Scoping and Conducting Data-Driven 21st Century Transportation System Analyses

January 2017

U.S. Department of Transportation
Federal Highway Administration

Cover Image Sources (clockwise from top left)
Thinkstock.com
Cambridge Systematics
Thinkstock.com
Thinkstock.com
Thinkstock.com
U.S. Department of Transportation
Thinkstock.com
Center Image: Federal Highway Administration

Technical Report Documentation Page

1. Report No. FHWA-HOP-16-072	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Scoping and Conducting Data-Driven 21st Century Transportation System Analyses		**5. Report Date** January 2017
		6. Performing Organization Code
7. Author(s) Karl Wunderlich (Noblis), Vassili Alexiadis (Cambridge Systematics), and Peiwei Wang (Noblis)		**8. Performing Organization Report No.**
9. Performing Organization Name And Address Noblis, 600 Maryland Ave., SW, Suite 755, Washington, DC 20024; Cambridge Systematics, Inc., 4800 Hampden Lane, Suite 800, Bethesda, MD 20814		**10. Work Unit No. (TRAIS)**
		11. Contract or Grant No. DTFH61-12-D-00048-T-5008
12. Sponsoring Agency Name and Address U.S. Department of Transportation, Federal Highway Administration, Office of Operations, 1200 New Jersey Avenue, SE, Washington, DC 20590		**13. Type of Report and Period Covered** October 2014 to December 2016
		14. Sponsoring Agency Code HOP

15. Supplementary Notes

The Government Task Manager (GTM) for FHWA is John Halkias

16. Abstract

Improvements in commercially supported analytic software, increasingly powerful advanced computational platforms and a newly data-rich environment enables the 21st Century analyst to model large and dynamic surface transportation systems. However, a case can be made that the gains made by the individual analyst have outstripped the gains made by the organizations that manage transportation systems to capitalize on new analytic techniques. These analysts remain largely isolated from the mission of improving surface transportation system performance. Often, projects are defined and delivered to the analyst as an accomplished fact. The analyst is not frequently involved in diagnosing transportation system problems or using data to assist in analytic project scoping. Data and models developed for past projects are discarded, lost, or documented so poorly that they cannot be leveraged for future projects. With rare exceptions, there is a lack of advanced institutional models to systematically and consistently leverage the power of transportation analytics embedded within the broader transportation system management mission.

This guidance document defines a Continuous Improvement Process (CIP) to integrate data-driven time-dynamic operational analyses within transportation systems management, featuring:

- **System Diagnostics**: Systematic methods for generating, integrating and prioritizing candidate analytical projects using a portfolio management approach.
- **Project Scoping**: Integrated analytical project and data acquisition scoping procedures (including a scoping tool).
- **Data Preparation**: Management practices for the preparation and analysis of integrated transportation system performance data with contextual data for the identification of operational conditions.
- **Analysis**: Best practices for data-driven analytical project execution.
- **Management of Analytical Capital**: A standard documentation procedure that supports project continuity from inception to lessons learned and preserves analytical capital (data and models) for the benefit of future analyses.

17. Key Words Simulation Modeling, Transportation Analysis Tools, Operational Conditions, Analytic Project Scoping, Analytic Project Screening, Transportation Systems Management, Continuous Improvement Process	**18. Distribution Statement** No restrictions.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	136	N/A

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

EXECUTIVE SUMMARY

The 21st Century analyst can readily draw upon improvements in commercially supported analytic software, increasingly powerful advanced computational platforms, and a data-rich environment to model large and dynamic surface transportation systems. However, have the gains made by the individual analyst outstripped the gains made by the organizations that manage transportation systems to capitalize on new analytic techniques? When analysts work in relative isolation from the mission of improving surface transportation system performance, they are frequently not involved in diagnosing transportation system problems or using data to help scope an analytic project. Data and models developed for past projects are discarded, lost, or documented so poorly that they cannot be leveraged for future projects. Too often, there is a lack of advanced institutional models that systematically and consistently leverage the power of transportation analytics embedded within transportation system management's broader mission.

The Federal Highway Administration (FHWA) has developed guidance on how 21st Century transportation analytic resources (data, tools, and computational platforms) can be systematically embedded within the transportation system management process, with four significant results:

- Enhanced characterization of transportation system dynamics and problem diagnosis.
- Improved analytic project scoping.
- Data-driven experimental design that limits risk and maximizes analytic insight.
- Systematic execution and documentation of analyses to preserve accrued analytical capital.

The FHWA guidance supports transportation professionals—including data managers and transportation analysts—at different levels of technical expertise and in a wide variety of uses across short-, medium- and long-term decision horizons. As shown in Figure 1 (on the next page), the 21st Century Analytic Project Scoping Process consists of a four-module Continuous Improvement Process (CIP):

- **Module 1: System Diagnostics**. Characterize system dynamics and diagnose problems.
- **Module 2: Scoping**. Perform data-driven transportation analysis project scoping.
- **Module 3: Preparing Data**. Collect and organize the data needed to conduct a transportation analysis.
- **Module 4: Analysis**. Conduct and document transportation analyses.

In Module 1, the analyst uses system performance measures and early diagnostic activities to develop preliminary analytic problem statements and prioritize the identified problem statements with a risk-reward project-screening approach. The Analytical Problem Statement(s) associated with a high-priority concept—the final product of this module—is the foundational connecting document that initiates the second major step in the process.

The project scoping in Module 2 includes a more detailed project definition, the identification of project-specific performance measures, a refinement of mitigation strategies and data needs, tool selection, and cost and schedule estimation. This guidance includes a scoping tool that helps the analyst complete the cost and schedule estimation. Module 2 ends when the Project Scoping

Summary is completed. The summary provides enough information to initiate data preparation and analysis work (Module 3).

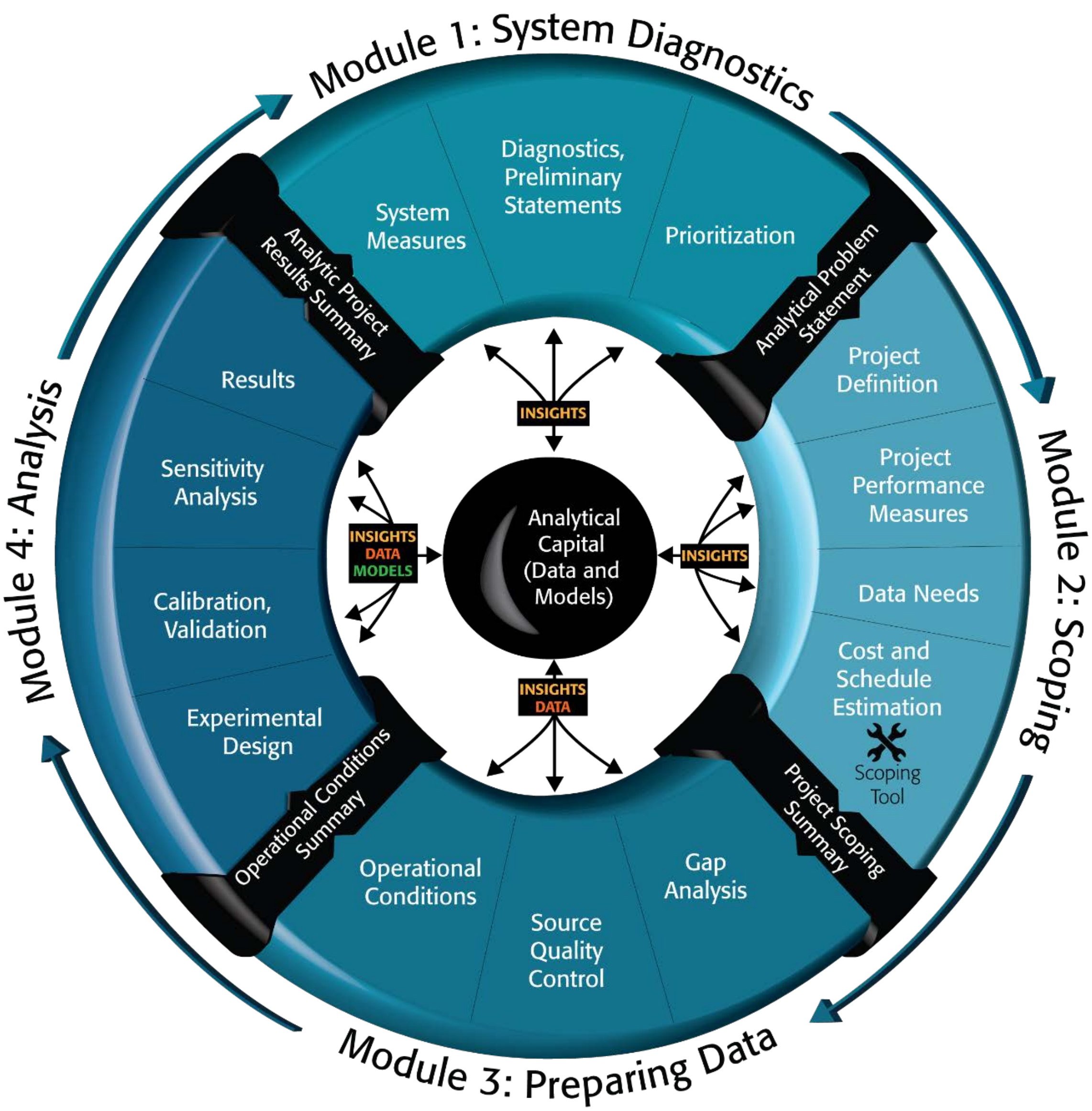

Figure 1. Diagram. 21st Century analytic project scoping process.
(Source: Federal Highway Administration.)

In Module 3, a data analyst verifies the consistency and quality of the available data and outlines the data collection plan to fill up the gap between the data needs and availability. Depending on the nature and scope of the project, the data analyst may identify and summarize a representative set of operational conditions that are critical for creating a strong analytical plan (Module 4).

To execute an analytical project (Module 4), the analyst creates a detailed design with experimental and control cases, as well as calibration and validation of models under various operational conditions and a sensitivity analysis. Documenting the results is the final step—the analyst documents project findings to inform decision-making and also captures lessons learned to improve the agencies' implementation of the 21st century analytic project scoping process (diagnostics, scoping, data preparation, and analytics) for the next cycle.

Following this four-step process enables transportation system management organizations to successfully achieve their objectives and realize a number of important positive outcomes related to greater insight, better analyses, and reduced costs and risks:

- Increased understanding of how the system as a whole operates and changes over time.
- More relevant and targeted transportation analytics.
- Minimized redundancy and technical risk in a portfolio of analytical projects.
- Reduced costs of conducting analyses over time.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
AASHTO	American Association of State Highway and Transportation Officials
AMS	Analysis, Modeling, and Simulation
ATIS	Advanced Traveler Information Systems
AVL	Automatic Vehicle Location
B/C	Benefit/Cost
BSM	Basic Safety Message
CATT	Center for Advanced Technology in Telecommunications
CCTV	Closed-Circuit Television
CIP	Continuous Improvement Process
COG	Council of Governments
CONNECT	Communications Networks, Content & Technology
COTM	Contracting Officer Technical Manager
DCM	Data Capture and Management
DMS	Dynamic Message Signs
DOT	Department of Transportation
DSRC	Dedicated Short Range Communications
EPA	Environmental Protection Agency
FHWA	Federal Highway Administration
GIS	Geographical Information System
GPS	Global Positioning System
HOT	High Occupancy Toll
HOV	High-Occupancy Vehicle
ICM	Integrated Corridor Management
IMO	Integrated Mobile Observation
INCOSE	International Council on Systems Engineering
ITS	Intelligent Transportation Systems
JPO	Joint Program Office
LPR	License Plate Recognition
MLIT	Ministry of Land, Infrastructure, Transport, and Tourism
MMITSS	Multi-modal Intelligent Traffic Signal Systems
MOE	Measure of Effectiveness
MPO	Metropolitan Planning Organization

Abbreviation	Definition
MTP	Metropolitan Transportation Plan
NOAA	National Oceanic and Atmospheric Administration
OBU	On Board Unit
ODME	Origin-destination matrix estimation
RCTO	Regional Concepts for Transportation Operation
RDE	Research Data Exchange
RITIS	Regional Integrated Transportation Information System
RSE	Roadside Equipment
RSU	Roadside Unit
RWMP	Road Weather Management Program
SAE	Society of Automotive Engineers
SAFETEA-LU	Safe Accountable Flexible Efficient Transportation Equity Act
SHRP2	Strategic Highway Research Program 2
SOV	Single Occupancy Vehicle
STIP	State Transportation Improvement Program
TAT	Traffic Analysis Tools
TIP	Transportation Improvement Program
TMC	Transportation Management Center
TOPS-BC	Tool for Operations Benefit/Cost
TPM	Transportation Performance Management
TRB	Transportation Research Board
TSMO	Transportation systems management and operations
USDOT	U.S. Department of Transportation
V/C	Volume-to-capacity ratio
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Device
VMS	Variable Message Sign
VMT	Vehicle Miles Traveled
WSDOT	Washington State Department of Transportation

INTRODUCTION

BACKGROUND: A BRIEF HISTORY OF SURFACE TRANSPORTATION SYSTEM ANALYSES FOR OPERATIONS

Conducting a time-dynamic operational simulation analysis of a transportation system was a novel endeavor 35 years ago. With personal computers, continuous loop detector data, and new software tools, codifying basic driver behaviors enabled analysts to understand and optimize the transportation system. A new capability emerged—we could examine a transportation system as it evolved over time over the course of a day. Analysts could consider animations of vehicles approaching and departing intersections, choosing lanes, and accelerating and decelerating to maintain safe distances across a virtual roadway. Structured analyses could examine the onset, nature, and duration of congestion within a virtual system in a way that resonates with the actual experience of the transportation system user.

New notions of optimal performance influenced by travel time and travel-time reliability became central concepts of system management. The environment for the 20th Century pioneers was exciting and intellectually challenging, but these emerging analytic capabilities had significant limitations for broad practical application to support transportation system management. Software tools were in primitive states of maturity, often derived from academic research efforts. Traveler behavior models used simple relationships based on what limited data were available, or filled in with best-available assumptions where empirical experiments had not been conducted. Field data to calibrate models of roadway systems were often scarce, incomplete, and riddled with errors. Computer platforms were unable to support the representation of large, complex and multi-modal networks—or required impractical amounts of time for these models to run to completion.

Since the early days of exploration and invention, there has been an evolution in the environment for 21st Century transportation systems analysts. This transformation has been driven by, among other factors, improvements in commercially supported software, low-cost access to advanced computational platforms, and a data-rich environment. Conducting some less complex experiments has evolved into standardized practices that inform some routine decisions (e.g., intersection warranting and interchange justification analyses). Forward-thinking analysts have attempted to push conventional boundaries of network scope and analytic purpose (e.g., multi-modal analyses and corridor and sub-regional analyses). While a history of partial successes and near-failures can be recounted, collectively, the field eventually cast aside a number of these limitations. Now, the scope of operational analyses is much larger than it was in the past—and the demands on analysts to address more difficult questions with even more complex modeling approaches continue to grow.

While these technological changes have empowered the 21st Century analyst, the gains made by the individual analyst may have outstripped the ability of the organizations managing transportation systems to capitalize on new analytic techniques. The transportation analyst began work in isolation as a researcher/super-user; to some extent, this notion of isolated “nerds in the basement” conducting analyses still persists—with analysts ensconced in specific consulting firms, analytic departments, and graduate school laboratories. A sociologist might easily distinguish a subculture of modeling and simulation experts gathered around a specific tool and associated networks, largely insulated from the mission of improving surface transportation system performance. The analyst is rarely involved in diagnosing transportation system problems or using data to support analytic project scoping. Isolated champions with the vision and institutional connections to bring analytics to bear in decision-making have made important contributions, but these connections tend to last only as long as the champion remains engaged. With rare exceptions, there is a lack of advanced institutional models to systematically and consistently leverage the power of transportation analytics embedded within the broader transportation system management mission.

Funding transportation analysis can provide valuable insight into the potential benefits and costs of transportation investments. The general value of conducting analysis is the extent to which it assists stakeholders to:

- **Invest in the right strategies**. Analysis offers transportation agency managers a predictive forecasting capability that helps them determine which combinations of transportation investments are likely to be most effective and under which conditions: Analysis helps decision-makers identify technical and implementation gaps, evaluate different strategies, and invest in the combination of strategies that would minimize congestion and produce the greatest benefits. Comprehensive analysis increases the likelihood of transportation improvement success and helps minimize any unintended consequences. It affords an enhanced understanding of existing conditions and deficiencies, improving our ability to match and configure proposed strategies to the situation at hand.

- **Proceed with confidence**. With reliable analysis, transportation agencies can "see around the corner" to identify optimal strategic combinations, as well as the potential conflicts or unintended consequences of certain combinations of strategies that would otherwise be unknown before full implementation. Effective analysis enables managers to estimate the benefits across different transportation modes and traffic control systems and to align these estimates with specific assumptions about existing conditions and improvement strategies. If they cannot predict the effects of proposed strategies, transportation agencies may not take the risk of making the institutional and operational changes needed to optimize transportation operations.

- **Reduce risk associated with implementation**. Armed with this analysis, stakeholders can develop detailed concepts of operations and requirements. The information also helps managers define and communicate key analysis questions, project scope, and partner roles and responsibilities. For operational improvement projects, developing an analysis plan may help identify flaws or technical issues in the Implementation Plan or Concept of Operations that would have been otherwise overlooked. Following the analysis methodology helps to communicate the scope of the project and appropriately set expectations among differing project stakeholders—planners, operators, data analysts, modelers, and agency management from state, local, and/or regional transportation agencies, etc.—and provides a clearer definition of expected roles and responsibilities. Analysis also helps managers identify and prioritize resources to project objectives, allowing for the effective and efficient allocation of resources and sounder project management.

OBJECTIVES AND VALUE OF TRANSPORTATION ANALYSIS FOR OPERATIONS

Objectives of This Document

This document provides guidance on how 21st Century transportation analytic resources—data, tools, and computational platforms—can be systematically embedded within the transportation system management process, resulting in:

- Enhanced characterization of transportation system dynamics and problem diagnosis.
- Improved analytic project scoping.
- Data-driven experimental design that limits risk and maximizes analytic insight.
- Systematic execution and documentation of analyses to preserve accrued analytical capital.

Transportation system management organizations that successfully achieving these objectives will realize a number of important positive outcomes related to greater insight, better analyses, and reduced costs and risks:

- Increased understanding of how the system as a whole operates and changes over time.
- More relevant and targeted transportation analytics.
- Minimized redundancy and technical risk in a portfolio of analytical projects.
- Reduced costs of conducting analyses over time.

Definition of Operational Analysis

This section defines the scope of an operational transportation systems analysis. Operational analyses are inherently time-variant, with explicit treatment of time within an analytical tool rendered in time steps bounded above by an hour and below by fractions of a second.

An operational analysis, therefore, includes all traffic micro- and meso-simulations, some time-dynamic macro-simulations, dynamic transit operations models, activity models, time-dynamic applications of statistical tools, prediction constructs, communications models, and integrated modeling frameworks combining these classes of tools.

Time Dynamics

Why are time-dynamic analytics critical? Because a busy modern transportation system is inherently a dynamic entity. It never exists in pure equilibrium, and is in a state of perpetual change at multiple temporal wavelengths: minute-to-minute, hour-to-hour, peak-to-nonpeak, day-to-day, seasonally, and year-to-year. A time-dynamic view of the network informs a fundamental element of transportations systems management, namely, that system management is essentially a task of managing change. Without time-dynamics, the view is critically skewed, and all decisions made regarding transportation system management will suffer from what is essentially a self-limited, stunted understanding of the true nature of the system.

The time-dynamic view most closely represents the way the traveler (or system user) experiences the transportation system. For the system user, each day represents a risk in missed appointments and wasted time. For many system users (both travelers and those who move goods from place to place), predictability of travel and the reliability of travel over repeated interactions with the system is paramount. The time-dynamic system view captures both the within-the-day experience of travelers seeking to use the system without excessive delay and the longer-term experience of system users seeking to manage uncertainty and variability in system performance.

Influence of Operational Analyses

Operational analyses can inform decision-making on multiple time scales: real-time, operational planning, planning for operations, and long-range planning. The incremental time-step within the time-dynamic representation of the network may be very small compared to the scale of decision-making. Understanding how congestion arises, propagates, and subsides within a system can inform decisions made in the next ten minutes, operational changes to be deployed over the next ten hours, investments made to improve operations over the next ten months, or major improvements to be phased in over 10 years or more. Later in this introduction, this guide will identify ways that operational analyses inform decisions at various time scales; we use examples throughout the document to show the breadth of transportation system management decision-making, potentially enhanced with time-dynamic analyses.

TRADITIONAL AND EMERGING CAPABILITIES

Emerging Capabilities

There are three critical aspects of the resources available to the 21st Century transportation systems analyst: new data sources and data volume, more powerful computational platforms, and increasingly complex and capable visualization and analytical tools.

The 20th Century analyses were constrained by the data, computational platforms, and tools available at the time. The abstraction of the system required to perform predictive analytics—particularly with respect to time-dynamics—were relatively severe. Lack of data forced analysts to focus on average data from disparate sources and create time-variant “normal” condition days that by definition excluded outlier days with incidents, weather or unusually high (or low) travel demand. For the largest geographic scope (e.g., four-step planning modeling), time-dynamics were removed all together.

With 21st Century capabilities, the analyst can cast aside severe abstractions to realistically and accurately represent the time-variant system in a new and powerful way. Continuous data collection captures outlier conditions as readily as any other operational condition. Computational platform improvements allow more runs in the same amount of time, rendered in higher detail—detail that can be analyzed and visualized in increasingly varied ways to serve multiple needs and provide insights. This richer representation of time-dynamics in the system allows the analyst to understand the transportation system the way a user experiences the system—over time, in repeated trips of varying nature, with good days and bad days—and can characterize performance across all conditions (both good and bad).

Data

Compared to the world 35 years ago, current analyses are conducted in a data-rich environment. Sources of data have expanded from infrastructure-based detection systems to include probe data (vehicles or mobile devices). More data are continuous—available at periodic intervals around the clock—rather than associated with a one-off data collection effort. More data sources reveal the system condition from the user-perspective, expanding the potential views from the traditional collection of facilities (infrastructure-based) to a collection of trip-making activities (user-based). Looking forward, there is no indication that these broad trends will decline, given the increasing number of data sources, more rapid data delivery, and a broader array of types of data. The advent of connected vehicle technologies, the Internet of Things (IoT), and an increasingly engaged and connected traveler will create new opportunities to understand system dynamics from multiple vantage points. Most importantly, these data sources allow analysts to characterize system performance in new and useful ways. Rather than simply examining average delay, they can characterize travel reliability, on-time performance, trip predictability, and other measures that capture the full range of options and experiences users encounter.

