest and therefore may need more reduction in the amount of information. But conversely, as Ann McDonald, an associate professor of design and interactive media at Northeastern University, asked questions in the Siqi lecture: if the general public has the right to vote on service decisions, should they also have such an explorable visualization to give them more comprehensive information?

Although my thesis is only discussing management, how visualization supports planner to make decisions, how to support operations, and how to help evaluate, but the management of transit services is not limited to the internal. When we initiate questionnaires for passengers to make assessments for services, should we think about what these passengers who responded are evaluating? As consumers, they estimate for their own experience of trips. This is useful for improvement of a specific service. But the transportation service is the social artifact of a city. Where and how can these people show their value as a member of the city?

Moreover, good governance will not only benefit the transportation agencies themselves. On the one hand, the ultimate beneficiaries of the results are the general public. Good public transit not just brings good experience at the moment, but also the improvement of the quality of life by providing safe, convenient and affordable transportation alternatives. For example, by supplied with mobility, people who can’t afford a car can have more employment options, and the senior can get around in their communities and stay connected to family and friends.
On the other hand, the process of engaging the public enhances the understanding and communication between them and agency. Living in Boston, I often hear people dissatisfied with public transit, and even have encountered a passenger on the Green Line who can’t help complaining that she couldn’t keep up to the scheduled commuter rail because this train was already delayed. I can understand the frustration and inconvenience of unreliability, but I can also give understanding. It not only because I was working at MBTA and saw employees dedicate to their work. But also because I accessed the real situation. For instance, once I was trapped on the subway because of a disabled train, so I turned on the Green Line Vehicle Location Viewer on my mobile phone and saw a picture of multiple trains jammed. I immediately understood how difficult the current dispatchers were. Thus, I believe that agencies and the general public can achieve empathy and better communication through visualization to eventually promote a more harmonious community.

Although I’m talking about transportation, it’s not just about transportation. When a transportation agency tries to implement a new platform, how do they reach a consensus with other stakeholders? How do they state their contributions and fight for their interests? How do they demonstrate what kind of support they can give to other construction groups? They are all based on the understanding and communication with the external.

The case and experience of my thesis are from MBTA. I have also been asked many times: how do you think your proposal can be applied to other cities, such as New York and Beijing? Indeed, the traffic conditions in each city are different. Compared to Boston, New York’s services are more complicated, and other U.S. cities may be more straightforward. In Beijing, China’s urban representative, the quality of service is not its ultimate goal, but rather management of congestion by providing better services with lower price to attract more people to use public transportation.

Moreover, each city has different levels of transportation technology. The subway in Beijing started from the middle of the 20th century, while Boston and New York began in the 19th century. It takes more effort and money to upgrade and maintain the subway system for hundreds of years. However, the governance of public transportation is an urban problem. It is essentially a social issue and not just a technical issue.
What I feel deeply touched is the Digital Matatus project introduced by Sarah Williams in her lecture “Build IT, Share it Hack it: Using Data for Policy Change” in 2017. She is an assistant professor of information technology and urban planning at MIT whose team collected data using cell phones, participated private Matatus drivers to chart over 3,000 Matatus stops and finally produced the first transportation system map in Nairobi. This project made me once again see the public transport system is social: our deployment in technology is based on people’s needs. The data generated by technology is essentially the sum of the participants’ behavior in the system.

That so-called technology is always evolving. It brings us more data faster, helps us with data mining and algorithms, and even makes more informed decisions on simple things. Some problems that can be visually resolved at present can indeed be solved by technology in the coming day. But as McCarthy and Wright propose, people don’t just use technology but live with it and we should not understate the felt life in their accounts of experience. (2007) In this case, the users of these tools not only improve the passenger’s experience by using them, but also embody it as their daily experience. Then, how do they meet and respond to this technological age? How do they perceive these massive data? How to understand these esoteric data models in mind? How to accept and trust these machine decisions emotionally? In turn, how can these technologies be strategically deployed and applied through these insights?

I believe, visualization is still our only way out. It efficiently gives us legibility. Because we can read and understand it, we are all informed in an equal way. Even though we deal with different issues, we still able to take advantages of what we read and understand to get what we need. But to thoroughly enjoy the benefits of technology development, a transportation agency requires more than just visualizing data. They also need to open up their system to provide a proper channel for information exchange and understand their management experience and the scenarios clearly, then design the visualization tools under the principle of legibility.

Finally, this perspective can also be applied to other complex organization or system, not only public transit. In the era of big data, many companies drive their business based on visualized data analysis, locating user preferences and providing personalized services. However, in fact, the potential of
application for information visualization is much more than business intelligence. As my thesis has proved, this application can provide or improve the legibility of their complex systems, help employees understand data and communicate more efficiently. It also opens the opportunity for communication with their customers. These ultimately will enhance their quality of service and make them more competitive.
Bibliography


Cooper, A. (1999). The Inmates Are Running the Asylum. Indianapolis, IN: Sams


Ware, C. (2013). *Information visualization perception for design*. Waltham, MA: Morgan Kaufmann