Computational Platforms

In 1965, Gordon Moore, co-founder of Intel, authored Moore's Law, a prediction that the computational power of analytical platforms will double every two years.[1] Fifty years later, this observation still holds true. The desktop or cloud-based computational tools at an analysts' disposal are more powerful and more ubiquitous.[2] That said, legacy transportation software packages are not inherently high-performance computational tools, and tend to be single-processer–based; the computational power available to the analyst may be limited in many cases since scalable solutions are not directly realized in most analytical tools. In this case, the improvement in power will be largely realized in data storage and visualization, which is inherently more scalable than discrete time-step simulations, for example.

Analytics/Visualization Tools

Increasingly complex tools will be increasingly capable. However, the time required to master and understand these complex tools has also increased. As the complexity grows, it becomes harder to manage and estimate all the detailed parameters needed to drive the tools effectively. Analysts must master individual tools at the same time that they understand the strengths and weaknesses of various tool types. In some cases, a simpler tool may actually provide greater insight than attempting to model at the smallest time step with the most parameters in play.

---

[1] Moore's Law and Intel Innovation, http://www.intel.com/content/www/us/en/history/museum-gordon-moore-law.html.

[2] Moore's Law Keeps Going, Defying Expectations, http://www.scientificamerican.com/article/moore-s-law-keeps-going-defying-expectations/.

CAPTURING SYSTEM DYNAMICS

The previous two sections describe how to define time-dynamics and why that is important. The other key concept to be clear about is the system—the thing that is to be managed, modeled, and—hopefully—optimized.

Why is a System-View Important?

A typical pitfall for analytic projects is that the underlying nature of the problem is not well known until the project is over. Often, only near the end of the project does the analyst focus on what specific system is being managed, modeled, and optimized. When it is unclear among multiple stakeholders what system is being modeled, the analyst must often "fill in the blanks" and define the system in modeling. An analytic project may be presented as a pure freeway merge/weave analysis, but it turns out that signal timing on nearby arterials produces dense platooning on the on-ramps that essentially create the merge/weave issue. The "system" here is the combined freeway-arterial roadway system, not just the freeway. An analysis of the freeway alone misses the essential point of the dynamics of the combined system. This example is quite tactical, but there are similar examples regarding individual elements of integrated multi-modal corridors, sub-regional analyses, and multi-state freeway corridor analyses.

What is the System?

The system is the collection of facilities, fleets, infrastructure, and trip-making users for which the system manager is responsible ***plus*** all interacting systems that influence the performance of the system for which the manager is responsible. When defining the system, it can be useful to examine a number of boundaries:

- **Geographic**. Any transportation system is always influenced by its neighbors. Treating a subset of the transportation system as an independent element is a typical and limiting assumption made by countless transportation analysts.

- **Temporal**. Under specific conditions or at times associated with particular events, other stakeholders may be in the transportation system itself and altering system controls, e.g., law enforcement.

- **Jurisdictional**. Even within a geographic boundary, localities and other organizations may play a critical role in influencing the transportation system. Simple examples include signals controlled by a locality in a long corridor, parking facilities, and large employers.

- **Functional**. The system is also defined by where and when managers have functional control, as well as by the limitations of these controls. Some functions may be automated, while others may offer the manager a set of number of pre-determined choices. Policies may require the manager to operate only within a defined range; dynamically priced toll facilities may be constrained to never fall below a particular figure or to exceed a maximum rate. Functional boundaries can also be identified considering how data flows to and from entities associated with a transportation system. In some cases, the issue is not limitations in system control but limitations in data. Data-related boundaries may limit the system manager's ability to understand the state of the system, or neighboring systems; these are sometimes referred to as situational awareness constraints.

- **Modal**. A transportation system may contain interacting sub-systems associated with alternative transportation modes—transit, freight, High-Occupancy Vehicle (HOV) lanes, pedestrian networks, and bike lanes. The system definition may focus on an individual transportation mode or the integration of all systems depending on the project scope. Whether the focus is at the subsystem or system level, a systems definition must recognize the interactions among these systems and subsystems.

Who Owns the System—What is the Span of Control?

After defining the system, it may become clear who really needs to be included in order to adequately address systematic issues and improve system performance. Once the full "system" is defined (in Module 4), we may end up with a new system concept with no single owner; this is highly typical for complex surface transportation systems that must interact. The presence of multiple overlapping spheres of control is a critical factor in nearly all system analyses. What gets optimized in each system relates to overall performance for everyone, but what is good for the goose may not always be thought of as good for her neighbor, the gander. Even if there is no clear system czar, there still can be an understanding of how systems interact. Collectively, managing the broader system is always preferable to uncoordinated local optimization.

Analysts can play a critical role in assembling a community of organizations and decision-makers that manage the shared system. They can help frame the system concept—using the system definition to bring together the community—and show interactions within and across boundaries. They can also use data to visually describe the system, which can be a powerful tool in energizing stakeholders by tangibly illustrating the concept of a collectively managed system. Analysts often use time-dynamic tools to put the conceptual system into motion, showing when and where the system performs poorly, and engaging stakeholders to work together to provide more effective collective system management. A notional system is much more powerful if it can be clearly defined, observed (using data), and studied (using both data and tools).

System of Systems Effects

One example where a system was positively influenced by analytics regards a mixed freeway/arterial corridor where the main arterial route had multiple jurisdictions, each in control of a particular subset of all the intersections in the corridor. Everyone agreed that congestion in the corridor, particularly on the arterial route, was a serious issue. Arterial travel had become so congested and so unpredictable that travelers complained and small municipalities along the route were worried that shoppers were avoiding doing business in their area because of the hassle of dealing with the congestion. The state department of transportation organized a stakeholder group and initiated a micro-simulation analysis. Only after the model was constructed and some initial runs conducted did it become clear that many signals in the system were optimized for access to/from shopping along the route, rather than for travel along the route, particularly from large parking facilities. Since the policies of these local sub-systems were the largest influencers of arterial performance, adapting other "major" intersection signal control would have limited influence. A workable solution was found by harmonizing turning movement release at major intersections to coincide with demand at/around the major parking facilities. This insight came somewhat late in the effort, so the original analytical model had to be extensively reworked and augmented to reflect the nature of the interacting systems. The system turned out to be highly dependent on the large parking facilities, and effective management of the system relied in large part on incorporating time-dynamic effects of this previously ignored parking sub-system.

THE 21ST CENTURY ANALYTICAL PROJECT SCOPING PROCESS

This guidance supports transportation professionals—including data managers and transportation analysts—at different levels of technical expertise and in a wide variety of uses across short-, medium- and long-term decision horizons. To effectively identify, scope, and conduct data analytics projects, a project scoping cycle provides a set of standard data-driven processes to reveal insights effectively (and cost-effectively). The defined cycle of steps in the 21st Century Analytic Project Scoping Process underscores the limitations of the past and offers a new way forward as a part of a Continuous Improvement Process (CIP). The process provides a common data-driven analytic framework adapted from modern systems engineering that is designed to enable the demands of modern reliability-focused analytics and alternatives analysis.

Data managers often fall into the trap of capturing, assembling, and managing data without a comprehensive understanding of the underlying analytical uses for these data or understanding that data itself can yield insights when properly integrated. Therefore, the scoping process begins by characterizing key system performance measures that reveal underlying transportation system dynamics, shifting day to day and year to year, subject to many external influences. Data needs are derived both to calculate these measures and to provide the context for the variation (e.g., variation in demand, incident patterns, and weather impacts). Data are then readied for analysis through a focused and resource-sensitive quality control process. These data also drive a series of characterizing analytics that inform and shape the project-scoping activity.

Project scoping includes problem statement and hypothesis development, spatial and temporal dimensions, and tool selection—choosing from the full gamut of potential analytical methods from statistical analysis to various alternatives among simulation techniques. Analytics includes creating a detailed experimental plan that accurately reflects a control (baseline) and experimental (alternative) structure; this plan is evaluated over a level playing field derived from a representative set of varied operational conditions.

Documentation of the data and analysis results is the key component to complete the analytics project cycle and to prepare for the next cycle. The next cycle stage is to circle back to identify what can be statistically derived from the analytics and how these results inform a new set of data needs and further analytic work. Figure 2 (on the following page) illustrates the 21st Century Analytic Project Scoping Process detailed in this guidance, a four-module CIP:

- **Module 1: System Diagnostics**. Characterize system dynamics and diagnose problems.
- **Module 2: Scoping**. Perform data-driven transportation analysis project scoping.
- **Module 3: Preparing Data**. Collect and organize the data needed to conduct a transportation analysis.
- **Module 4: Analysis**. Conduct and document transportation analyses.

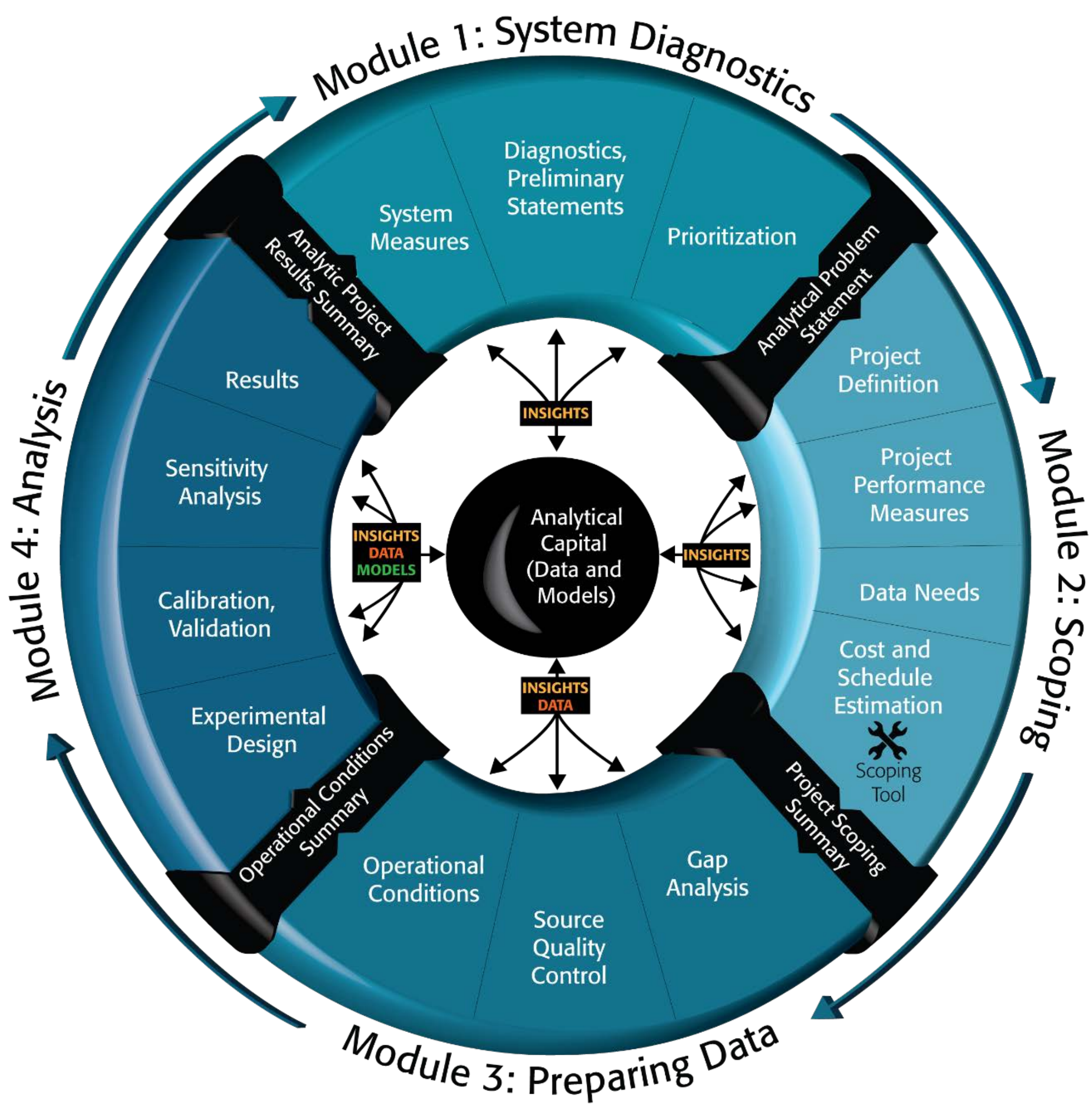

Figure 2. Diagram. 21st Century analytic project scoping process.
(Source: Federal Highway Administration.)

Module 1 discusses system performance measures and early diagnostic activities to help analysts develop preliminary analytic problem statements and prioritize the identified problem statements with a risk-reward project screening approach. The final product of this module—one or more individual Analytical Problem Statements associated with a high-priority concept—serves as the foundational connecting document for Module 2: Scoping.

Project scoping includes a more detailed project definition, the identification of project-specific performance measures, a refinement of mitigation strategies and data needs, tool selection, and cost and schedule estimation. A Scoping Tool to help the analyst complete this final step is included as a part of this guide.[3] Module 2 ends with the completion of the Project Scoping Summary, which provides enough information to initiate data preparation and analysis work (Module 3).

A data analyst in Module 3 ensures the consistency and quality of the data available and outlines the data collection plan to fill any gaps between data needs and availability. Depending on the nature and scope of the project, the data analyst may analyze the available data to identify and summarize a representative set of operational conditions, critical to the creation of a strong analytical plan in Module 4.

In Module 4, an analytical project is executed. The analyst creates a detailed experimental design with experimental and control cases, then calibrates and validates models under various operational conditions and conducts a sensitivity analysis. Documentation of analysis results is the last step—to document project findings that inform decision-making and to capture lessons learned to improve the agencies implementation of the 21st Century analytic project scoping process itself for the next cycle of analytic work.

In the center of Figure 2, the 21st Century analytic process includes maintenance of analytical capital built up from multiple projects. These can be simply thought of as improved analytical tools, system models created as inputs to these tools, and the data collected to assess system performance or characterize operational conditions. However, a third, often-overlooked element is the insight gained along the complete life cycle of the analytical project with respect to lessons learned in all four steps: diagnostics, scoping, data preparation, and analytics. If the individuals performing these steps do not record key insights and move on to new roles in other organizations, these insights are nearly always lost and a new generation of analysts are likely to repeat mistakes made in future analytic work.

Data-driven analytics enable improved decision-making at different time scales, short-term (transportation system monitoring and real-time operational management), medium-term (Transportation System Management and Investment Planning) and long-term (transportation system planning and long-term trends). Each of these time scales place different demands on the data manager and have distinct data needs.

Transportation System Monitoring and Real-Time Operational Management

The cycle of transportation system monitoring and real-time transportation operational management projects can be conducted in real time—every minute, hourly, or daily, depending on the scope of the project. Given the short feedback timeframe, most settings need to be automatic or semi-automatic. Examples are ramp metering and traffic signal control systems. The system can collect and store real-time data and built-in algorithms to conduct data analysis and can store the analysis results. A routine self-diagnostic function can be made automatically for real-time system adjustments or manually on a daily or weekly basis for short-term system adjustments.

---

[3] The scoping tool is available at the U.S. Department of Transportation Open Source Application Development Portal (OSADP) Web site: http://www.itsforge.net/.

Transportation System Management and Investment Planning

From the transportation system management aspect, it takes days, weeks, or months to repeat the project cycle. A work zone alternative analysis project—a type of transportation system management project very similar to the transportation system monitoring project—takes more time to complete the cycle. The system collects daily time-dependent traffic data and compares day-to-day traffic patterns, weather, and incident impacts with different alternatives. The system can identify the best alternatives based on the previous analysis results and current operational conditions. The data and the analysis results are stored properly for the next round of work zone activities to both save the cost and gain lessons learned. It could be documented as a standard procedure when conducting a work zone alternative analysis.

Transportation System Planning and Long-Term Trends

This long-term category includes one-to-three year transportation system planning (e.g., a freeway lane expansion) and over-ten-year long-term trends (e.g., Metropolitan Planning Organization's [MPO's] Metropolitan Transportation Plan [MTP]). The Federal Highway Administration (FHWA) published a guidebook of a data-driven, strategic approach for Departments of Transportation (DOTs), MPOs and other planning organizations to make investment and policy decisions to attain desired performance outcomes for the multimodal transportation system.[4] The data-driven/performance-based transportation plan includes a discussion of conditions and performance of the transportation system over several years. As this diagnostics occurs through multiple cycles, the plan serves as a baseline for developing and refining plan goals, objectives, and targets. Information from the performance report can support refinement of targets associated with the timeframe of the transportation plan, as well as near- or mid-term targets.

Relations between Different Time Scales

The MTP is a blueprint that contains a 20-30 year planning horizon and is updated every 3-5 years to identify current and future transportation corridors, and forecast transportation demand, needs and growth patterns. Based on MTP, a 3- to 5-year State Transportation Improvement Program (STIP) or Transportation Improvement Program (TIP) is developed to identify the highway and transit improvement plan and identify short-term priorities with funding sources. The performance targets set in the long-term plan are evaluated during both short-term and mid-term cycles. The analysis results from short-term and mid-term cycles also feed into future long-term plans as part of the system diagnostic process. Figure 3 illustrates interactions among these three time scales. The solid lines represent the project identification and development from transportation analysts. Dashed lines represent insights gained and quantitative feedback from relatively shorter-term analyses that influence relatively longer-term project identification and development.

[4] Federal Highway Administration, MODEL LONG-RANGE TRANSPORTATION PLANS: A Guide for Incorporating Performance-Based Planning, August 2014.

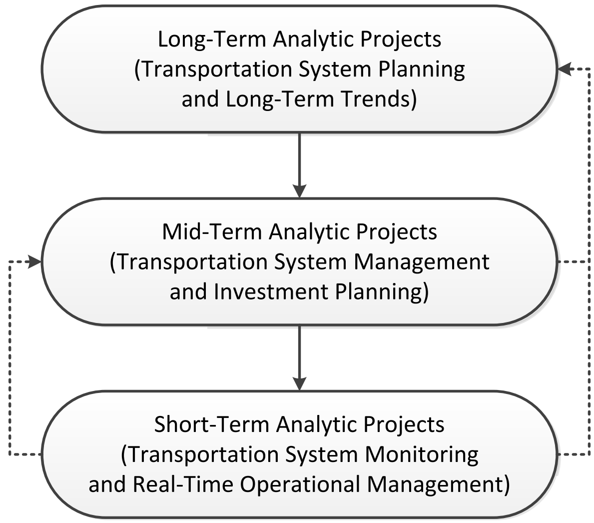

Figure 3. Flowchart. Interactions among analytic projects with different time scales.
(Source: Federal Highway Administration.)

A freeway segment expansion is identified with projected performance targets in the TIP based on the demand growth and need prediction in MTP. This decision then goes to the mid-term transportation system management plan to conduct the alternative analysis for the freeway construction period. During the construction period, the short-term system monitoring and real-time management takes place to adjust the system in real time or a short time period. Daily, weekly, or monthly system performance and adjustment reports serve as a basis to exam and adjust the system management plan for the next few months. The system management plan documents lessons learned and key findings serve the future TIP and MTP in the next scoping cycle.

RESOURCES

The following is a list of resources related to the planning, development, and implementation of an Intelligent Transportation Systems (ITS) project in the 21st Century. Each resource includes the date if available and brief summary of its content:

- *Programming for Operations: MPO Examples of Prioritizing and Funding Transportation Systems Management & Operations Strategies (September 2013).* This document focuses on the programming phase of transportation decision-making in metropolitan areas. By examining case studies from nine MPOs, it highlights best practices related to programming Transportation Systems Management and Operations (TSMO) strategies.

- *Transportation Planning For Operations: Quick Guide to Practitioner Resources (September 2013).* This document provides a quick reference for transportation professionals to find the best resource for planning management and operation of transportation facilities. References listed include guidebooks, case studies, and workshops. It provides sample questions that might be asked during the transportation planning process and suggested resources for finding the answers.

- *Getting the Most from Your Transportation System Investments: Operating for Peak Performance*. This brochure highlights the benefits and advantages of using a performance-based approach to integrate management and operations into transportation planning. It also provides questions that help guide the transportation professional in using this approach for their region.

- *Designing for Transportation Management and Operations: A Primer (February 2013)*. This primer describes tools and institutional approaches to help transportation agencies consider operations during the design process, known as "designing for operations." This process requires the development of design policies, procedures, and strategies that support both operations and transportation management. The primer presents design considerations for various types of strategies and collaborations between practitioners from multiple transportation agencies.

- *Systems Engineering for Intelligent Transportation Systems (January 2007)*. This document provides an introduction for transportation professionals to system engineering and how it can be applied to intelligent transportation systems projects. Included are step-by-step descriptions of the system engineering life cycle and guidance on how to apply these principles to ITS projects.

- *Systems Engineering Guidebook for ITS (November 2009)*. This guidebook applies system development process activities from information technology, Department of Defense, Mil-Aerospace, and the automotive industry to ITS projects. It was written to help support the development of ITS projects, including developing requests for proposals, assessing capabilities of potential systems managers, and supporting development teams in the implementation of ITS projects. All major phases of project development are extensively covered, as well as several case studies and lessons learned.

- *Integrating Demand Management into the Transportation Planning Process: A Desk Reference (August 2012)*. This desk reference provides information on integrating demand management into the transportation planning process. Two major areas of coverage include policy objectives and scope of the planning effort. This reference discusses how demand management might be integrated into four levels of transportation planning from the state down to the local level. It also provides information on tools available for evaluating demand management measures and on the known effectiveness of these measures.

- *Longitudinal Study of ITS Implementation: Decision Factors and Effects (April 2013)*. This study discusses decision factors and valuation of benefits for implementation of ITS technology to define considerations for next generation ITS and the connected vehicles. Data for this study was gathered through literature reviews, stakeholder interviews (public sector, automobile manufacturers, and trucking agency), post-hoc analysis, and workshops. The findings reveal underlying characteristics and factors for ITS technology adoptions, deployment, and benefits.

- *The Regional Concept for Transportation Operations: A Practitioner's Guide (July 2011)*. This guide shares successes and lessons learned from four metropolitan regions as they developed Regional Concepts for Transportation Operations (RCTO). This guide highlights specific practices to help overcome challenges faced during the RCTO development process. It is designed to help ITS professionals select the most effective methods for their own region's transportation performance.

- *Management & Operations in the Metropolitan Transportation Plan: A Guidebook for Creating an Objectives-Driven, Performance-Based Approach (November 2007).* This document provides guidance and strategies on the metropolitan transportation planning process with a focus on Management and Operations (M&O). These guidelines are designed to help fulfill Safe, Accountable, Flexible, and Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) requirements and also result in the most benefit for transportation investments. It describes planning frameworks, engaging stakeholders, using operations objectives to develop performance measures, and determining outcomes.

- *Operations Benefit/Cost Analysis TOPS-BC User's Manual: Providing Guidance to Practitioners in the Analysis of Benefits and Costs of Management and Operations Projects (June 2013).* This manual provides instruction on the use of the Tool for Operations Benefit/Cost (TOPS-BC), a tool developed as part of the FHWA Operations Benefit/Cost Desk Reference Project. TOPS-BC provides key decision support capabilities, including life-cycle cost estimating, spreadsheet-based Benefit/Cost (B/C) analysis, and others.

- *Regional Concept for Transportation Operations (RCTO): The Blueprint for Action—A Primer (June 2007)*. This document explains the RCTO, its development, and applicability. It introduces the RCTO as a management tool that can increase efficiency and effectiveness of collaboration during an ITS project. The essential components and potential role of the RCTO in the transportation planning project are laid out and demonstrated through multiple examples. Additionally, this primer highlights benefits and keys for success of an RCTO.

- *Statewide Opportunities for Integrating Operations, Safety and Multimodal Planning (May 2010).* This reference manual provides how-to information for transportation professionals on integrating operations, safety, and multimodal planning. It identifies various opportunities and approaches at various levels of decision-making. It also highlights the significant role of data collection and performance measures at the state, regional, sub-area, and project levels.

- *Performance-Based Planning for Small Metropolitan Areas (December 2014)*. This report outlines performance based planning by MPOs in urbanized areas with populations less than 200,000. Insights are collected from existing best-practices research and interviews with small MPOs and DOT partners engaged in metropolitan performance based planning across the country. Two case studies are included as models of the benefits of performance-based planning.

- *Intelligent Transportation Systems Benefits, Costs, and Lessons Learned: 2014 Update Report (June 2014)*. This report collects material from three related Web-based databases, known collectively as the ITS Knowledge Resources, into a group of factsheets presenting information on the performance of deployed ITS. There are a total of 20 factsheets covering the benefits, costs, and lessons learned of the performance of deployed ITS technologies and applications.

- *Traffic Analysis Toolbox Series*. The FHWA Office of Operations developed the Traffic Analysis Tools Program to balance efforts between developing new, improved tools in support of traffic operations analysis and efforts to facilitate the deployment and use of existing tools. There are two tracks—one for development and one for deployment. The Development Track focuses on developing and enhancing current models that are easier to use, more robust in their application, and more reliable in their results. The Deployment Track has 14 volumes of guidance, including an overview of traffic analysis tools, a selection guide, guidelines for use, and insights for best practices, methodologies, applications, benefits, and cost.

- *Traffic Analysis Toolbox: Volume XIII: Integrated Corridor Management Analysis, Modeling, and Simulation Guide (May 2012)*. The purpose of this guide is to help corridor stakeholders implement the Integrated Corridor Management (ICM) Analysis, Modeling, and Simulation (AMS) methodology successfully and effectively. This guide includes a step-by-step approach to implementation of the ICM AMS methodology and lessons learned in its application to the three ICM Pioneer Sites and a test corridor.

- *Traffic Analysis Toolbox: Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software (July 2004)*. This report describes a process and acts as guidelines for the recommended use of traffic microsimulation software in transportation analyses. Guidelines include an example problem walked through the seven steps of microsimulation analysis: (1) scope project, (2) data collection, (3) base model development, (4) error checking, (5) compare model Measures of Effectiveness (MOEs) to field data (and adjust model parameters), (6) alternative analysis, and (7) final report. Each step is described in detail and illustrated with examples.

- *Guidebook to Develop Methods/Analytics for Summarizing and Analyzing Data to Support Transportation Operations Decision-Making (anticipated late 2016).* This guide will present material on how to collect, store, manage, mine, analyze, and interpret "big data" to create valuable intelligence, to support the transportation operations decision-making process. This Guide will provide information on how to convert information from data into knowledge and how to explore data to find consistent patterns, associations, and relationships between variables. Material will include how to collect, store and manage data and how to analyze data, create data analytics, and make strategic predictions. The guide will then identify and define important variables for analysis in support of transportation operations decisions.

ISSUES/CHALLENGES

A complex transportation system—such as an interstate freeway corridor or a transit system—involves multiple jurisdictions (e.g., state and local agencies or a state highway administration and a transit administration) and multiple stakeholders (e.g., private companies, public agencies, and academic/ research institutes). One of the major challenges in managing such a system is the amount of effort spent coordinating stakeholders. For example, from a state point of view, optimizing a corridor including the surrounding neighborhoods may be a centerpiece of concepts to improve corridor performance. However, for local agencies, such as a county or city, local signal timing optimization is related only to localized (sub-system) optimization. System-wide optimization may be inimical to local optimization and vice versa.

Note that sub-system optimization can be technically challenging but potentially counter-productive. For example, ramp metering controls may seek to increase throughput on the freeway mainlines—even if this implies extra delay at nearby arterial intersections. Similarly, the arterial signal optimization may not take into account impacts on freeway throughput.

Managing the system does not mean running roughshod over the goals of some stakeholders to benefit others. Collective system management includes finding common ground regarding goals for the system and showing why *ad hoc* local sub-system optimization is worse for everyone than coordinated, optimized system management. Armed with a solid system definition, the data to describe and illustrate the system, and the insight gained on the time-dynamic nature of the system, the analyst can play a critical role in addressing these challenges.

NEEDS-DRIVEN, DATA-DRIVEN ANALYTICS

An underlying theme in this document is the importance of limiting the isolation of transportation analysts. Analyst can bring valuable insights to the overarching process of systems management. With better integration, analytical needs are better defined, better analytical projects are defined, and more insight are gained per dollar spent.

This guidance provides an analytical project scoping cycle to connect analysis results and system characteristics and diagnostics. Both the transportation decision-makers and the data analysts must fully understand the nature of system performance issues and participate in identifying and refining analytic project concepts. In order to use data and transportation analyses as an integral part of a continuous system improvement process, analysts must play an integral role in systems management. They can be engaged in identifying the nature of the problem and possible root causes and then conduct analyses that return insight to the decision to change practices or invest in longer-term solutions. Agencies that engage in a continuous, modern, data-driven analytic project scoping process can expect to see fewer redundant analytical projects, fewer failed analytical projects, and improved decision-making fueled by the insights from relevant and targeted analytical efforts.

NAVIGATING THIS DOCUMENT

This guidance document describes a four-element CIP best deployed as a complete process with all four modules. In general, the authors suggest that readers consider the guidance in typical serial order, that is, reading the introduction first, then the four modules in order. If readers want to refer to this guidance for current projects that are already in latter stages (e.g., data analysis or execution), we suggest reading the introduction before moving to the relevant module to gain insights for these projects. The strongest use of this guidance is not for a single project—but rather for managing a portfolio of analytic projects over time and leveraging past data and models in developing, scoping, designing, and executing future projects.

MODULE 1. CHARACTERIZING SYSTEM DYNAMICS AND DIAGNOSING PROBLEMS

Module 1 characterizes the performance of the transportation system and identifies and diagnoses system problems (see Figure 4 on the following page). Is the system performing as expected? Is travel in the system safe, reliable, and predictable? Module 1 also considers the role of analytical projects: What types of analytical projects can be successfully conducted in support of improving system performance? Analysts are often not directly involved in this part of the system management process. In many cases, the analyst becomes involved after an analytic project concept has been developed. More precisely, an analyst may be asked only to respond to an assignment to conduct a specific analytic project. Too often, they are not engaged in characterizing the performance of the system or offering insights on the problems within the system. This is unfortunate, since analysts work directly with the data and tools that provide insight into the underlying nature and dynamics of the transportation system.

If an analyst has just completed a reliability analysis of trips within the system and potential mitigation strategies, the insight gained from this analysis may be critical for understanding the complex interaction of factors driving delays and outlier travel times in the system. System users may experience a lack of travel-time reliability, and system manager may observe that the system has “bad days”. However, the analyst who has systematically assessed the root causes of unreliability, can point to specific conditions (low visibility, spotty slick conditions, incidents in particular hot spots, and shifted travel demand based on delayed area-wide school start times) and the resulting impact on travel-time variability. Any discussion of improving system reliability occurs at a deeper, more informed, more data-driven level. To implement the 21st Century Analytic Project Scoping Process as a continuous process, key inputs for characterizing the system dynamics include the data used in related (previous and current) analyses, the hypotheses and findings of these studies, and the insights gained by the analysts who conducted these studies.

As illustrated in Figure 4, this module describes a step-by-step process that engages the 21st Century analyst and brings analytical capital to bear to enhance and improve subsequent analyses, system management, and system performance. First, we discuss the nature of system performance measures that reflect system products and by-products—and how to reconcile differing stakeholder viewpoints of what “good” performance looks like. Second, we discuss how time-variant profiles of key measures can be useful in bringing depth and clarity to deliberations on system performance and how early diagnostic activities can guide the development of preliminary analytic problem statements. Third, these concise statements of underlying need and analytic intent are a critical building block used to intelligently integrate, characterize, and prioritize analytical projects prior to larger-scale scoping and execution. It is here, in the earliest origins of individual analytical projects, that poor decisions are often made that greatly reduce the value of downstream analytical activity. Effective planning for analytics is a project portfolio management effort, ensuring that the most critical needs are met while resources and technical risk are managed comprehensively across all projects.

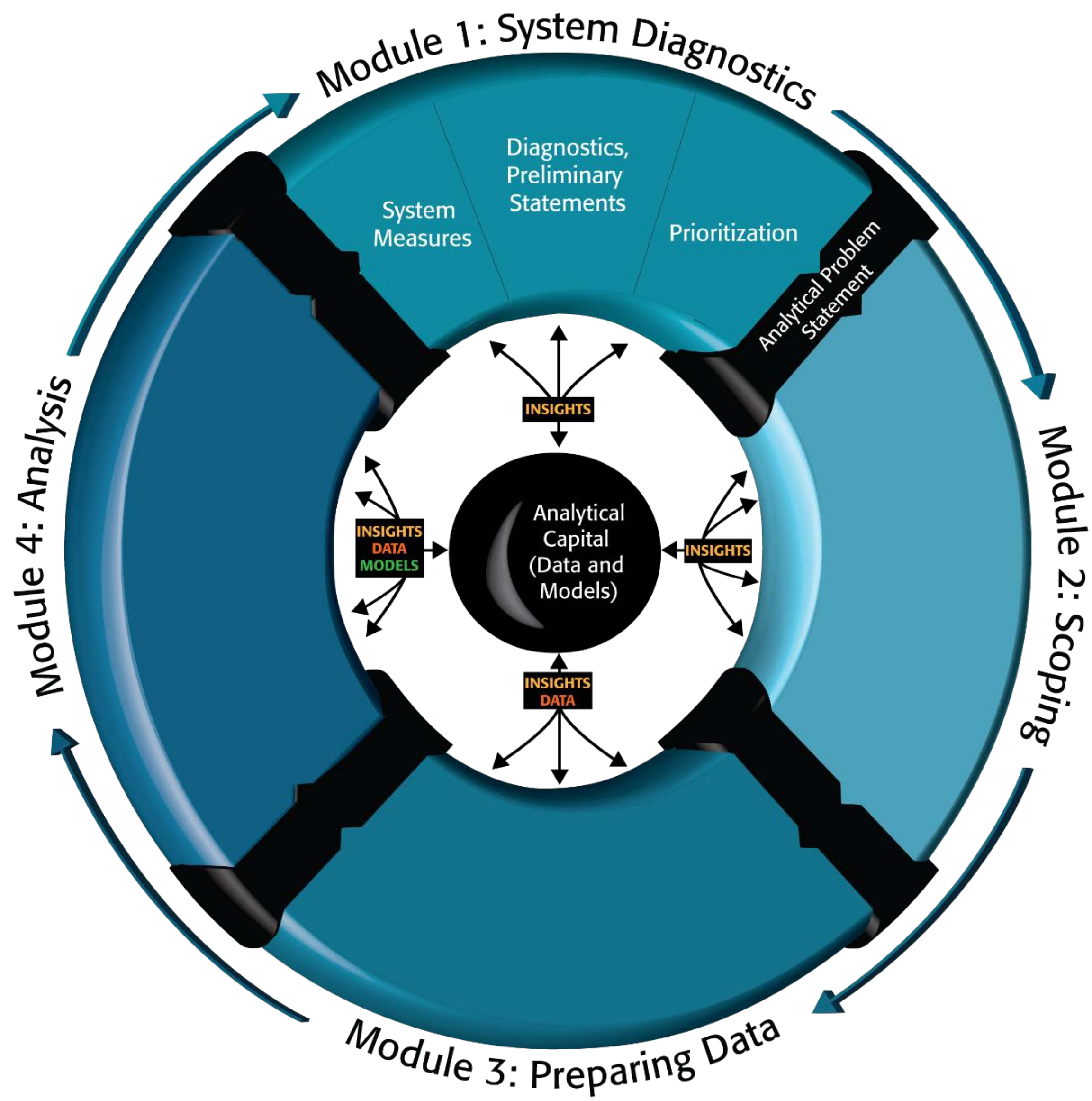

Figure 4. Diagram. System characterization and diagnostics within the 21st Century analytic project scoping process. (Source: Federal Highway Administration.)

ENVISIONING TARGET SYSTEM PERFORMANCE

Recently, performance measurement has received a great deal of attention as a focal point for improved transportation system management. The Federal Highway Administration (FHWA) Office of Transportation Performance Management[5] offers a Web site with resources that help transportation system managers create and refine system-wide Transportation Performance

[5] Federal Highway Administration (FHWA) Office of Transportation Performance Management, http://www.fhwa.dot.gov/tpm/.

Management (TPM) capabilities (see Figure 5). Many agencies are working to enhance their TPM capabilities in advance of expected national rulemaking related to performance management.

FHWA Definition of Transportation Performance Management

Transportation Performance Management, a strategic approach that uses system information to make investment and policy decisions to achieve national performance goals…

- Is systematically applied, a regular ongoing process
- Provides key information to help decision makers allowing them to understand the consequences of investment decisions across multiple markets
- Improving communications between decision makers, stakeholders and the traveling public.
- Ensuring targets and measures are developed in cooperative partnerships and based on data and objective information

Figure 5. Graphic. Definition of transportation performance management.
(Source: Federal Highway Administration.)

Sayings such as “manage to what gets measured” ring especially true for complex systems-of-systems like surface transportation systems. Increased focus on performance measurement is broadly helpful, but the trend towards more quantitative performance measurement is of particular relevance to the 21st Century analyst. First, many aspects of performance management are directly applicable to both system management and analytic project management. Second, the results and insights gained from conducting analytic projects informs performance management aspects of system management. Because of the insight gained from working with field data and assessing system time-dynamics in modern analytical tools (such as a microscopic simulation tool), analysts should have a seat at the table whenever system performance measures are discussed, defined, or used. This is broadly true in the characterization of system performance and decisions related to system management (e.g., which candidate investment strategies are most cost-effective). The 21st Century analyst has much to offer in this arena. Analyst participation is most critical for the decisions made regarding how alternatives are compared, and how prospective analytical projects are conceptualized, prioritized, and documented.

Products and By-Products of the Transportation System

The primary function of a transportation system is to facilitate the movement of people and goods. The inherently positive “products” of the system are completed trips, goods moved from one place to another, and travelers delivered to their destinations. For this document, the notion of a system “product” is only positive. One aspect in peak performance then, is a characterization and quantification of both personal mobility and positive economic activity generated. At the same time, the transportation system also produces a set of negative by-products associated with the movement of people and goods (crashes and fatalities, delay, and environmental impacts).

Figure 6 provides a graphical illustration of this distinction between system products and by-products.

Figure 6. Illustration. Transportation system products and negative associated by-products. (Source: Federal Highway Administration.)

In the remainder of this section, we discuss the inherent products of the transportation system and why they are as important as by-products in characterizing system performance—although they are often ignored or left as unquantified (and therefore unimportant) notions. The 21st Century analyst plays a key role in helping define, quantify, and estimate system products (and by-products) to inform a 21st Century concept of system performance.

The Trap of Managing a System Using By-Products Alone

By-products of the movement of people and goods are inherently negative: e.g., crashes, fatalities, fuel consumed, and emissions. The by-products are no more (or less) important than system products. However, if by-products alone are used to characterize system performance, there are nuanced effects that can result in irrational investments.

The problem of managing by-products alone is that any alternative examined that reduces system use is characterized as having at least some positive effect. Taken to an extreme, one can simply point out that a transportation system with no users has no negative by-products: zero crashes, zero fatalities, zero delay, and no environmental impacts. Clearly, this is a self-evident negative outcome, wherein the transportation system is completely unused and moves neither people nor goods. There must be some counterbalancing measurement of the system product so a more complete picture of system performance can be made.

Even for cases less extreme than the “empty system,” there can be subtle negative effects. Essentially, any considered alternative that suppresses use of the transportation system will accrue benefits associated with reduced by-products. Hidden deep within complex analyses, the benefits of improving system capability to reliably deliver people and goods are not valued—while small changes in delay and other by-products are measured, monetized, and used to inform decision-making.

Why has system performance traditionally focused on system by-products? Simply put, it is because system by-products are easier to observe and measure than system products. When a crash occurs, it can be counted, as well as associated fatalities and delay measured or estimated. Legacy 20th Century technologies (e.g., loop detectors) used to monitor a system are geared to provide snapshots of roadway speed, transit system ridership, and congestion. These systems only detect individual vehicles and travelers as indistinct entities. We may be able to count and classify vehicles by size, but only where we find them. The notion of the snapshot of conditions in time fails to reveal the reality of trip-making and goods delivery within the system.

The 21st Century analyst can introduce a more balanced performance measurement approach, quantifying both products and by-products. Mobile phone technologies and wirelessly connected vehicles are increasingly capable of generating data that can quantify efficient trip-making within a network. Modern "smart tag" technologies associated with cargo can have a similar impact on characterizing both the volume and reliability of goods movement. Further, modern time-dynamic analytical methods can be used to characterize the amount and quality of mobility delivered by a transportation system, either as a part of a retrospective analysis or in the analysis of competing future alternatives.

Managing System Performance: Maximize Products, Minimize By-Products

A 21st Century characterization of managing a transportation system can be described as a two-ledger system wherein system products (people and goods delivered reliably) are maximized and system by-products (crashes, fatalities, delay, and environmental impacts) are minimized. Some agencies understand this important point and have begun to include mobility product measures as a part of their routine performance reporting. For example, Florida includes aggregate freight deliveries as a measure of performance in their performance indicators.[6] Other agencies calculate broad annual estimates of economic impact, including access to jobs. The American Association of State Highway and Transportation Officials (AASHTO) New Language of Mobility (2013)[7] recommends AASHTO "green light" language for describing transportation system performance to the public, including sustainable mobility, economic impact, and access to jobs. However, these system product examples are far less prominent, less quantitative, less precise, and broadly aggregated in nature than associated by-product measures.

The 21st Century analyst can bring data and tools together to provide increasingly robust and quantitative measures of system performance. Some useful measures of system products over a period of time (e.g., a peak period, or a day, or a month) include reliably completed trips and total value of goods delivered. While these products may be hard to measure directly, using time-variant travel time data and supporting estimates of ridership and volume data, the 21st Century analyst can conduct travel reliability analyses that are a key first step in the measuring a system product. Reliability data are a key element in characterizing trip-making since, if a trip takes much longer than expected, the disbenefits associated with disrupting travel plans outweigh the

[6] Florida Transportation Indicators, http://www.floridatransportationindicators.org/.

[7] The New Language of Mobility: Talking Transportation in a Post-Recession World, http://downloads.transportation.org/ANewWayToTalkAboutTransportation/NewLanguageofMobility.pdf.

benefits. This is particularly true for goods movement within a supply chain, but the same basic principles hold for person-trips. For example, if a trip home from work takes longer than expected and a change to child-care arrangements is required, this can have direct and measureable financial consequences.

The Congested Port Example

Consider a collection of roadways connecting a port facility with goods distribution points in a large metropolitan area (see Figure 7). Access to the port facility may be plagued with unpredictable delays at certain times of day. To keep the port active and competitive with other ports along the seacoast, the region might consider a system that rations port facility access to reduce delays and make trips more reliable. Another alternative under consideration is to make a series of related capacity improvements at known bottlenecks in port access (e.g., freeway interchange movements, signal timing plans, and port-entry processing procedures). Data are then collected, and a well-constructed delay analysis is conducted. The Access Rationing alternative reduces delay by 35%, while the Capacity Improvement alternative reduces delay by only 7% (and costs more). Based on delay alone, a rational decision-maker would choose the Access Rationing alternative. An analyst could also calculate a measure of the number of containers reliably delivered in a peak three-hour period—a Peak Reliable Throughput measure. Using this measure, the Access Rationing alternative does not change peak reliable throughput while Capacity Improvement improves peak reliable throughput by 10%. Now the relative merits of the two alternatives can be weighed using a balanced discussion of a tradeoff in the economic benefits of higher peak throughput versus lower delays. It may even lead to a discussion of how a third alternative combining the most valuable aspects of both alternatives can be conceptualized and evaluated.

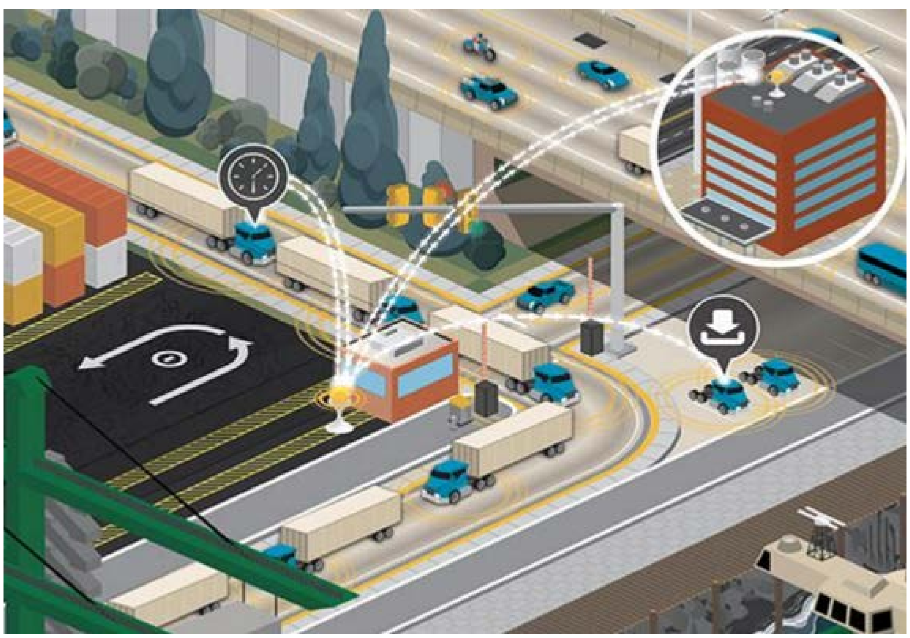

Figure 7. Illustration. Example of a measuring system product.
(Source: U.S. Department of Transportation.)

What Does "Good" Look Like—Multiple Stakeholders, Multiple Views

Using a balanced scorecard approach for performance measures can also help engage stakeholders in a discussion of what "good" looks like for a particular system. The 21st Century analyst has a key role to play here as well. If a high-level description of measures (e.g., delay or crashes) are listed on a single page, stakeholders and decision-makers only engage at the highest level in discussions of tradeoffs. When the time-dynamics of the system are explored in terms of more precise, quantitative measures, this sharpens the focus and discussion to reveal underlying differences in stakeholder viewpoints—viewpoints that very rarely represent a consensus. The analyst does not represent a specific viewpoint in this discussion; instead, he or she facilitates a reconciliation of give and take among stakeholders that can lead to a more informed, nuanced consensus about "good" system performance. This is critical, since if an analytic project is conceptualized, scoped, designed, and delivered with an underlying flawed conception of system performance, regardless of the technical skill of the analyst, the analysis will carry these flaws forward. These flaws can threaten the value of any analytic project, potentially increase analytical cost, and threaten the quality of any decision-making informed by such an analysis.

CREATING SYSTEM CONGESTION, RELIABILITY, AND SAFETY PROFILES

With enriched data sources, time-variant profiles of key performance measures are indispensable for bringing depth and clarity to deliberations on system performance. Before jumping to scoping an analytic project, it is necessary to understand current system performance and review existing studies associated with the defined system. This step can lead to more effective analytical project development, more focused and valuable analytical results, and a more efficient use of resources associated with analysis.

To start building system profiles, with identified problems and defined system boundaries (see the Introduction), data analysis is used to identify and collect/assemble the existing data and analysis results for diagnostics. The profiles include the data used in related (previous and current) analyses, the hypotheses and findings of these studies, and the insights gained by the analysts who conducted these studies. For example, an intersection with a high crash rate is identified within a transportation analysis. Before scoping a project to solve this problem, a data analyst pulls out the time-variant crash data at that intersection and searches relevant studies—such as intersection design studies, traffic signal design studies and crash attribute studies—to create the system profiles and document findings from the previous studies.

Standardizing Methods for Performance Measurement

Using existing information to identify/diagnose system problems, the system performance profiles are developed and maintained in a continuous process based on system management needs and the availability of data and tools. Performance measures should be consistent over time; within a day, month to month, or year to year. Standardizing how data are collected and processed is the key to ensure quality and consistency of performance measures identified in the system. Guide on the Consistent Application of Traffic Analysis Tools and Methods[8] listed the main benefits of consistent analysis throughout the project development life cycle as follows:

- Analysis cost and time are reduced because repeated effort is not needed.
- Analysis results can be refined and reapplied based on the results of prior studies.
- The project analyses enjoy a longer shelf life.
- The credibility of analysts, managers, and the project development process as a whole improves.
- Agency decision-making is more effective and consistent.

Because the transportation system is in a constant state of change, we need to describe the system in a way that the performance measures and the methods by which those values are calculated are consistent, the measures are comparable, and how the system performs is consistently understandable over time. When a new data set or new sensors are added, data analysts need to revise the calculation methods so the new data can be added properly, and document the changes so a meaningful comparison can be made in the future. That way, when monitoring and diagnosing system performance, confounding factors can be eliminated to the minimal effect and the subjective performance measures can be evaluated constantly through and even beyond the project cycle.

System Profiles

The purpose of building system profiles is to characterize system performance (the system is getting better or worse) and to identify anything missing in the profile so the profile can be improved over the long term. With a large volume of data from different sources, using traditional data analysis techniques to develop system profiles can be cumbersome when an analyst attempts to find patterns in the data and establish relationships among variables. Using modern data mining and visualization techniques can lead 21st Century data analysts to insights that would be difficult to discover using traditional methods. This section offers a series of example system profiles (including their key components): congestion profiles, reliability profiles, and safety profiles.

Congestion Profiles

Time-variant congestion measures may contain travel time, vehicle delay, bottleneck throughput, queue length, vehicle stops, and other attributes, depending on the nature of the problem and the system features. For example, to create a freeway corridor congestion profile, travel time and bottleneck throughput may be selected as performance measures; to analyze an intersection performance, queue length and vehicle delay may be used throughout the study. Figure 8 presents an example of congestion maps from Washington State Department of Transportation (WSDOT).[9]

---

[8] FHWA, *Guide on the Consistent Application of Traffic Analysis Tools and Methods*, November 2011. http://www.fhwa.dot.gov/publications/research/operations/11064/11064.pdf.

[9] Traffic Map Archive, http://www.wsdot.wa.gov/data/tools/traffic/maps/archive/?MapName=SysVert.

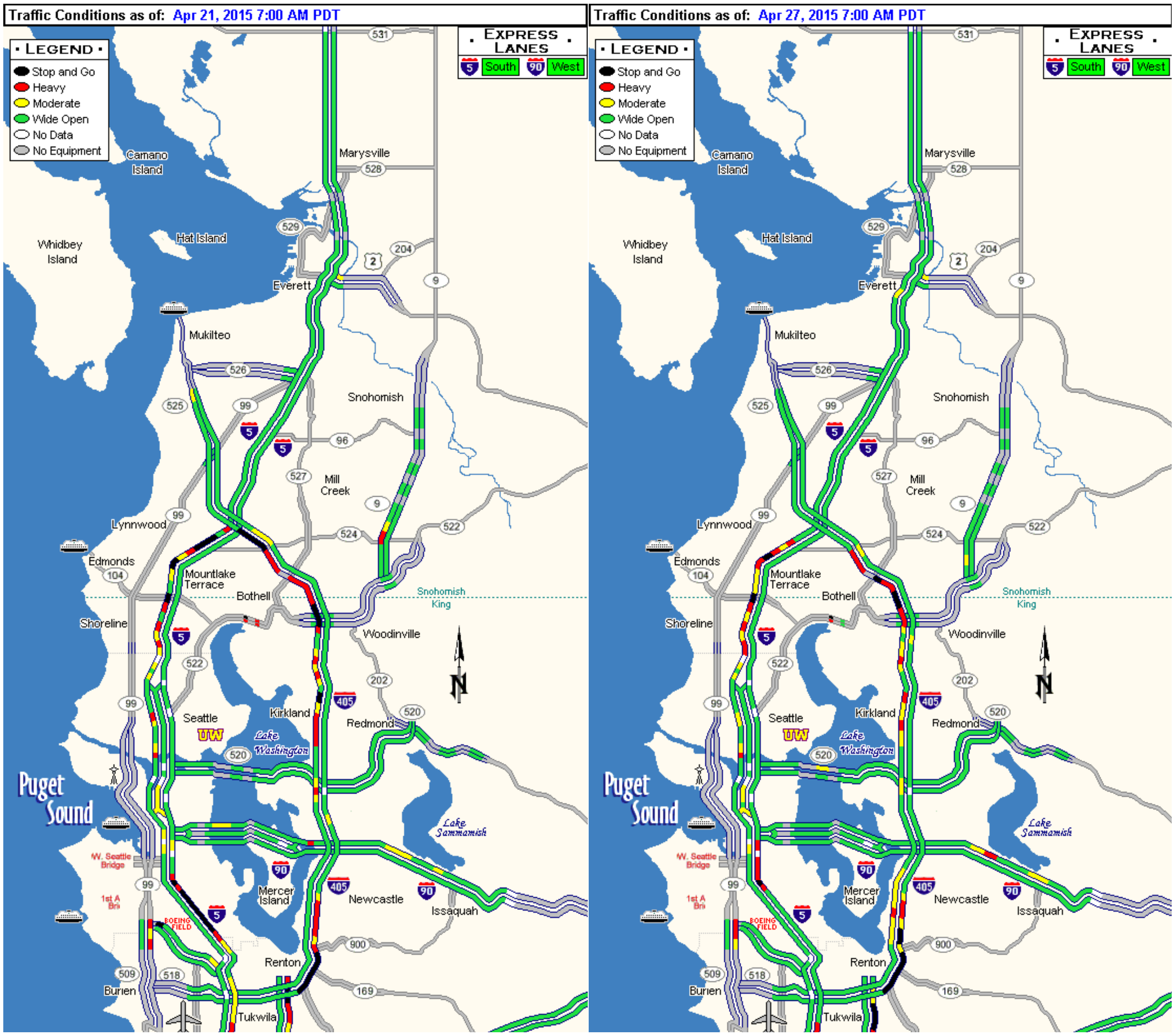

Figure 8. Maps. Washington State Department of Transportation congestion maps.
(Source: Washington State Department of Transportation)

Using speed data, WSDOT has archived a system-wide congestion map on ten-minute intervals since January 2013. These maps can be used in sequence in a simple and intuitive way to identify problem or bottleneck locations, as well as other changes within the system over time. Figure 9 illustrates another example congestion profile from the Kansas City Scout Traffic Management System (KC Scout).[10] KC Scout developed a traffic dashboard to graphically display speeds and travel times by roadway segment in 15-minute, 30-minute, or 1-hour resolution.

[10] Kansas City Scout Traffic Management System, http://www.kcscout.net/.

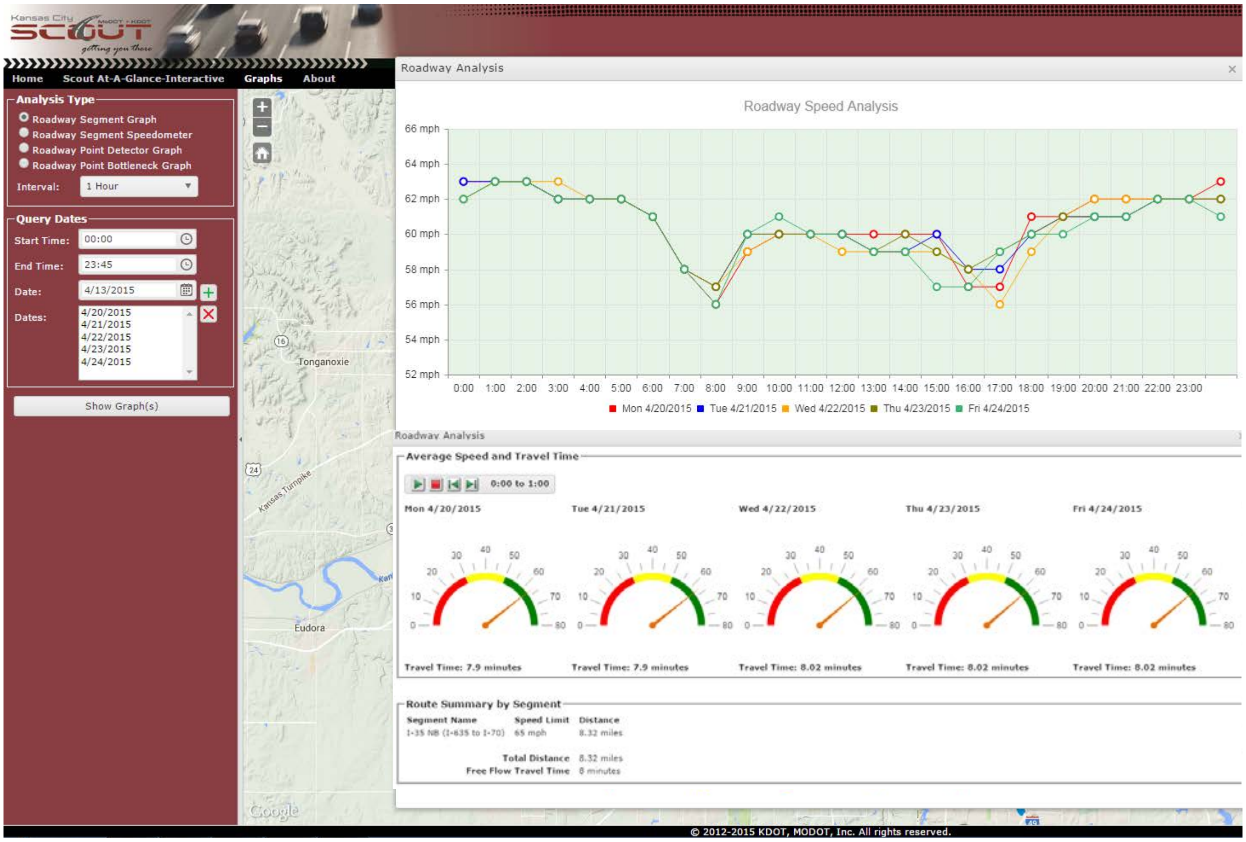

Figure 9. Snapshot. Kansas City Scout traffic dashboard.
(Source: Kansas City Scout.)

Reliability Profiles

A within-day time-variant travel time chart is an effective way to convey system-wide or individual-route travel-time reliability. When combined with the number of delivered trips, the data can also be used to measure and report reliable throughput. The FHWA Travel Time Reliability Measures Guidance[11] suggests a set of performance measures used to quantify travel time reliability: 90th or 95th percentile travel time, buffer index, planning time index, and frequency that congestion exceeds some expected threshold (see Figure 10). According to the guide, travel-time reliability measures the extent of this unexpected delay and provides this definition: *the consistency or dependability in travel times, as measured from day-to-day and/or across different times of the day*. Figure 11 presents a plot of a within-day roadway segment travel-time distribution from Monday to Friday from KC Scout. The average and variation of travel time can be easily identified from this graphical expression.

[11] Travel Time Reliability Measures Guidance, http://ops.fhwa.dot.gov/publications/tt_reliability/.

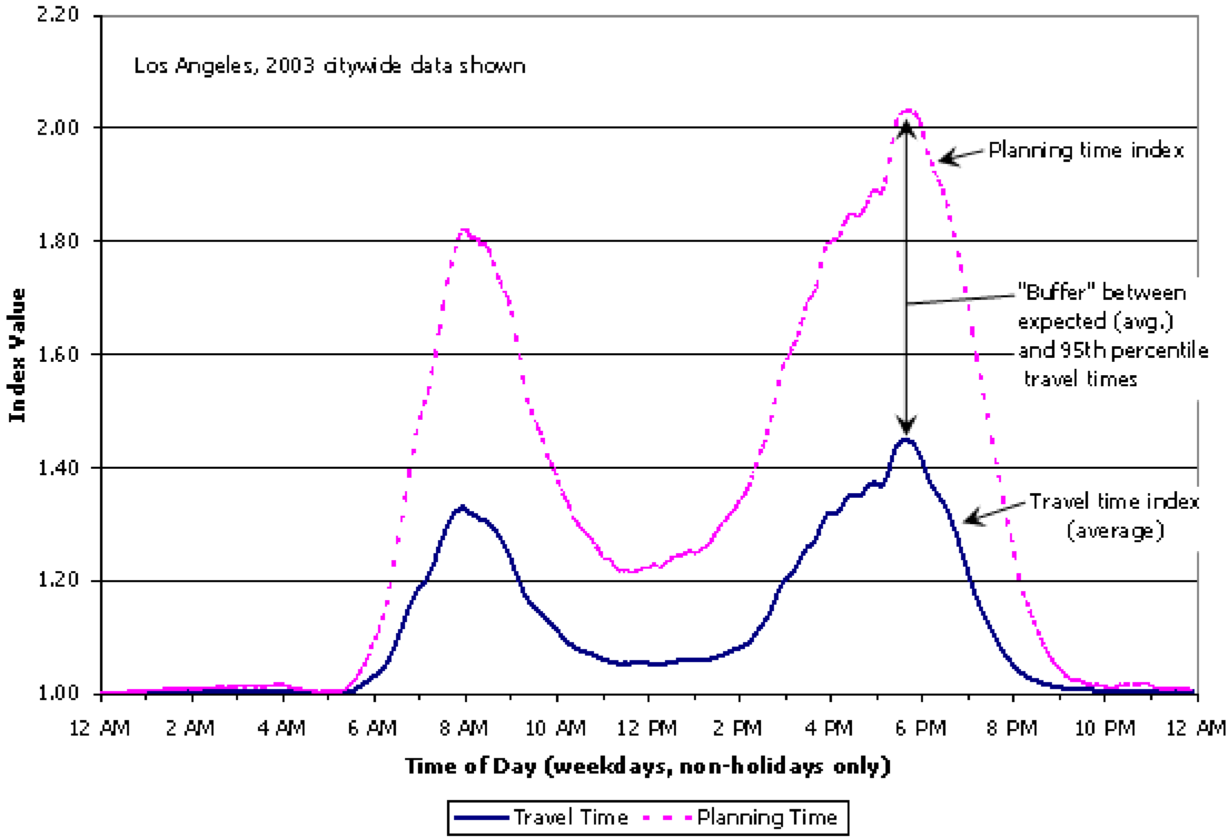

Figure 10. Chart. Reliability measures are related to average congestion measures.
(Source: Mobility Monitoring Program, http://mobility.tamu.edu/mmp.)

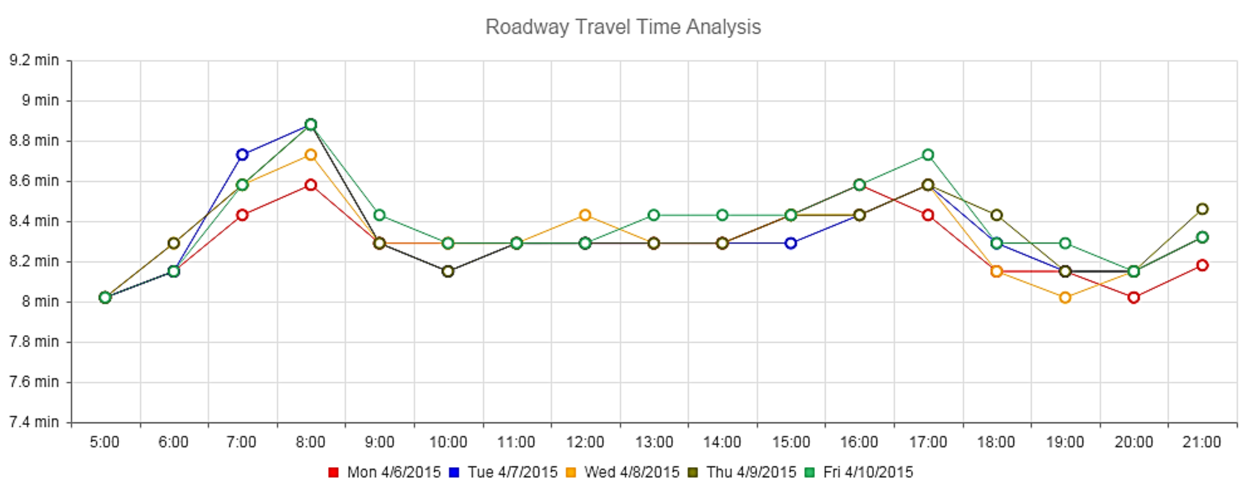

Figure 11. Chart. Roadway travel time distributions.
(Source: Kansas City Scout.)

Safety Profiles

Common safety measures include crash rates and number of fatalities, which can be acquired directly from state or local agency databases. Similar to a congestion profile, a geographic illustration can be an effective way to identify the problem locations. In the KC Scout Annual Report[12] (see Figure 12), KC Scout uses a heat map to illustrate the locations of multiple-vehicle incidents in 2013 to intuitively identify severe/high-incident rate locations. Figure 13 gives another example of geographically illustrating archived incident records from the Regional Integrated Transportation Information System (RITIS) developed and maintained by University of Maryland's Center for Advanced Transportation Technology Laboratory (CATT Lab).[13]

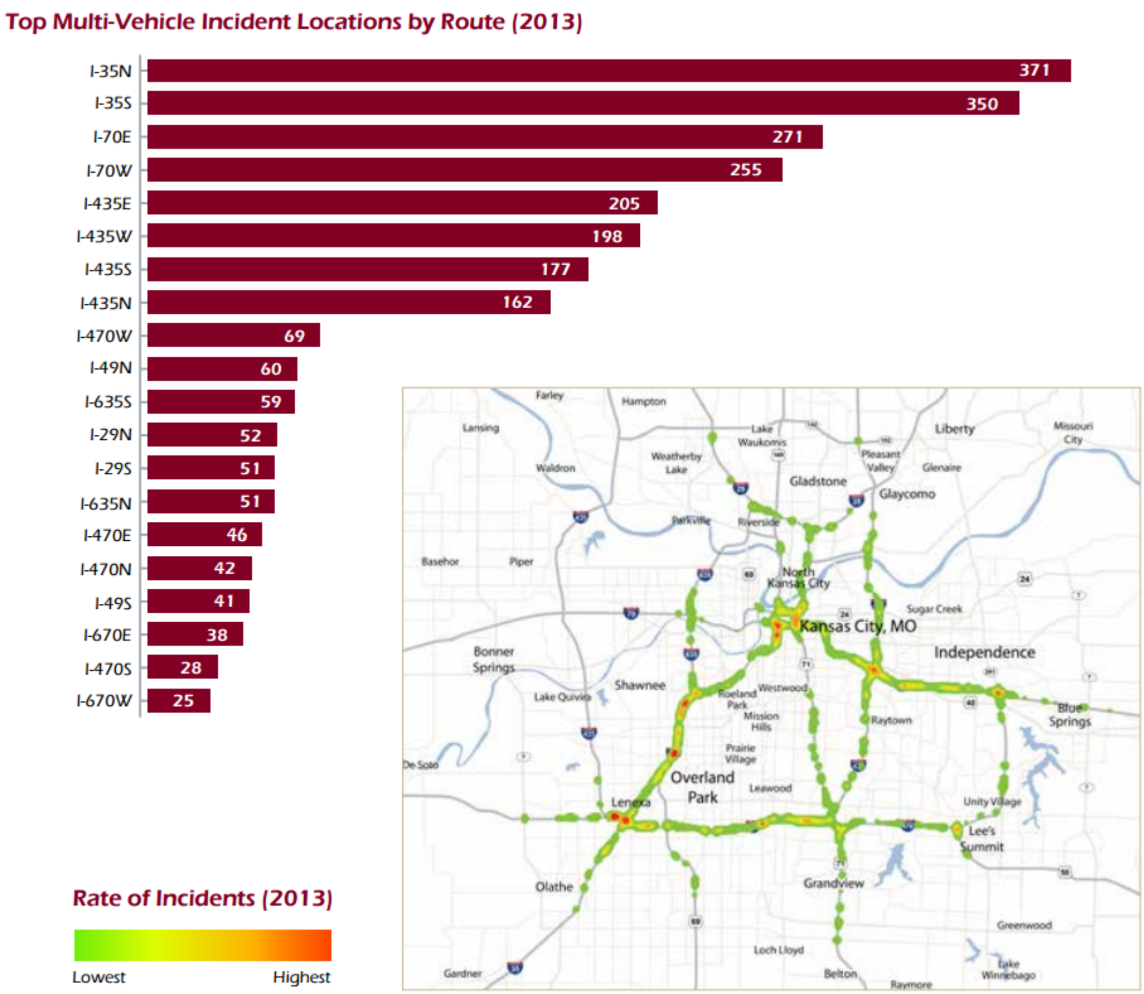

Figure 12. Snapshot. Top multi-vehicle incident locations by route.
(Source: Kansas City Scout.)

[12] KC Scout FY 2013 Annual Report. http://www.kcscout.net/downloads/Reports/Annual/AnnualReport2013.pdf.
[13] University of Maryland CATT Lab. http://www.cattlab.umd.edu/.

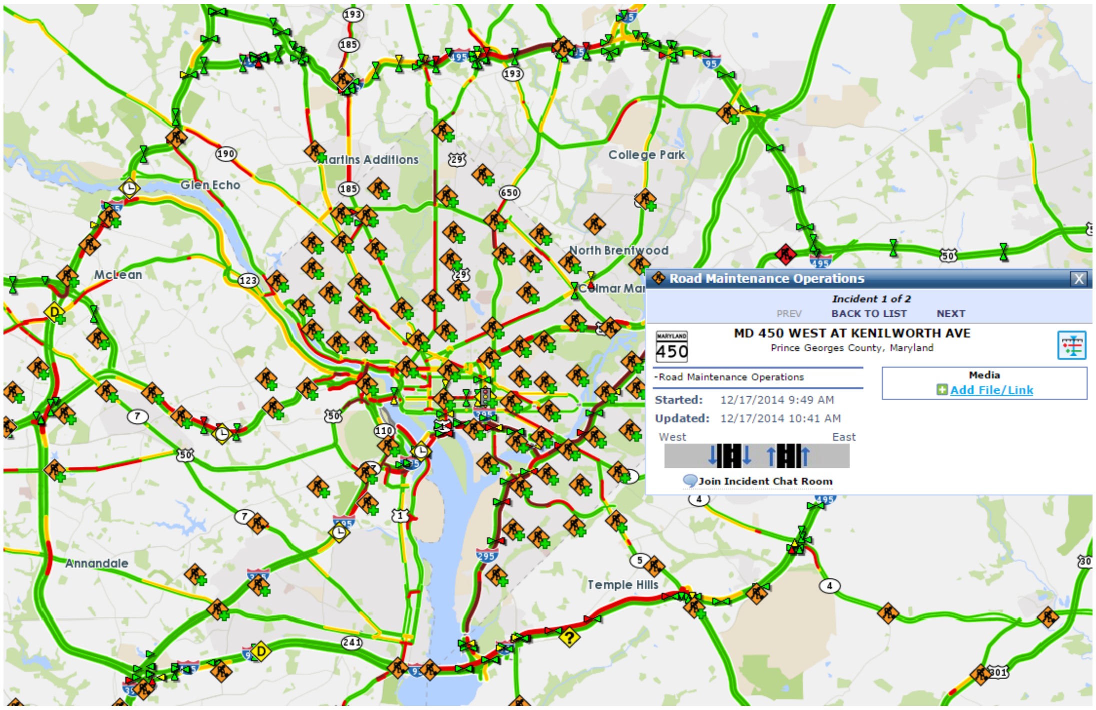

Figure 13. Snapshot. Maryland average annual incident frequency during morning peak hours by location.
(Source: Regional Integrated Transportation Information System.)

Special Considerations for System Profiles

When creating system profiles, there are several considerations that require special attention to ensure that the data are treated properly and the resulting profiles are analytically sound. Additional discussion regarding the treatment and analysis of data can be found in Module 3.

Inconsistent Data

When preparing the system profile based on existing data and studies, analysts may face the situation that the relevant study data are neither comprehensive nor collected consistently over time. While data may be available on a subset of intersections, incident data may only be available for the past five years, and/or travel time data were collected before a new interchange was built. Data in these cases need not be thrown out—with some adjustment or identification of the limitations of these data, useful system profiles may still be constructed.

Outlier Events

Abnormal traffic conditions due to outlier events (such as sports events or extreme weather conditions) cause bias if they are not separated from regular traffic conditions. Traditionally, these types of conditions are either averaged out in analysis or treated as outlier data and removed from the analysis. These types of operational conditions can actually be identified from different data sources using cross reference approaches (such as data mining) or statistical approaches (such as cluster analysis).

Seasonality and Cyclical Trends

Seasonality is a pattern in time series data that repeats every year. This repetitive pattern in transportation can be attributed to a variety of causes. For example, an examination of monthly congestion data could offer a view of the systematic intra-year movement or seasonal patterns of congestion. There are variations in these patterns among different systems. These trends can be obtained by examining weekly, monthly, or seasonal averages of demand, congestion, and safety measures. Figure 14 illustrates an example of monthly travel-time reliability trends in the National Capital Regional Congestion Report.[14] Taking advantage of the availability of rich data and analytical tools, the Metropolitan Washington Council of Governments (COG) produces the quarterly updated traffic congestion and travel-time reliability performance measures to summarize the region's congestion and to examine reliability and non-recurring congestion.

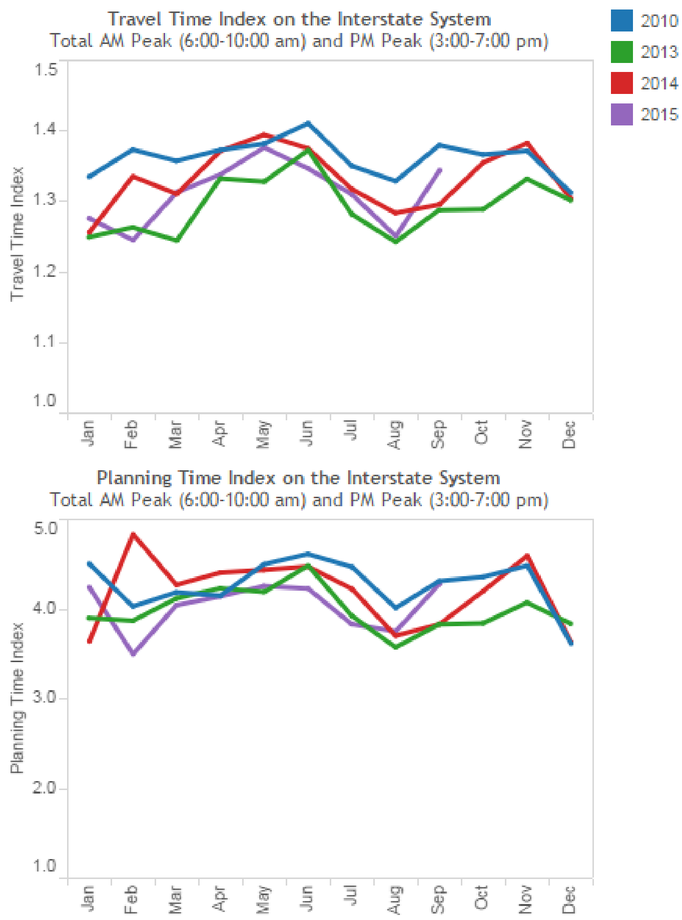

Figure 14. Chart. National capital region congestion report.
(Source: Council of Governments.)

[14] National Capital Region Congestion Report, https://www.mwcog.org/documents/2016/07/04/national-capital-region-congestion-report/.

Preparing Predictive Profiles over Different Time Horizons

With comprehensive information gathered, a data analyst can predict trends over different time horizons. In a short term, without changing system service (such as opening a new ramp or a new bus route, or starting new construction), the profile should remain the same or follow the previous trends. In the long term, statistical methods and transportation planning prediction models can help analysts predict the trends. Analysts need to highlight the changes and key factors influencing the performance measures and select appropriate methods to prepare predictive profiles.

DIAGNOSTICS

Having a set of relevant and dependable system performance profiles is a key resource for the 21st Century analyst to both understand patterns of system dynamics and to recognize when something new (and potentially problematic) is happening in the system. Much like a doctor takes measurements of a patient's blood pressure, pulse rate, and other data over time to establish a historical "healthy" range of vital statistics, the informed analyst has an analogous historical record of system dynamics to rely upon when assessing the nature, root causes, and severity of issues within the transportation system. Keeping data at a relatively disaggregate level is key in establishing what a consistent "healthy" range for these data may be. If we calculate an average blood pressure over two years, a doctor may find nothing of note. However, if an underlying pattern is changed (say the presence of some disorder), then blood pressure in the last two months may be much higher, falling outside a statistical range (e.g., a one-sigma band) taken using prior historical data. This is a clear signal to the doctor that something has changed, quite possibly for the worse, and that the system performance has deviated from previous patterns.

In this section, we refer to diagnostics as the process where an analyst considers the set of historical system performance profiles to identify anomalous or problematic elements of current system performance. This diagnostic activity can be either direct or indirect in nature. Direct diagnostics occurs when an analyst engages directly with the data to make observations and hypotheses about the nature, root causes, and severity of system performance issues. Indirect diagnostics occurs when a stakeholder suggests an observation or hypothesis regarding system performance that the analyst may be asked to investigate. Both direct and indirect diagnostic activity are useful. A successful 21st Century analyst often finds highest value from a balanced slate of direct and indirect diagnostics.

Direct diagnostics allows the analyst to develop new insights derived solely from the data. In some cases, no other stakeholder may be considering the full range of these performance data, and the analyst provides a valuable and unique viewpoint to inform system management.

Indirect diagnostics can reveal important insights regarding potential issues in the system from stakeholders with different experiences in the system, and therefore different viewpoints that are not data-driven. These insights are nearly always valuable in their own way, and sometimes reveal limitations in the data quality, data coverage, and selected performance measures.

System profiles, and system performance data, are not perfect indicators of actual system performance. Successful analysts fully engage with a set of stakeholders with different vantage points from which to assess system performance. If an analyst is disconnected from these stakeholders, the value of purely active diagnostics is lower because it is not informed, tested, or enhanced by these alternative viewpoints. Further, the analyst should keep in mind that data itself cannot always be an infallible barometer of system performance. The ways that data are collected, processed, and analyzed can always be improved.

The remainder of this section presents tactical guidance on how the analyst can engage with stakeholders to leverage their insight, reconcile perception and observation, and create concise, targeted hypotheses regarding system performance. These hypotheses, when considered together, are the raw material used to generate preliminary analytic problem statements.

Leveraging Data and Stakeholder Insight

Observations about the performance of the system can come from many different viewpoints: from system managers who observe the buildup and dissipation of congestion through graphical displays and camera images, from agency staff who maintain the system, from the public safety officers who traverse the system, and from travelers themselves. When something isn't "right" or when system dynamics change, there are likely stakeholders within the system who notice that something out of the ordinary is happening. Whether this insight comes from direct, hypothesis-oriented engagement with data or the experience of observing or traversing the transportation system, it is important to establish processes and relationships that pull together insights from diverse stakeholders. The 21st Century analyst actively cultivates a set of performance-oriented individuals who engage with the transportation system in diverse ways. This cadre of contacts can provide invaluable observations that, when combined with data-driven observations from direct diagnostics, can lead to more substantive candidate hypotheses and ultimately more meaningful analyses.

A hypothetical example where data-driven and stakeholder insight can be usefully combined is shown in Figure 15. Consider an analyst who notices a change in a reliability profile report covering the last quarter for northbound afternoon travel. This might be a one-time anomaly associated with a transient event (e.g., a temporary change in demand pattern), part of a longer-term persistent trend (e.g., the effect of new housing or employment center opening), or simply a misleading artifact related to erroneous data (e.g., loop detectors or speed sensors having been recently recalibrated). Around the same time, a staff person directing maintenance crews and pavement inspections notes that getting around the network in the afternoon in the northbound direction is more tedious than any time in recent memory, with longer and more unexpected delays. This same staffer observes that the pavement quality in some of the northbound sections has markedly deteriorated over the recent bitter cold winter months. The staffer might observe that something was amiss and might attribute the change to the pavement conditions but couldn't be sure since other factors (e.g., demand patterns, incidents, and weather patterns) might be in play. Independently, the two observations regarding the northbound system performance might be shrugged off or not considered critical or never brought together. However, just considering the two observations together proves much more powerful than if the observations were only considered independently.

Figure 15. Flowchart. Leveraging direct (data-driven) and indirect (non-data) observations. (Source: Federal Highway Administration.)

In this regard, the two forms of observations can be powerful evidence when trying to pin down the nature, severity and root causes of poor system performance. Like a detective, the analyst can pull together independent observations to better establish the context for a particular issue and create a candidate hypothesis (see Figure 16). This doesn't mean that just because the analyst observed a change in travel-time reliability and the staffer wondered about the effect of the pavement conditions that we can immediately jump to the conclusion that there is a simple and straightforward reason for recent poor system performance. The analyst is still obligated to assess the validity of the underlying data to ensure that the change in system performance is not related to changes in data collection and processing. Many state DOTs with large loop detector systems perform routine validity checks on data quality to screen for issues, but the analyst cannot blindly assume that these are 100 percent effective.

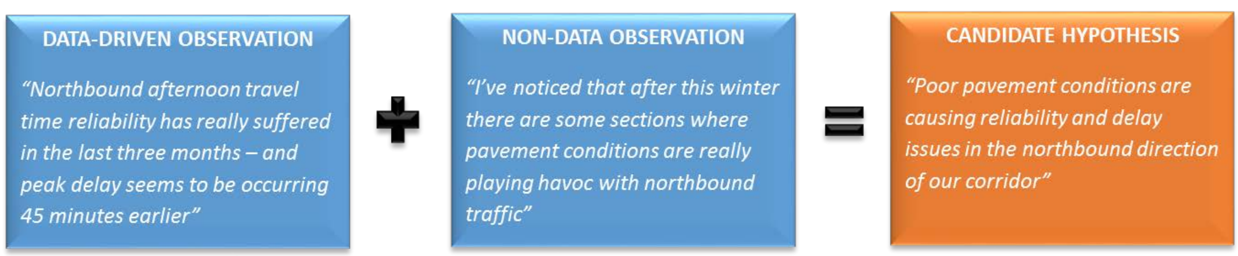

Figure 16. Flowchart. Cross-validating observations to create a candidate hypothesis. (Source: Federal Highway Administration.)

Reconciling Perception and Observation

Combining insights from individual observations and the analysis of data can lead to powerful and effective outcomes with respect to understanding system dynamics and developing candidate hypotheses. Figure 17 provides an example to how perception—supported by targeted data analysis—can lead to key insights regarding system dynamics and help to focus efforts to mitigate and resolve system performance issues.

The "Underutilized" Arterial Example

In an urban traffic management center several years ago, Transportation Management Center (TMC) staff had an excellent birds-eye view of a local freeway bottleneck from the windows of their facility (see Figure 17). Every day, they observed the buildup and dissipation of congestion where a major on-ramp fed into the freeway just ahead of a lane drop. Every weekday morning, congestion would develop at this location and stretch for a distance back onto the freeway, which was elevated from the surface street below. The arterial network in the vicinity was also visible to the TMC staff, and these roadways had no visible congestion; in fact, they appeared almost deserted in comparison with the densely packed freeway system. The TMC staff felt that if only the travelers knew how empty these surface streets were, they would flock to them, making use of the arterial system and freeing up the bottleneck on the elevated freeway. A data analyst working with the TMC staff had access to freeway counts and travel times but no arterial data. To better diagnose the nature of the "freeway bias" problem, a short-term travel-time study on the arterials was conducted to see how much travel time might be saved for popular origin-destination pairs along this part of the system. As it turned out, in the morning peak period, the travel times on the freeway, although at a slower speed, were quite comparable to the travel times on the arterial system. The arterial travel times were slowed both by signals and delay in accessing the surface street network from the elevated freeway system. In the end, the observed differential in vehicle density did not equate with a difference in vehicle travel time. The team then turned their attention to analyzing alternative methods of improving traffic flow in the bottleneck section, rather than on the provision of more detailed traveler information on changeable message signs.

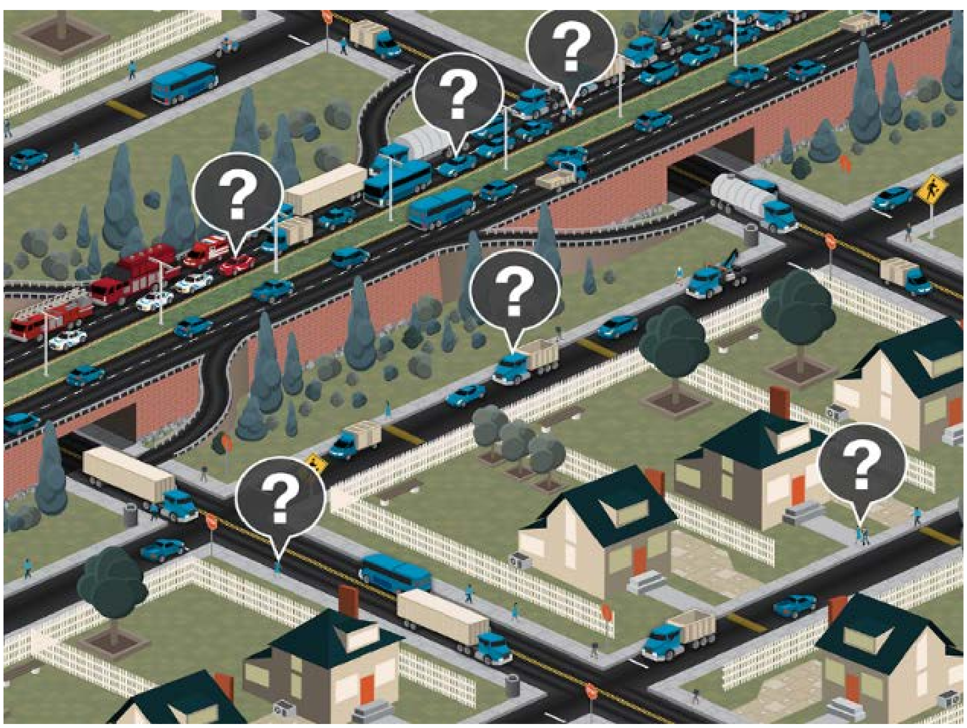

Figure 17. Illustration. Example of reconciling perception and observation.
(Source: U.S. Department of Transportation.)

PRELIMINARY ANALYTICS PROJECT STATEMENTS

This section discusses how to characterize and categorize a list of hypotheses/problem statements developed from the diagnostics step, provide the possible strategies to fix the problem, and consolidate the analytics projects based on available resources.

Characterize Problem Statement(s) and Identifying Candidate Strategies

With the problems identified and verified from the system profiles and the hypotheses regarding the system performance developed from the diagnostics step, an analyst then tags the problems/hypotheses with attributes. The attributes may include system-level; location-specific or with geographical constraints; temporal; with certain conditions; dealing with certain conditions, such as work zones; and dealing with certain improvements, such as traveler information. Some problems/hypotheses may be well defined (like a bottleneck at a weaving area), and some may be more generic (such as system travel-time reliability). Each of the problems/hypotheses can be assigned to one or more analytics projects. From the hypotheses, an analyst merges some hypotheses together if they are related or break down into some subsets if the hypotheses are system-wide.

For example, there is a bottleneck on a freeway segment. From diagnostics, the bottleneck occurs during peak hours due to the large merging and diverging activities at the weaving area. This is a well-defined problem/hypothesis with clear attributes. The problem could be a stand-alone project or could be combined with other problems with the similar attributes, such as the timing issue of the ramp metering near the bottleneck location or the traffic signal optimization issue at the interchange near the bottleneck location.

To form a problem statement for an analytics project, the next step is to add possible strategies to fix the problem (alternative identification); possible approaches/methodologies to analyze the problem, (e.g., simulation or field test), and data needed to conduct the project. To solve the bottleneck issue at the freeway weaving area, a simulation model may be needed to evaluate all possible alternatives, such as adding one auxiliary lane, creating a Collector/Distributor (CD) lane, or optimizing the on-ramp metering with a new set of optimization criterion that takes into account the freeway throughput or speed. An analyst could suggest performing a field test on the later strategy but not on the first two strategies due to the risk and cost-effectiveness issue. Some of the strategies might be developed by the analyst and some of the strategies may be already suggested by the stakeholder at the diagnostics step. Using the bottleneck at the weaving area as an example, Table 1 presents a hypothetical example of initiating a problem statement.

Table 1. Steps in the generation of an analytic problem statement.

Steps	Products/Items/Descriptions
Observation	Long congestion duration and queueing on freeway segment.
Diagnostics	The bottleneck location is a weaving area with a large amount of weaving and diverging activities.
Characterizing the Problem	This problem occurs at a freeway weaving segment downstream from an on-ramp with ramp metering control during peak hours. Data have been checked for validity and other factors (e.g., pavement, glare) do not seem to play a major role.
Identifying Strategies	Some possible mitigating strategies are: • Optimizing fixed metering rates. • Deploying adaptive fixed metering capability. • Creating a CD lane. • Adding one auxiliary lane.
Techniques/Approaches	Primarily simulation with possible field test on the first strategy.

Consolidating Analytics Projects

A key role for the 21st Century analyst revolves around deriving ideas from their own direct observations, talking to the stakeholders, and sorting things out. A purely passive stance of waiting for problem statements to come from stakeholders is not recommended because this may yield a set of incremental or partial concepts that may be individual facets of a larger problem. One major contribution from the 21st Century analyst is to consolidate analytics projects, where several small issues can be grouped and analyzed together to better use resources and gain a better value. For example, pedestrians crossing an intersection, lack of progression at an arterial and a rise in unprotected left-hand turns mid-block on the same arterial can be consolidated into one project or related projects that can share resources and maximize the system performance.

Another key item an analyst needs to guide the selection of which analytical projects to move forward with is a high-level technical risk and reward (value) analysis for each candidate project before prioritizing the problem statements.

Risk Assessment

An analyst considers all the possible strategies to provide a high-level technical risk assessment, resulting in a rating of low, medium, or high. Risks could be related to the size of the network, the availability of supporting data, or the complexity of the most appropriate analytical methods. Risk is related to all the things that can make doing the analytics work difficult, costly, or create the potential for technical failure. The following bullets are some examples to be considered:

- General Risks:
  - Resource issues, such as lack of data or no model.
  - Poorly defined project concept, vague objectives.
  - Identified problem is actually a data quality issue.

- Technical Risks:
  - Large geographic scope required involving multiple jurisdictions to make decisions.
  - Difficulty in representing alternatives from existing models or approaches.
  - Difficulty in predicting strategic traveler behavior (as in mode or route shifting).
  - Requiring detailed modeling or involving new technologies.
  - A requirement to model rare (or previously unobserved) scenarios such as a new upcoming special event and or severe weather scenarios.

Value Assessment

Similar to risk assessment, an analyst lays out all values/benefits from each analytics project ranging from low to medium to high. The values include immediate values and potential or long-term benefits. The following are some examples for consideration:

- Problem scope and potential influence on the overall larger-scale system.
- Visibility and media attention.
- Urgency.
- Severity of the issue.
- The cost associated with potential alternatives.
- The use of sketch planning tools may also be considered in assessment of potential value.[15]

Risk/Reward Assessment

A simplified risk/reward assessment could be conducted to help the later prioritizing step. Moreover, it could be another way to group some projects with low risk and low values if the combination produces higher benefit due to the natural relations between the projects. Figure 18 illustrates a simplified version of a risk-reward assessment chart where the identified projects are placed in one of the boxes to address the risk and reward based on the list of the risks and the values listed above.

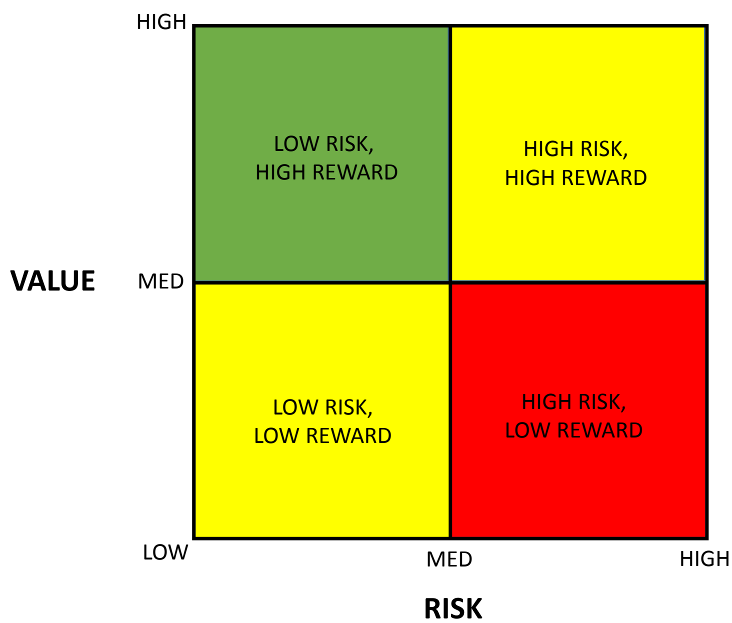

Figure 18. Chart. Risk/reward assessment chart.
(Source: U.S. Department of Transportation.)

[15] NCHRP 07-22 Planning and Preliminary Applications Guide to the HCM, expected December 2015.

Where Transportation Analysis Provides Unique Value

Some straightforward problems can be done by pure retrospective data analyses or by conducting a simple field test. Other problems may involve multiple strategies that cannot be tested in the field and require a modern simulation technique to evaluate the strategies—or a hybrid approach using a small-scale field test to verify the values of the parameters and input the observed data into a virtual network to simulate possible strategies. The 21st Century analyst has a valuable and unique insight into system dynamics, underlying system issues, and potential mitigating strategies. The analyst also brings an understanding of where data alone may be able to answer critical questions and when more complex modeling may be required. These are critical insights into making resources deployed in analytics less risky, more accurate, and more useful.

PRIORITIZING ANALYTIC PROBLEM STATEMENTS

One key step in getting the highest value from analytical resources is making strategic decisions on which analytical ideas to move forward and scope as projects. We recommend doing this periodically, considering a complete portfolio looking across all the candidate ideas generated, rather than individual projects one at a time as the ideas come up.

This may not always be possible—in some cases, urgent needs may arise off-cycle that must be dealt with individually. However, a process that is purely reactive runs the risk that high-value analyses do not have resources available because lower-value projects have already claimed all available analytical resources (staff time and budget).

Figure 19 illustrates a portfolio of ten potential analytical projects considered within a portfolio analysis. In this case, the analyst has rated each project as LOW-MED-HIGH on each of the two assessment axes (RISK and VALUE). Clear winners to move ahead include Project D (High Value, Low Risk), Project G (High Value, Medium Risk), and Project J (Medium Value, Low Risk). Possible projects to consider include Projects I, E, F, and B, where risk and value have equivalent assessment ratings. Projects A, H, and C are not good candidates because each has higher risk ratings than value ratings.

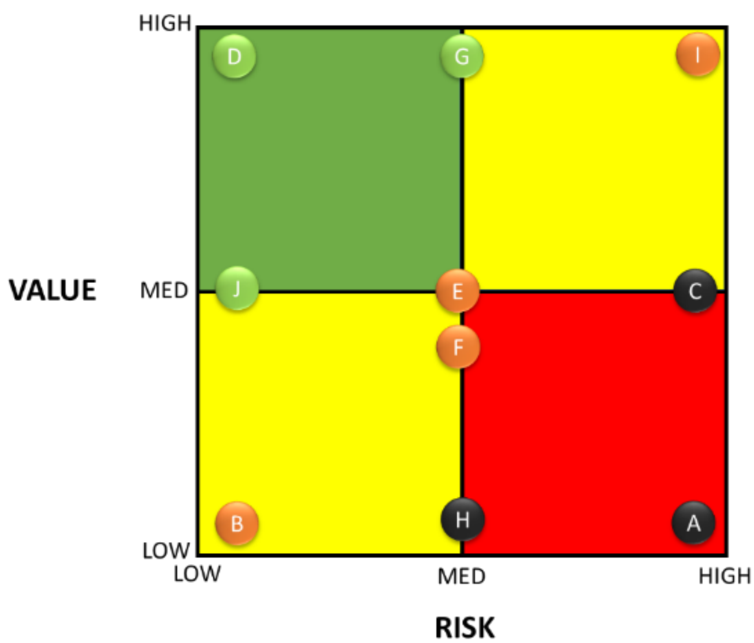

Figure 19. Chart. Risk-reward assessment of candidate analytical projects. (U.S. Department of Transportation.)

The process of optimizing a portfolio of analytical efforts is the next step in the process. The goal is to commit resources to scope only the highest value collection of potential projects while minimizing risk. While pure optimization methods do exist for portfolio analysis, in this case, our recommendation is for a qualitative prioritization using the gathered information rather than a strict quantitative optimization.

At this point, we are not committing funds to conduct the analysis, we are only moving forward a set of promising candidates to formal scoping. Why do we limit the number of projects to scope? Because even at this preliminary level, the effort to scope a project effectively is not trivial. In fact, a common trap for analysts is to have many analytical project concepts then try to jump directly into preparing a formal data collection plan and full analysis plan. This aggressive form of project development is problematic on two fronts. First, key details of the analysis have not been completely thought through, so there is a good chance of wasted effort in data collection and in the analysis phase. Second, the resources incurred to prepare a detailed plan may be wasted when at some point in the effort it is determined that the project is too large, too risky, or infeasible for various reasons that would have surfaced in a formal scoping effort. The goal is to choose a focused number of high-value projects to move forward into scoping, and to allocate sufficient resources and attention to these projects so they are properly and thoroughly scoped.

Consider a total analytical budget representing some combination of staff time, staff experience, and budget available in this round of projects selected for scoping. Clearly, this is not an unlimited resource. The analyst assesses the effort associated with thoroughly scoping each of the projects in terms of HIGH, MEDIUM, and LOW effort. Taking these into consideration, the analyst identifies the target number of projects that can be well-scoped in each category, say one HIGH, one MEDIUM, and two LOW projects. The prioritization process begins by selecting the project with the highest value over risk rating. In our example, this is project D, with HIGH value and LOW risk. Assume this is a MEDIUM effort project, and it takes the MEDIUM slot in the available scoping portfolio. The next two high-value projects (J and G) are rated HIGH and LOW effort for scoping and take the next two slots. Among the remaining moderate value/risk projects (B, E, F, and I), the analyst should choose among projects rated as LOW effort.

As discussed above, this process is qualitative, not strictly quantitative. If there are two HIGH value/HIGH effort projects, the analyst may decide to pursue the second HIGH effort scoping in lieu of the one MEDIUM and two LOW efforts. The goal is to make this process systematic and clear but not so inflexible that low-value/risk projects are consistently selected over higher value/risk projects. The process can also be documented so that all stakeholders have buy-in on the portfolio of projects selected for scoping.

Another consideration at this phase is that just because there are analytical resources to move forward with these projects, are they worth moving forward? Just because staff time or funding is available, low-value, high-risk efforts may not be good analytical investments. At this point, the analyst should be cognizant that it may be time to identify new project ideas or to modify the current ideas so they can be more valuable and/or less risky.

MOVING TO MODULE 2: THE ANALYTIC PROBLEM STATEMENT

In order to move to Scoping (Module 2), a minimum set of information generated in this step of the process needs to be documented for each project as an analytic problem statement. As indicated in Table 2, this can be represented in a single page.

Table 2. Example analytic problem statement.

Steps	Products/Items/Descriptions
System Performance Issue	Long congestion duration and queueing on freeway segment
Summary of Observations	The bottleneck location is a weaving area with a large amount of weaving and diverging activities
Characterizing the Problem	This problem occurs at a freeway weaving segment downstream from an on ramp with ramp metering control during peak hours. Data have been checked for validity and other factors (e.g., pavement, glare) do not seem to play a major role.
Candidate Hypothesis	A reduction in uncontrolled merge and weave behavior in this section will significantly reduce congestion duration and queueing
Candidate Mitigating Strategies	Some possible mitigating strategies are: • Optimizing fixed metering rates. • Deploying adaptive fixed metering capability. • Creating a Collector/Distributor (CD) lane. • Adding one auxiliary lane.
Candidate Analytical Approach	Primarily simulation with possible field test on the first strategy
Value Rating	HIGH
Risk Rating	MEDIUM
Scoping Effort Rating	MEDIUM

- The *system performance issue* provides a concise overarching problem statement and identifies the performance measure or measures that are the focus of the analysis.
- The *summary of observations* describes the nature of the issue providing context for why this particular performance issue was flagged for attention.
- The *characterization of the problem* documents cross-validation and other insights from diverse stakeholders.
- The *candidate hypothesis* represents the leading underlying cause of the system performance issue.
- *Candidate mitigating strategies* documents the range of potential mitigating strategies suggested by stakeholders, including the analyst.
- The *candidate analytical approach* documents the initial thinking of stakeholders and analyst regarding a feasible method for evaluating the effectiveness of the mitigating strategies in resolving the system performance issue.
- The initial ratings for *value, risk,* and *scoping effort* are also documented.

These problem statements are a critical resource for improving the analytical process over time. The analyst can demonstrate stakeholder engagement and involvement. Early statements can be compared against completed projects to assess the accuracy of initial estimates of effort, risk, and value. This can prove invaluable in improving the quality of early project identification over time and can help focus resources on the most critical near-term projects.

To some, the development of problem statements and performing a risk/reward analysis may not seem like time well spent—perhaps the time could be better spent running models and grinding up data. **In fact, this early process of diagnostics and systematic problem statement development may be the most critical element of creating a more efficient organization-wide analytical capability.** The amount of effort is low; it is essentially a bare-bones effort to document a process that occurs in an *ad hoc* fashion in many organizations. When a project idea is identified, the person with the greatest insight may not be there when the project is scoped, data collected, or analysis performed. The insights must then be documented so that the original intent of the work (as defined by the stakeholders) is not lost. Otherwise, there is a risk that a different, lower-value, higher-risk effort is scoped and performed—all because of a lack of documentation at this early stage.

MODULE 2. DATA-DRIVEN TRANSPORTATION ANALYSIS PROJECT SCOPING

Following the characterization of system dynamics and the development of an analytic project statement (presented in Module 1), Module 2 provides guidance on how to develop a transportation analysis scoping plan. Transportation agencies can use this scoping guidance and accompanying scoping tool to develop preliminary scopes and budgets for their transportation analysis projects. Information about the scoping tool is provided later in this module. Figure 20 illustrates the orientation of this module with respect to the diagnostic efforts that precede it and the data preparation efforts that are initiated after the completion of the project scoping summary. Actions in this module include project definition, identification of performance measures to be used in the analysis, data needs for the analysis, and cost and schedule estimation.

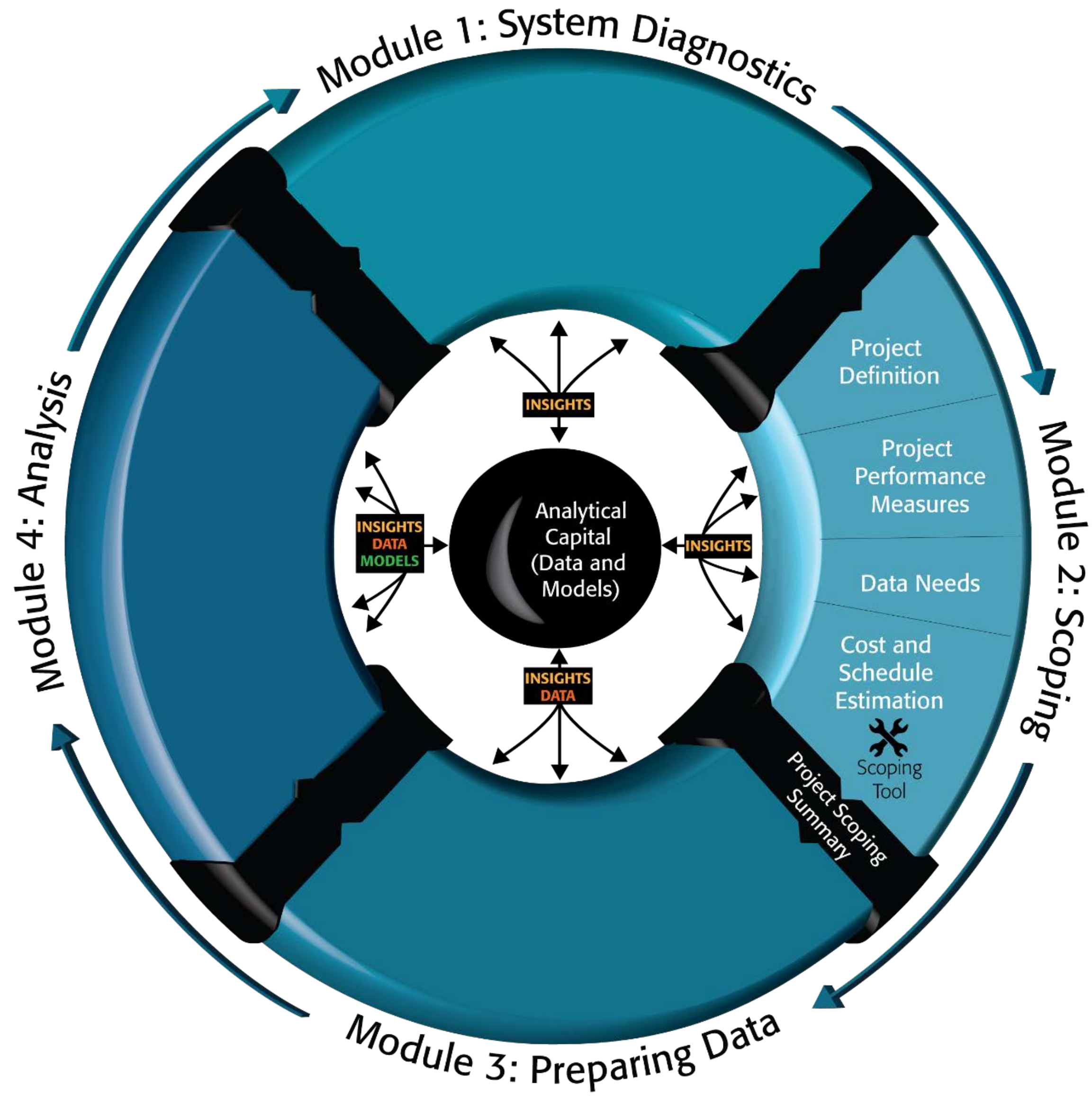

Figure 20. Diagram. Project Scoping within the 21st Century analytic project scoping process. (Source: Federal Highway Administration.)

NEED FOR AND CHARACTERISTICS OF AN ANALYTIC PROJECT SCOPING PLAN

A scoping plan provides a valuable tool for communicating the project scope, a critical element—indeed, the foundation—of success in an analytic project. The plan includes initial planning and scoping, as well as iterative updates to assumptions, scope, and agreements as the project moves forward. Scoping plan development is the primary mechanism for securing a clear and mutual understanding of expectations and assumptions among stakeholders. It helps identify flaws or technical issues in the project concept that may have been otherwise overlooked.

The scoping plan confirms both the stakeholder agreements regarding the scope of the analysis and the most appropriate approach for the analysis based on an enhanced understanding of project objectives, the study area conditions, the strategies being implemented, and the available tools and data. The benefits of developing a scoping plan include a better allocation of resources appropriate to the study objectives; a clear and shared understanding of roles, responsibilities, and expectations among project participants; and the ability of project participants to effectively communicate the project vision to the broader stakeholders. It also helps maintain agreement and project continuity as stakeholders leave positions and new staff comes in mid-stream.

The scoping plan documents the common understanding among all stakeholders regarding the purpose, nature, and extent of the proposed analytical project. It compiles project information and understanding developed to date and provides a single-source document for the proposed analysis approach. While the scoping plan must be sufficiently detailed to provide practical guidance on the actual conduct of analysis, it should retain some flexibility to adapt to project contingencies as they are encountered. The following are signs that a transportation analysis project is on track and likely to be successful:

- A detailed scoping plan is prepared that identifies appropriate analysis tools, data required, performance measures, and all parties responsible for various parts of the analysis.
- The purpose of the analysis is clear.
- Key stakeholders are engaged from the start of the analysis to its completion.
- Tools used in the analysis can convincingly demonstrate their ability to replicate observed traffic conditions using quality-checked, internally consistent, observed data.
- Interim and final results can be independently reproduced.
- Analytical results can be clearly communicated relative to analytical objectives.

A scoping plan should include the following interrelated components (the order may vary depending on project characteristics):

- Project definition.
- Geographic and temporal scope.
- Selection of the appropriate analysis tool type.
- Performance measures to be used in the analysis.
- Analysis data requirements.
- Preliminary list of alternatives to be studied, including analysis scenarios and transportation mitigation strategies.
- Expected cost, schedule, and responsibilities for the analysis.

DEFINING AND SCOPING THE PROJECT

The Analytic Project Summary prepared in Module 1 does not contain sufficient detail to perform project scoping. Additional details are needed to focus and clarify the nature and objectives of the proposed analytic effort. Project definition includes the problem statement and project goals and objectives, existing operational conditions, problem diagnosis, affected stakeholders, analysis time horizon, analysis periods, study area and geographic scope for the analysis, and modes and facilities affected.

Problem Statement and Project Goals and Objectives

This section of the scoping plan should describe the purpose of the project, provide the project background, and present the problems and issues that the analysis is intended to address. The project understanding should include clear descriptions of the transportation project purpose and need, elements that relate to the transportation problem to be analyzed, the project study area, affected communities and stakeholders, and traffic analysis objectives and hypotheses.

The overall goals and objectives for the mitigation strategies being considered should be assessed and used to shape the goals and objectives of the analysis effort. A clear definition of the "what" and "why" for conducting the analysis provides a foundation for the analysis plan. The objectives should be "SMART" (specific, measurable, actionable, realistic, and time-bound).

Document Existing Operational Conditions

Existing traffic conditions should also be documented, including average daily and peak traffic levels; directionality of traffic flow; variability of traffic flow; status of construction activities; known bottlenecks; queuing conditions; free flow and average peak speeds; and summary incident and accident statistics for the study area.

Available documentation should be compiled and reviewed by the analysis team to familiarize themselves with the operating characteristics of the study area and identify substantial issues. Individual interviews with project partners can help analysts understand study area conditions and the strategies being considered. Previous studies, archived data systems, and accident/incident data reports can all provide valuable insight into the current operational characteristics of the study area. The analysis team should seek out information on the mitigation strategies being considered and the impact these strategies have had in other regions.

The scoping team should visit the study area to gain a better understanding of traffic conditions and characteristics. The site visit(s) should include a comprehensive review of the different facilities, modes of transportation, and major mode transfer locations throughout the study area. The site visit may include visits to the regional traffic management center or toll authority, as appropriate. Depending on the characteristics of the mitigation strategies being considered, the scoping team may want to plan to visit the site on multiple occasions (e.g., peak period versus off-peak, or good weather versus inclement weather) in order to gain further insight into how traffic characteristics vary.

In documenting existing conditions, it can be useful to analyze and document the factors that influence congestion in the study area (e.g., frequency of special events). This analysis activity eventually feeds into the identification of analysis scenarios completed in a subsequent step. The analysis of influencing factors can include demand variations, incidents, or weather. This analysis helps project analysts and stakeholders to better understand the causes of congestion in the study area and the frequency with which these causal events occur. Data related to these factors can be compiled and analyzed to illustrate the effects of the factors on existing traffic conditions in the study area.

In the scoping activity, the description of operational conditions can be more qualitative than the systematic methods of characterizing operational conditions discussed in Module 3. However, if past projects have conducted more detailed and quantitative characterizations, these can serve as references to assist in the high-level description.

Identification of Underlying Causes

Key outcomes of the assessment of the existing operating conditions can be thought of as a "problem definition" and a "problem diagnosis." The scoping team should carefully evaluate any needs assessments and problem definitions included in other documents to see that they are consistent with the existing conditions data and material compiled. If a modified or more discrete problem definition is required, the scoping team should work with the stakeholders to firmly define the problem being addressed. The problem diagnosis should include a more thorough assessment of the study area conditions to ensure that the needs are properly defined, underlying causes are documented, and a small number of relevant hypotheses are created about what will help resolve the issue or mitigate the problem. Reviewing results from previously conducted assessments and comparing these with high-level assessments of existing conditions data helps identify the likely cause(s) and extent of the identified problem. Additionally, the scoping team should carefully assess any project goals and objectives identified to date and map these to the problem diagnosis to evaluate their applicability. Beginning to assemble and evaluate influencing factors in this way affords an opportunity to identify the best combinations of multiple scenarios that are most representative of actual conditions.

Affected Stakeholders

It is essential to identify a complete set of stakeholders and partners who fully represent the agencies and organizations affected by the project. Stakeholders and partners include representatives of agencies from different jurisdictions managing parts of the study area, and components and modes impacted by the strategies (e.g., highway or roadway agencies, transit agencies, program managers and stakeholders, freight industry groups, bike/pedestrian groups).

Developing a stakeholder and partner database from the beginning is important because it can provide a mechanism for tracking contact information, special concerns, and stakeholder engagement. At a minimum, it can be helpful to track stakeholders and partners by name, organization, segment (state/local/private sector), title, role on the project, mailing address (with state, city/county noted in manner that can be sorted), and the individual's contact information.

Project-specific stakeholders and partners may include more technical stakeholders and could include nontraditional members, such as emergency responders, toll authorities, and media representatives, depending on the priorities and objectives of the project (and proposed alternatives to be assessed). Guiding principles help stakeholders and outside interests better understand the focus and boundaries of the analysis effort and help ensure that any key stakeholder concerns are honored as part of the process. A key principle that should be applied in all analyses is "The overall analysis effort must take place within the budget and timeframe specified in the Scoping Plan." To minimize the risk of having to redo parts of the analysis late in the process, agencies with reviewing and/or approving authority over the analysis should be at the table from the start of the project. At the very least, project partners should have an opportunity to review and comment on the project statement as early as possible.

Analysis Study Area, Time Periods, Time Horizon, and Modes and Facilities Affected

Once an initial understanding of the study area and its operating characteristics has been defined, stakeholders refine their understanding in light of the area's major transportation issues. These issues form the foundation for analysis scenarios and mitigation strategies likely to be considered in the analysis. Analysis scenarios are developed based on the study area's geographic scope; infrastructure and facilities; and causes and patterns associated with recurrent congestion (capacity, weave zones, etc.) and non-recurrent congestion. The scenarios also consider specific causes of non-recurrent congestion, such as traffic incidents (number of incidents per day in study area, number of lanes blocked, response time, high-frequency crash locations, root cause where known [e.g., merge or weave zones, lane drops, physical characteristic such as a blind curve], incident response protocols for various incident types, etc.). The problem definition and diagnosis documented in the assembly and analysis of the existing conditions completed in the previous steps form the foundation for the identification of suitable alternatives.

Although the alternatives identification discussed in this section is presented as a linear process, in reality it is typically an iterative process. The initial alternatives identification takes place in close concert with the design phase—formulating likely strategies and combinations of strategies, then mapping these to the existing conditions outputs generated in the previous stage. As the analysis continues and the initial results are reviewed and shared with the design team, modifications or new alternatives may be proposed as existing alternatives are found to be impractical or result in unforeseen negative impacts. For example, an analysis of a strategy may reveal that the strategy is creating a bottleneck at a downstream location that was not foreseen prior to analysis. This unexpected result may promote a change in the strategy that causes the analysis to be re-run for alternatives containing the strategy. In reality, the alternatives definition and design process continue in an iterative manner throughout the process. Please refer to Module 1 for more information on system performance.

Scenarios should be developed for the range of operational conditions of greatest interest to the site in light of its analysis objectives. While traffic incidents are often the single largest cause of non-recurrent congestion, stakeholders are encouraged to investigate and understand other influences, including special events, weather, fluctuations in demand, and work zones. This initial analysis includes developing a preliminary understanding of both supply side (infrastructure/capacity) and demand-side influences on the study area across all modes (including underlying causes of demand, such as directionality or day of week). The goal is to identify potential issues (where demand exceeds supply to an extent believed to interfere with study area performance) and opportunities (underused capacity/supply that could potentially help absorb demand).

Practitioners should compile data on the frequency and severity of conditions linked with elevated congestion levels. This data creates the foundation for identifying “operational clusters” that characterize the operational conditions found in the study area and can later be used to organize the analysis (described in Modules 3 and 4). Comparisons or distributions of various sources of delay should be assembled and evaluated to identify the relative frequency of events/conditions related to congestion. From this assessment, practitioners should critically assess the potential impact of various scenarios. Scenarios identified as having a low frequency of occurrence and limited impact should be considered with lower relative priority for the analysis effort, since in the final analysis these scenarios has little overall impact on system performance. This problem diagnosis based on severity and frequency of scenarios may reveal needs previously unknown to the practitioners, help to weed out inconsequential scenarios, and greatly assist in targeting resources to provide the greatest expected value from the analysis.

The data analysis required in this step includes identifying the frequency and likely impact of the scenario. Scenarios recording the greatest frequency and the greatest impacts should be given the primary priority. Scenarios with a low likelihood but major impact (e.g., major snowstorms) and scenarios with a frequent occurrence but limited impact (e.g., minor incidents occurring on otherwise normal days) should be assigned secondary priority. Scenarios with low frequency and low impacts should be assigned tertiary priority. Often impacts for tertiary scenarios can be inferred or estimated from other results, so direct modeling of these scenarios may be eliminated from the project scoping plan. There are no set thresholds for the inclusion or dismissal of scenarios. The analysis team needs to apply engineering judgment and common sense at this point in the process. This section of the scoping plan should include the following components that help define the analysis as clearly as possible:

- **Study Area, Facility Types, and Affected Modes**. The spatial extent of the study area includes intersections, highways, and other facilities to be analyzed. The study area must cover beyond the end of the full spatial extent of queues and congestion in the baseline and future years of analysis. All transportation modes in the study area (heavy or light rail, bus transit, High-Occupancy Vehicle [HOV] or High-Occupancy Toll [HOT] lanes, bicycle, etc.) should be included in the analysis as deemed necessary; the consideration of mode shift may be important in the analysis of near- and longer-term mitigation strategies in the study area.

- **Analysis Time Period**. Define the project analysis time period, (AM/PM/Midday peak hour and/or peak period, off-peak period, etc.). This time period must cover the beginning and end of full temporal extent of queues and congestion in the baseline and future years of analysis.

- **Alternatives Definition**. Scenarios should include geometric and operational alternatives to be analyzed and compared to the baselines. Generally, it is a good idea to define a finite set of alternatives to be analyzed but also allow for a small number of presently unknown alternatives to be identified in the future. Module 4 is devoted to Alternatives Analysis.

- **Analysis Time Horizon**. A future baseline model (or future no-build alternative) is an essential part in the analysis process. It is the basis for comparison between alternatives but in a future time horizon as required by analysis needs. For the purposes of transportation analysis, a common methodology for developing future demand forecasts is a regional Travel Demand Model. This model takes into account regional growth due to land use, demographics, and socioeconomic activity. In cases where a demand model does not exist, it is acceptable to use a trend projection of travel demand.

SELECTING THE APPROPRIATE ANALYSIS TOOL TYPE

Once the analysis scenarios and strategies are identified, the analysis team begin to explore and select the appropriate analysis approach to be applied. This is a critical step in the analysis methodology; selecting the appropriate approach and tool type (and the selection of a specific tool) ensures that the analysis meets the needs of the study and streamlines the analysis process. This step follows the preliminary scoping steps because before they can select a tool, project partners must first understand the mitigation strategies, scenarios, and operational conditions of interest for the analysis and have a general grasp of the available data. One critical assessment is whether the questions posed in the project are inherently retrospective or if a predictive or prospective analysis is required. Useful retrospective analyses may be conducted through the analysis of archived performance data to differentiate the effectiveness of distinct, current operational practices. However, for new operational practices, a prospective analysis using a predictive tool is likely required. If partners want to model the potential impact of some proposed (but not yet implemented) road pricing strategies, it is important to select a tool that can accommodate this. Modeling of traveler diversion requires a combination of tools to produce results that can be assimilated to build the desired understanding. This is often a multi-iteration process where the analysis team initially focuses on identifying a high-level category of tool to use, then focuses on identifying key capabilities provided by different tool types, and ultimately selects the specific tool to apply in the analysis. Since transportation analysis science and applications are continuously evolving, as are the performance measures used in transportation analysis, the contents of this scoping guide will evolve in light of new developments in the transportation analysis field. Key steps in the evaluation and selection of the analysis tools are identified below:

- **Research and identify available analysis tool type(s) for the study area**. In this step, the analysis team researches and compiles information on models and analysis tool type(s) currently in use in the region. This may include models used on a continual basis in the region and individual models used for specific, one-time analysis in or near the study area. For each model or tool, the analysis team should identify the following:
  - The analysis package or tool (name and version of the software).
  - The year of the analysis.
  - The time periods available for analysis.
  - Facilities represented in the mode.
  - Modes represented in the tool.
  - Any special scenarios available in the model (e.g., incidents, special events, weather).
  - High-level assessment of capabilities and limitations.

The availability of a model or tool in a region is a first step; it should not be used as the sole determinant for selecting the models and tools used in the analysis. In subsequent steps, the analysis team assesses their individual needs and maps these to an appropriate tool or combination of integrated tools. This scan of available tools helps educate project partners about the range of tools available and their relative advantages and disadvantages. It is useful for selecting the appropriate set of tools and for identifying potential data sources for the model development. The Federal Highway Administration (FHWA) Traffic Analysis Toolbox is a useful resource for this step.

- **Identify factors for selecting tool type(s)**. The analysis team identifies and performs a critical analysis of the key factors that determine the required robustness of the analysis toolset selected. The resulting analysis is used in a subsequent step to map project needs to an appropriate tool or combination of tools.

- **Select the appropriate tool type(s)**. Once the scan of available tools is complete and the analysis team has had the chance to identify and critically analyze the selection criteria based on the project needs, the team is ready to select the tool type(s). The FHWA Traffic Analysis Toolbox[16] initiative includes a spreadsheet-based decision support tool that the team can use to weigh various factors to determine the appropriate tool(s). The decision support tool does not recommend a specific software vendor's tool; it identifies an appropriate category of tool based on the input factors. The team must evaluate specific vendor products within these categories. Figure 21 provides an overview of the basic factors the FHWA Traffic Analysis Toolbox method considers.

The tasks completed to develop the scoping plan to this point provide most of the input necessary to complete Categories 1 through 6 in the FHWA Traffic Analysis Toolbox Decision Support Methodology, shown in Figure 21. Particular care should go into assessing the seventh category on Tool/Cost Effectiveness. Assessing this final factor involves evaluating the cost-performance tradeoffs associated with qualifying tool options that satisfy the first six criteria. Once a qualifying set of candidate tools has been identified, the analysis team determines which tools can deliver the greatest value for the estimated cost in software and configuration/calibration load. This determination is influenced, in part, by the general understanding of available data (and workload required to render that data useful for the modeling effort, which may vary by tool in light of their various capabilities), staff skills, previous modeling efforts that can potentially be leveraged for this effort, and so forth.

In assessing the needs of project analysis, multiple tools may be required. A single tool may not be sufficiently robust to handle the analysis needs, and the analysis team may need to consider integrating the analysis capabilities from multiple tools to achieve the necessary abilities. In this step, the analysis team also specifies requirements for interfacing between different tools.

[16] Traffic Analysis Tools, http://ops.fhwa.dot.gov/trafficanalysistools/index.htm.

Once appropriate tool categories have been selected, the analysis team uses the documentation provided with the FHWA Traffic Analysis Toolbox Volume III to research the range and capabilities of individual software packages and tools within the selected category.

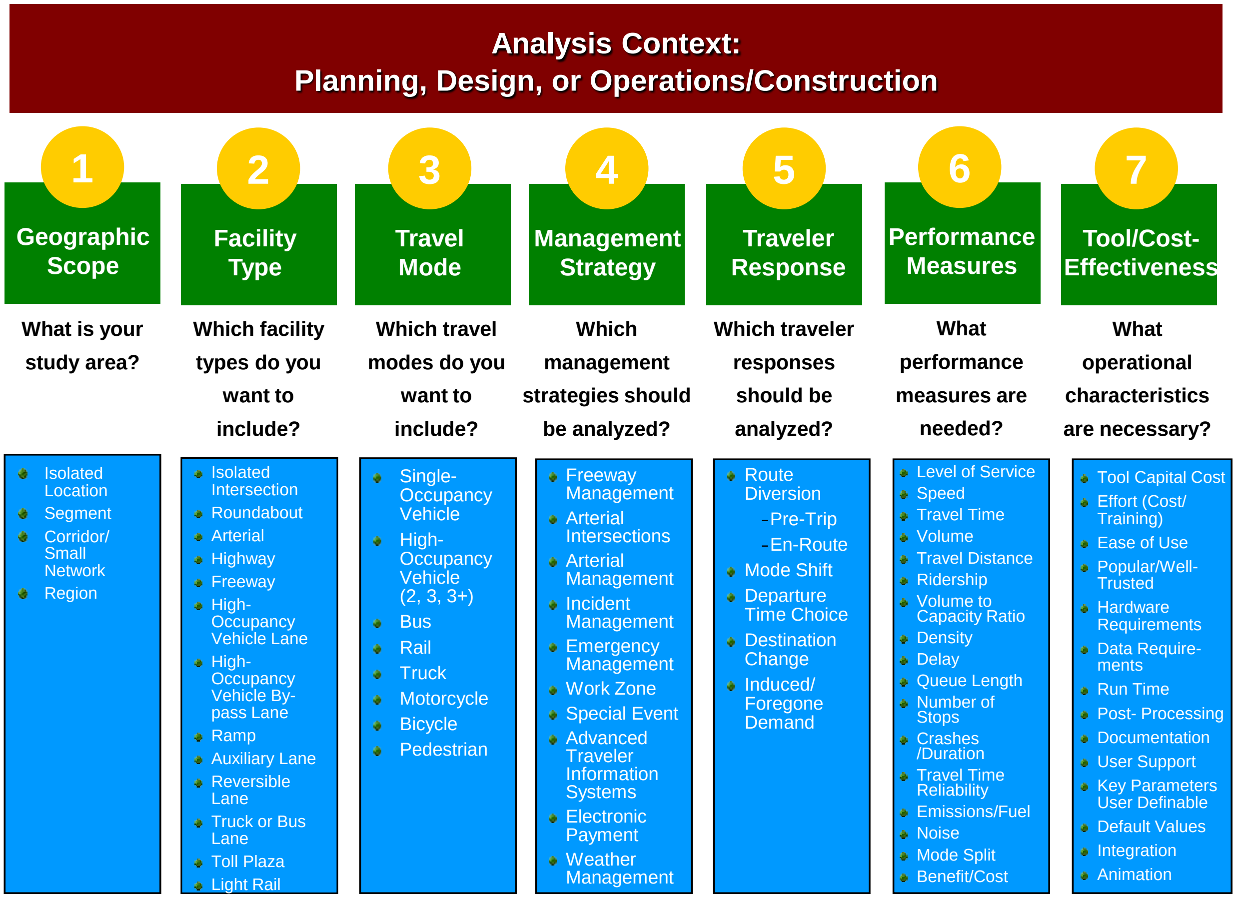

Figure 21. Flowchart. Federal Highway Administration Traffic Analysis Toolbox: Overview of analysis factors to be considered in selecting appropriate analysis tools.
(Source: U.S. Department of Transportation.)

The U.S. Department of Transportation (USDOT) documentation includes web links to individual research organizations and vendors supporting the various packages for more information; peer research can be valuable as well. From this research, the analysis team can make a high-confidence decision regarding the specific tool—or combination of tools—that best meet the requirements of the analysis project's needs.

Accuracy and Precision Requirements for the Analysis

Generally, there are three sources of uncertainty related to the conduct of transportation analysis:

- Uncertainty associated with **forecasting** future travel demand, future economic factors, and/or other traveler-related characteristics, such as a traveler's willingness to pay for use of various components of the transportation system. The analyst has little control over "unknowables" in the future, but techniques can be used to minimize or bound future uncertainty; for example, the future location of jobs will determine commute patterns and the congestion along certain corridors. These techniques typically involve sensitivity or risk analysis implemented by varying analysis parameters and assumptions using probabilistic distributions of these parameters based on "back-casting" methods. These methods result in a probabilistic representation of future impacts that represent a statistical likelihood that a project will have a certain impact. As long as analytical assumptions and parameters do not vary across different alternatives, the analysis can produce valid conclusions on the **relative effectiveness** of different alternatives to, say, improve mobility.

- A second type of uncertainty in transportation analysis relates to **traveler responses** to different operational conditions and deployed mitigation strategies. Traveler choices are being affected by different operational conditions (e.g., rain versus no rain) and by different mitigation strategies deployed to address a transportation problem; depending on the prevailing operational conditions, travelers may leave their origin earlier or later or even choose a new destination. There are very limited data on this type of traveler responses and the motivation behind each response. Until such data become available, transportation analysis must rely on a robust sensitivity analysis that would test the analysis assumptions by varying the relevant analysis parameters until a satisfactory solution has been reached.

- A more manageable type of uncertainty in transportation analysis relates to the ability of analysis tools to **replicate observed conditions and traveler behaviors**; this practice is usually referred to as "model calibration." The remainder of this section focuses on how this type of uncertainty can be mitigated so that more trusted models can be produced to analyze the effects of different mitigation strategies.

Model calibration is the process of systematically adjusting model parameters so that the model can reproduce the observed traffic conditions. The process continues until the error between the performance measures taken from the field data and the performance measures calculated in the modeling is less than a predetermined margin of error. Once the model reproduces observed conditions, model calibration can focus on specific performance measures, such as volumes, speeds, travel times, and bottlenecks. In the model calibration process, there needs to be a tradeoff between required precision and available resources to collect data and conduct modeling.

The goal of calibration is to make the model represent local, observed traffic conditions as closely as possible. However, since traffic may vary greatly from day to day, it is not possible for one model to accurately represent all possible traffic conditions. Most modeling software are developed using a limited amount of data. The model parameter values are estimated using the limited data.

Driver behavior differs by region, and it may differ significantly from normal to non-typical days. For example, poor visibility, severe weather, incidents, presence of trucks, and pavement conditions all have an impact on driver behavior. We do not recommend using a model developed using data from one region to represent future traffic conditions in another region. Investment decisions made using a model that has not been calibrated to local field conditions will be flawed.

Guidance on the overall model calibration process is presented in Volume III of the Traffic Analysis Tools suite. The overall model calibration process can be divided into four main steps, as described below:

- Identify the performance measures related to the analysis objectives and critical locations for the models to be calibrated against.
- Determine the strategy for calibration, consistent with Volume III of the Traffic Analysis Tools suite.
- Determine the statistical methodology to be used to compare modeled results to the field data. Identify Tolerance E (measure of precision) and Confidence Level (measure of accuracy or variability among samples).
- Conduct model calibration runs following the strategy and conduct statistical checks. When statistical analysis falls within acceptable ranges, then the model is calibrated.

The updated Volume III of the Traffic Analysis Tools suite update allows for the analysis of projects under different operational conditions with a better-expected statistical fit than a "typical day" trying to represent a reliability space including operational conditions such as different incidents, work zones, weather events, special events, etc.

During the model scoping process and the development of the data collection plan the types of required field measurements are determined, including speeds, volumes, queuing, and other congestion observations at different locations in the network. Since traffic conditions fluctuate daily, it is important to obtain field data from multiple days. The multiple days of data serve as a database in which field variations are used to determine the tolerance of error in the simulated results.

During model development, it is recommended that the spatial and temporal model limits extend beyond where and when the congestion in the field occurs. The statistical calibration should not necessarily include every link in the model, but it should focus on the critical design elements within the primary study area in the model.

The number of locations selected for comparing the performance measures in field data against model outputs needs to be balanced against the quality and location of the available data, the desired level of statistical confidence, and the availability of resources. The model outputs and reporting for these statistical tests should be similar to the performance measures that will be used later on in the analysis of the study area.

The selection of the number of data days to be used should be based on an analysis of available data in terms of what data is available and cost-effective to collect. In an urban area where freeway sensor data are archived and readily available, more days of data can be used. In areas where there is no surveillance and where manual or temporary data collection devices are used, the amount of data to be collected is more resource-intensive.

Cost Implications

Generally, building and calibrating analysis models for large or more complex study areas requires a disproportionately greater level of effort compared to smaller projects. This is attributed to the increased complexity of modeling travel demand and traffic operations in a larger geographic area and to more extensive project reporting, alternatives analysis, and presentation requirements. Seemingly similar projects may require different levels of effort for a number of reasons:

- Experience of project manager, analysis team, and reviewers.
- The project purpose, objectives, and scope.
- The availability of good data for model calibration.
- Temporal and spatial resolution requirements for the analysis.
- The number and complexity of the alternatives being analyzed.
- Performance measures used.
- Software used.
- The amount of documentation, meetings, and presentations required.
- Number and effectiveness of project reviews conducted.
- The extent of stakeholder involvement.

SELECTING PERFORMANCE MEASURES

Early in the project, the analysis team defines performance measures in line with the project objectives, mitigation strategies under consideration, analysis scenarios, and operational conditions (shaped by the understanding of available data) identified for the project. This provides the "home" where these measures are documented in the scoping plan, and helps the analysis team crystallize its vision and scope for the analysis effort.

In order to begin to understand how the proposed transportation improvements perform and whether they meet stakeholder expectations, and even to help stakeholders develop realistic expectations for the improvements, project managers must first be willing to articulate them. The analysis effort helps illuminate which expectations are realistic, which may be unrealistic, and why. It identifies opportunities that optimize the study area's transportation network by allowing analysts to experiment with adjusting mitigation strategies for the price of a model run rather than myopically making such adjustments to the actual deployed system, where the cascading second and third-order effects are more difficult to perceive in real-time.

Performance measures should be closely tied to the identified overall project goals and objectives and the expected traveler responses. For many improvement strategies, it is important to consider a set of performance measures that are sensitive to recurring and nonrecurring congestion. The scoping plan should identify the selected performance measures and the approach for calculating the performance measures based on the expected model capabilities and available data.

An effective way to identify appropriate performance measures is to test one or more specific hypotheses for each objective. These hypotheses can indicate a change in travel conditions (such as: *The strategies will reduce travel times during an incident by 5 percent.*) or can be neutral in the prediction of an impact (such as: *The strategies will not result in a change in emissions rates*.). Performance measures that support the testing of the formulated hypothesis should be identified. Using this method ensures that the performance measures are appropriately mapped to the project goals and objectives.

To compare different investments within a study area, it is important to define and apply a consistent set of performance measures. The performance measures should:

- Provide an understanding of travel conditions in the study area, including localized and system-wide metrics representing impacts in the immediate vicinity of the proposed improvement and in the larger study area.
- Be consistent with lead agency overall performance measures used to evaluate all sorts of transportation improvements.
- Demonstrate the ability of improvement strategies to improve mobility, throughput, and travel reliability based on current and future conditions.
- Help prioritize individual investments or investment packages within the study area.

To the extent possible, the measures selected should be reported for the overall system and by:

- **Mode—**Single Occupancy Vehicles (SOV), High Occupancy Vehicles (HOV), transit, freight, etc.
- **Facility Type—**Freeways, expressways, arterials, local streets, etc.
- **Jurisdiction—**Region, county, city, neighborhood, and study area-wide.

Transportation analysis performance measures typically focus on the key areas described below. However, customized measures may be selected based on unique impacts of individual mitigation strategies. There are five key performance areas:

- **Mobility**. Mobility describes how well people and freight move in the study area. Mobility performance measures are readily forecast. Three primary types of measures are used to quantify mobility: travel time, delay, and throughput. Travel time and delay are fairly straightforward to calculate using model outputs. Throughput is calculated by comparing travel times under the incident scenarios to those under no incident. By comparing the percentage of trips under the same threshold travel time in the pre- and post-mitigation scenarios, the relative influence of the strategies on reducing extreme travel times can be estimated.

- **Reliability and Variability of Travel Time**. Reliability and variability capture the relative predictability of the public's travel time. Unlike mobility, which measures how many people are moving at what rate, reliability/variability measures focus on how mobility varies from day to day. Travel-time reliability/variability is reported in terms of changes in the Planning Index and changes in the standard deviation of travel time.

- **Transportation Safety.** This is another performance area that may be of interest to transportation analysis. Safety is typically measured in terms of accidents or crashes in the study area, including fatalities, injuries, and property-damage-only accidents. Currently available safety analysis and prediction methodologies are not sensitive to transportation improvement strategies. At best, available safety analysis methods rely on crude measures such as Volume-to-Capacity ratio (V/C) or empirical comparison methods, such as identifying safety benefits that result from implementing a certain type of mitigation strategy and applying the same expected improvement rate to a future implementation of the same or similar strategy. Clearly, this is an area deserving new research.

- **Emissions and Fuel Consumption**. Emissions and fuel consumption rates are used to produce estimates based on variables such as facility type, vehicle mix, and travel speed.

- **Cost Estimation**. Planning-level cost estimates are based on life-cycle costs, including capital, operating, and maintenance costs. Costs are typically expressed in terms of the net present value of various components. Annualized costs represent the average annual expenditure that is expected in order to deploy, operate, and maintain the transportation improvement and replace equipment as they reach the end of their useful life.

In attempting to optimize a system's performance across all these measures, the analyst may be faced with conflicting objectives and constraints as they relate to different performance measures. For example, minimizing travel time or maximizing travel speed are often in conflict with minimizing emissions of fuel consumption. In order to overcome this conflict, the stakeholders and analysts must make tradeoffs that typically result in less-than optimal system elements for a specific measure but also result in a system that strikes a balance between multiple performance measures, hopefully optimized in different ways for different operational conditions.

ANALYZING DATA REQUIREMENTS

The precise data requirements for developing and calibrating a transportation analysis model vary depending on the software tool selected; however, they all require five basic types of input:

- Roadway geometry.
- Traffic control data.
- Travel demand, traffic volumes, and intersection turning movements.
- Performance data, such as queue locations, queue lengths, travel times, and speeds.
- Data on vehicle characteristics, such as vehicle classifications or vehicle mix.

As models are calibrated to the data collected for the study, the true validity of any model is dependent on the quality of the data that goes into it. To get an early estimate of effort required, data requirements should be evaluated during the initial stages of a project. Quality/variability of existing data affect sample sizes, so early statistical evaluations of the data (Margin of Error) can prove highly valuable for subsequent development of effort required. Key model calibration performance measures should be identified early on to help determine the data needed to estimate these performance measures.

Since data collection can be one of the most important components of an analytical study, it is critical to identify the key data needed for the study early in the process. Often, it is possible to use existing data, but the data may be outdated or from different timeframes for different parts of the network. In that case, resources need to be allocated for new data collection. If there is limited funding, resources need to be spent judiciously to collect sufficient quality data to conduct a study that helps inform decision-makers of the potential implications of their proposed transportation investments.

Given the importance and costs of data for the modeling process, it is important to document the information. Documentation methods should include a clear file-naming structure, an explanation of the sources of data, and a database structure that can be readily incorporated into model inputs and reports. Efficiency of managing the data can be further enhanced through the development of automated procedures. Module 3 in this guide provides more detailed guidance on data storage, quality assurance processes, and metadata requirements.

Types of Data

Types of data can be categorized in three main areas: travel demand, traffic control, and physical geometry. Travel demand data include traffic counts, vehicle classification counts, speeds, travel times, congestion, and queuing observations; these data require the majority of the data collection effort. Traffic control data include signs, signal control, and timing plans. Physical geometry can be obtained from rectified aerial photography and base mapping files that may be prepared as part of the design effort for projects. The tenets of data collection and data management that are important for conducting an effective analytical study include the following:

- The type of analysis being conducted should dictate the quantity of data collected.
- The required accuracy should drive the quantity of data collected.
- Since the quality of the analysis and the resulting decisions depend on the data used, it is important to use data that is recent, internally consistent, and relevant.
- Collection of traffic data should capture the temporal variations in demand and performance and should be finite enough to allow for an appropriate analysis to be conducted.

Helpful tips on data collection include:

- Do not simply rely on the veracity of historical data; allocate resources to verify the data through current field inspection.
- Measure flows and estimate or forecast demand—procedures for estimating demand are given in Volume III of the FHWA Traffic Analysis Toolbox.
- Check flow conservation—entry flows should match the exit flows.
- Always check for data quality—use the most recent Geographical Information System (GIS) files, maps, and drawings.
- Use counts from before the onset of congestion to after congestion dissipates.
- Broaden the geographic scope to cover bottlenecks and spatial and temporal extent of queues.

Data Sources

- **Travel Demand**. The basic demand data needed by most transportation analysis software are the entry volumes—the travel demand entering the study area—at different points of the network. At intersections, the turning volumes or percentages should be specified.

- **Origin-Destination (O-D)**. O-D trips can be estimated from a combination of travel demand model trip tables and traffic counts. While O-D data can be obtained from the local Metropolitan Planning Organization's (MPO) regional travel demand model, these are generally 24-hour estimates, so they must be adjusted and refined to produce hourly or peak period estimates for use in simulation or other transportation analysis models; typically, these estimates are further disaggregated to represent 5- or 15-minute estimates required for more detailed analysis. License plate matching surveys can be used to estimate hourly trips, but this is resource-intensive. Depending on project complexity, a cost-effective way is to adjust the travel demand model O-D trip estimates based on field counts. If the study area has transit, HOV, and trucks in the vehicle mix or if there is significant interaction with bicycles and pedestrians, corresponding demand data would be needed. In addition, vehicle dimensions and vehicle performance characteristics (e.g., maximum acceleration and deceleration) are required. Even if only the peak periods are being examined, demand data should be collected before the onset of congestion and should continue until after the congestion is dissipated. Also, to capture the temporal variations in demand, it is best not to aggregate demand data to longer than 15-minute intervals.

- **Vehicle Characteristics**. This data can be obtained from the state DOT or air quality management agency. National data can be obtained from car manufacturers, the Environmental Protection Agency (EPA), and the FHWA.

- **Traffic Control Data**. Data from traffic control devices at intersections or junctions are required. Control data refers to the type of control device (traffic signal, stop sign, ramp meter, etc.), the locations of these control devices, and the signal-timing plans. Traffic control data can be obtained from the agencies that operate the traffic control devices in the given study area. Traffic operations and management data on links are also needed. These include location and type of warning signs—for lane drops, exits, guide signing; type and location of regulatory signs. If there are HOV lanes, information on the HOV lane requirement (HOV-2 versus HOV-3), their hours of operation, and the location of signs are needed; if there are HOT lanes, information on the pricing strategy is required.

- **Operational Conditions**. If there are Variable Message Signs (VMS) in the study area, the type of information displayed, the location, and—where possible—the actual message displayed are needed. Most of this information can be obtained from the public agencies responsible for operating the VMS. The type of signs and locations can be obtained from GIS files, aerial photographs, and construction drawings. Event data can be received from public agencies, such as the Traffic Management Center (TMC) logs. Crash databases should be verified since data may not always be recent and may not be for the specific study area. The team can usually obtain workzone data from state DOTs and weather data from the National Oceanic and Atmospheric Administration (NOAA). Overall, the data should be from concurrent timeframes so that it is easier for the analysis to establish cause and effect.

- **Transit Data**. This data can be obtained from the local and regional transit operators. It can include schedules and stop locations. Calibration data could include Transit Automatic Vehicle Location (AVL) data, boarding and alighting data, and dwell time at stops.

- **Mobile Source Data**. Mobile source data include data derived from mobile phones, blue tooth devices, and other mobile sources can be used to augment other data collected. The primary type of data obtained from mobile sources are speeds and travel times. In most cases, the mobile source techniques use samples of vehicles in the traffic stream, but they may not be reliable sources for traffic counts and vehicle composition. Mobile source data are also being used to develop Origin-Destination (O-D) matrices or trip tables. The mobile source data are typically obtained, stored, and sold by private vendors. Before purchasing these types of data, it is a good idea to have a demonstration of the data and a means to compare the data supplied by the vendor with a real observation of what is happening in the field. Consider how the data are provided (format, structure, and software) and take into account how the data will be used for traffic modeling and other purposes.

Challenges with Data

Going through a systematic process of collecting the critical data, verifying data quality, and documenting any assumptions are key to justifying the results of a study to decision-makers and the public. A statistical analysis of collected and previously available data can be helpful in determining the statistical data variability and the margin of error contained in the data.

Data Comprehensiveness. Comprehensive data cover different performance measures (volumes, speeds, bottlenecks, queuing, and congestion data) across freeways and arterial streets, as well as transit data and incident data. Traffic counts should be taken at key locations in the study area; key locations include major facilities (freeway segments, major intersections, and interchanges) and major on- and off-ramps. If possible, data collection should be done simultaneously at all key locations. Otherwise, the counts should be taken at least during similar timeframes with similar demand patterns and weather conditions.

Data Reliability. Automated data sources are often best for collecting the long-term data needed to develop and calibrate transportation analysis models. However, many existing automated data collection systems lack the robustness or reliability to effectively compile relevant data sets. A thorough assessment of the data quality from all sources is recommended to identify any potential problems early in the process and establish methods to address any deficiencies.

Accuracy is the measure of the degree of agreement between data values and a source assumed to be correct. It also is defined as a qualitative assessment of error. It is important to have accurate, internally consistent, and recent data. If information on future traffic conditions is not available, it could be helpful to study data from other regions that may have the bottlenecks and traffic patterns that are envisioned for the study area. This is critical since the goal of model calibration is to not only see if the model can represent observed conditions, but also to examine if the model can handle the future congestion. With respect to data filtering and fusion, it is crucial to adopt standard ways to accept or reject field data and to address data gaps and missing data. Overall, data accuracy is paramount for an analysis project since it percolates throughout the analysis process.

A small margin of error in the collected data is required to increase the validity of analysis results. It is necessary to collect enough data points for each performance measure (volumes, speeds, etc.) so that the sample is an accurate representation of the mean and standard deviation of this performance measure. Depending on the mean to standard deviation ratio and the desired margin of error, the required sample size may vary greatly. Module 3 in this guide presents a discussion on cluster analysis and how it can be used to improve the analytical accuracy.

Documentation of Data Gaps

At the end of the data assessment task, the analysis team should develop a summary of what data needs are covered by in-house or available data and what primary data collection may be needed. This is a critical item because data collection is going to be costed out since estimates for data collection, quality control, and analysis will need to be developed separately from tool acquisition, from base model development and calibration, and from alternatives analysis.

REFINING ALTERNATIVES AND MITIGATION STRATEGIES

Interviews with project partners can be useful in understanding study area conditions and the improvement strategies being considered. Additional documents can provide valuable insight into the current operational characteristics of the study area: previous studies; transit data and studies on topics such as ridership and parking occupancy (supply and demand); archived data systems; and crash/incident data reports. Regional and long-term transportation plans provide insight into congestion hotspots, but these must be supplemented by more detailed studies.

The analysis team should also seek out information on the improvement strategies being considered and the impact these strategies have had in other regions. Likewise, developing an understanding of previous capacity and operational strategy projects in the study area—including an understanding of the expected and unexpected results of those projects—can be beneficial. Peer-to-peer contact with agencies that have undergone similar planning and deployment is a valuable way to gain this insight.

Once the analysis scenarios have been identified, the next step is to identify the improvement strategies and define under which analysis scenarios the strategies will be activated. It is also critical to understand precisely when (under what conditions) the strategies will be applied and how their application may vary under different conditions. To better understand what factors influence congestion and the frequency in which these factors occur, the analysis team works with the stakeholders to identify the combinations of travel demand, incidents, special events, and weather events that affect study area operations. A transportation improvement project is typically concerned with nonrecurring congestion on a level equal to or greater than typical recurring congestion levels. Therefore, it is critical that the analysis team recognize the non-typical factors that affect nonrecurring congestion.

The analysis team begins by exploring preliminary scoping hypotheses and assumptions and by identifying possible opportunities and constraints associated with the application of improvement strategies identified under specific operational conditions. For example, freeway managers may be interested in opportunities to divert drivers from the freeway to arterials as an incident management strategy. They talk with local arterial managers to understand whether the local jurisdiction can accommodate this diversion and activate these strategies to accommodate the desired performance in a sufficiently timely manner. This step also includes discussion of the potential to avoid problems—such as changing signal timing or VMS to avoid queues that may lead to collisions at critical locations. If there are contributing circumstances that can be avoided that lead to greater non-recurrent congestion, the team must devise strategies that may avoid these contributing circumstances. They also seek to understand and address any concerns or constraints the local jurisdiction may have. Is the technology in place to accommodate the needed signal timing? Can the timing strategies be activated in a sufficiently timely fashion to make the strategy feasible?

Likewise, if local agencies are interested in exploring opportunities to divert freeway or arterial traffic to transit, they will engage in collaborative dialogue with the regional transit managers to understand possibilities for creating available transit capacity, including parking facilities, to accommodate the possible influx of demand under certain scenarios.

These discussions begin to yield initial insights into the eventual performance measures for the analysis effort. Because the models have not yet been run, these are preliminary scoping discussions at this phase, the purpose of which is to begin to shape possibilities and identify limiting factors that will form the foundation of the more detailed planning documented in the analysis plan. As discussions progress and more quantitative data becomes known, stakeholders update and refine assumptions and hypotheses, bringing further clarity and detail to the hypotheses and assumptions associated with operational conditions and strategies to be analyzed. The analysis team identifies which parts of the envisioned strategies they may want to make dynamic (strategies that could be manipulated in response to changing operational conditions, such as ramp metering, HOV, or pricing strategies, or traveler behavior such as mode choice).

COST, SCHEDULE AND RESPONSIBILITY FOR THE ANALYSIS

Analysis roles must be defined and clarified among the various project stakeholders. Table 3 provides a high-level example of how key project roles can be displayed. The project team also includes a summary of the estimated budget and timeframe in the analysis plan. These will be updated regularly as the analysis effort moves forward.

Although the scoping plan is designated as a single deliverable, there may be multiple sub-deliverables (e.g., technical memos and presentations) that are generated as the plan is developed, depending on workflow and individuals responsible for compiling the necessary information. The scoping and analysis plans undergo some major revisions prior to finalization to allow for additional levels of detail as the process moves forward and to allow for stakeholder input and comment. Key output of the scoping and analysis plans includes Memoranda of Understandings/Agreements (MOUs/MOAs) among initiative stakeholder organizations, documenting project scope and anticipated roles and levels of effort.

Table 3. Example high-level allocation of analysis responsibilities.

Work Step	Analysis Project Manager	Operations Manager	Planning Manager	Modelers	Systems Manager	Stakeholders
Develop Analysis Plan	**X**	**O**	**O**	**O**	**O**	**O**
Develop Data Collection Plan and Collect Data	**X**	**O**	**O**	**O**	**O**	**O**
Model Setup and Calibration	**X**	**O**	–	**X**	–	–
Alternatives Analysis and Documentation	**X**	**O**	**O**	**X**	**O**	–

Note: × represents primary responsibility and O represents secondary responsibility.

There are many variables that affect the estimated level of effort for the analysis, including the existence of—and level of precision in—available project documentation; quantity, quality and availability of needed data; cohesion in stakeholder vision for the analysis effort; and experience of the project staff with modeling tools in previous efforts. Although each analysis varies due to these factors, a rough order of magnitude estimate of the proportion of analysis resources that may be required of the different analysis steps includes:

- Develop Analysis Plan—20 percent.
- Develop Data Collection Plan and Collect Data—20 percent.
- Set up and Calibrate Model—30 percent.
- Analyze and Document Alternatives—30 percent.

While the analysis steps are presented sequentially, in actuality the analysis process is iterative and requires flexibility in its application. The implementation of these steps must be carefully configured to the individual needs of each analysis effort and appropriately readjusted throughout the process as conditions and needs change.

The FHWA publication FHWA-HRT-13-026 (March 2014) “Guidance on the Level of Effort Required to Conduct Traffic Analysis Using Microsimulation” provides analysts and modeling managers with a frame of reference on the level of effort required to complete a transportation analysis using microsimulation.[17]

[17] Guidance on the Level of Effort Required to Conduct Traffic Analysis Using Microsimulation, http://www.fhwa.dot.gov/publications/research/operations/13026/13026.pdf.

The reader of this guide should use the labor-hour estimates as a point of reference and not as an absolute number to apply to projects. Seemingly similar projects can require different levels of effort for a number of reasons:

- The project purpose, objectives, and scope.
- The availability of data (how much data had to be collected).
- The number and complexity of the alternatives being analyzed.
- Performance measures used.
- Software used.
- The amount of documentation required.
- Experience of project manager, analyst, and reviewers.
- Number and effectiveness of project reviews conducted.
- The amount of stakeholder involvement.

A comparison of the total level of effort and the amount of effort expended specifically for building a base model and the calibration process reveals a trend that is a consistent order of magnitude regarding the model building and calibration effort with the model size. Larger models require a disproportionately greater level of effort compared to smaller projects. This is attributed to (1) the increased complexity of a larger analysis model as it relates to modeling travel demand and traffic operations in a larger geographic area, and (2) extensive project reporting, alternatives analysis, and presentation time expended.

THE SCOPING TOOL

As part of work on this guide, a software tool was developed to produce ballpark estimates of staff-hours to complete the tasks needed to support a transportation analysis consistent with discussions developed within the Scoping Guide. The objectives of the scoping tool are (1) to inform a procurement decision made by a transportation agency before procuring transportation analysis services; (2) to provide a rough order labor hour estimate for conducting the transportation analysis; and (3) to incorporate an estimate of relative risk associated with various analysis tasks.

The focus of the scoping tool is on traditional, off-line project analysis—not on analysis conducted in real-time to actively managing aspects of the transportation system. The focus of the tool is also on time-dynamic analytics, such as all types of simulation and Highway Capacity Manual analysis, but not including travel demand modeling.

The scoping tool is a simple spreadsheet tool that produces ballpark estimates of labor hours required for analysis based on a small number of user inputs. The tool first asks users to choose among small number of options related to the expected analytical complexity and risk. Based on user inputs, the tool uses factors to assess different complexity and risk levels and to produce outputs, including ranges of labor hours by task for three labor categories plus total labor hours by task. The Scoping Guide team used records from 40+ analysis projects to inform the calculation engine inside the scoping tool. The scoping tool methodology was validated through iterative stress tests. Assumptions and methodology were vetted in several review cycles, including both internal and external teams of analysis experts. User inputs and options provided by the scoping tool are identified below:

- Number of intersections in the study area.
- Number of on- and off- freeway ramps.
- Baseline model availability (No, Partial, Full).
- Baseline model calibrated? (No, Yes).
- Data collection requirements (Low, Medium or High Effort).
- Number of time periods to be analyzed (e.g., AM, mid-day and PM peak periods).
- Number of alternatives to be included in the analysis.
- Number of operational conditions to be analyzed.
- Number of analysis horizons (Baseline year plus X number of future years).
- Complexity of analysis scenarios (multiple modes, pricing, etc.—Simple or Complex).
- Complexity of analysis methodology (Deterministic or Dynamic/Stochastic).
- Complexity of output performance measures (Simple or Comprehensive).
- Analyst Experience (Some or Considerable).

Complexity and risk factors taken into account in the scoping tool include:

- A large number of intersections and/or on- and off-freeway ramps in the study area.
- No baseline model available.
- Baseline model not calibrated.
- High data collection requirements.
- Complex analysis scenarios (multiple modes, pricing, etc.).
- Complex analysis methodology (Dynamic).
- Reliability analysis (multiple operational conditions).
- Complex output performance measures (such as reliability of travel time, or emissions).
- No significant analyst experience.

At this time, there is no documentation provided with the scoping tool. Figure 22 and Figure 23 provide example screenshots from the Scoping Tool.

Transportation Analysis Project Costing Tool

Press This Button to Start

Summary of User Inputs:

1	Name of Study Area:	Standard TIS
2	Number of Intersections:	5
3	Number of Freeway Ramps:	10
4	Base Model Availability:	Yes
5	Is the Base Model Calibrated:	Yes
6	Number of Analysis Horizons:	2
7	Number of Alternatives:	2
8	Number of Representative Days:	2
9	Number of Peak Periods	2
10	Data Processing Requirements:	Low
11	Complexity of Analysis Scenarios:	Simple
12	Complexity of Methodology:	Deterministic
13	Complexity of Outputs:	Comprehensive
14	Analyst Experience:	Considerable

Note: This Transportation Analysis Costing Tool is provided "as is" without warranty of any kind, and without any documentation, user's guide, or mai

Figure 22. Screenshot. Analysis scoping tool—summary of example user inputs.
(Source: Federal Highway Administration.)

Transportation Analysis Project Costing Tool

OUTPUT REPORT
Estimate of Labor Hours Required to Complete the Analysis of:
Standard TIS

	Project Task	Manager Hours	Engineer/ Planner Hours	Technician Hours	Total Hours	Lower Bound	Upper Bound
1	Develop workplan, analysis plan, and project management	30	30	10	**70**	**60**	**80**
2	Select analysis tool	10	10	-	**20**	**20**	**20**
3	Develop data plan and process data	-	-	40	**40**	**40**	**40**
4	Define clusters and representative days	30	50	-	**80**	**70**	**90**
5	Develop and calibrate baseline model(s)	20	40	70	**130**	**120**	**170**
6	Develop future baseline model(s)	30	30	30	**90**	**80**	**100**
7	Analyze alternatives	90	180	90	**360**	**320**	**400**
8	Reports and presentations	30	30	10	**70**	**60**	**80**
	Total Labor Hours	**240**	**370**	**250**	**860**	**770**	**980**

Source: Cambridge Systematics for U.S. Department of Transportation

Figure 23. Screenshot. Analysis scoping tool—example output.
(Source: Federal Highway Administration.)

MOVING TO MODULE 3: SUMMARIZING THE PROJECT SCOPING PLAN

A scoping plan should, at a minimum, include the elements listed in Table 4. These components of the scoping plan can guide the data preparation, gap analysis, and data collection planning and execution described in Module 3, and the analysis activities described in Module 4 of this Guide.

Table 4. Project scoping summary elements.

Project Definition	A concise statement of the overall system problem includes cross-validation and other insights from stakeholders on the nature of the issue and potential solutions.
Geographic Scope	The geographic area to be covered by the analytical project includes a statement of the required detail of representation within this geographical area.
Temporal Scope	The times of day, days of week, seasonality, and years of operation are assessed in the analytical effort. This includes an assessment of the simulation horizon.
Candidate Hypothesis	The *candidate hypothesis* represents the leading underlying cause of the system performance issue.
Analytical Approach	This element describes of the proposed method for evaluating the effectiveness of the mitigating strategies in resolving the system performance issue.
Selected Tool Type(s)	The one or more tool types will be used in the analytical approach. This section should identify if existing models are to be employed, or if new models must be developed.
Data Requirements	A summary of data will be used to characterize operational conditions, represent alternatives, and model the geographic and temporal aspects of the system.
Preliminary List of Alternatives	High-level description of the alternative solutions and/or operational practices will be assessed within the analytical project.
Key Operational Conditions	The set of travel demand, incident, and weather conditions under which a meaningful examination of alternative impacts must be conducted.
Selected Performance Measures	The measures of system performance selected for the effort. These measures should be most suited to differentiate alternatives, be meaningful to stakeholders, and can be well-represented/estimated within the proposed analytical approach.
Expected Costs	The projected cost of the analytical project, including data collection.
Expected Schedule	The projected time to conduct the analysis, including data collection.
Expected Assignment of Responsibilities	An assessment of responsibilities related to the project and how those responsibilities are allocated among departments, contractors, and other organizations engaged in the effort.
Risks	A summary of risks comprising risks in data collection, technical risks, and non-technical risks.

MODULE 3. PREPARING DATA TO CONDUCT A TRANSPORTATION ANALYSIS

Module 3 describes a key element of the data-driven analytic project scoping cycle. Diverse data are integrated while ensuring temporal and geographic consistency so that the analysis team is ready to move forward to conduct and document analytics projects. Figure 24 highlights this module within the 21st Century Analytic Project Scoping Process. During the scoping process, the team identified the performance measures that reveal underlying transportation system dynamics. Data needs are then derived to calculate these measures and provide the context for their variation (e.g., variation in demand, incident patterns, and weather impacts).

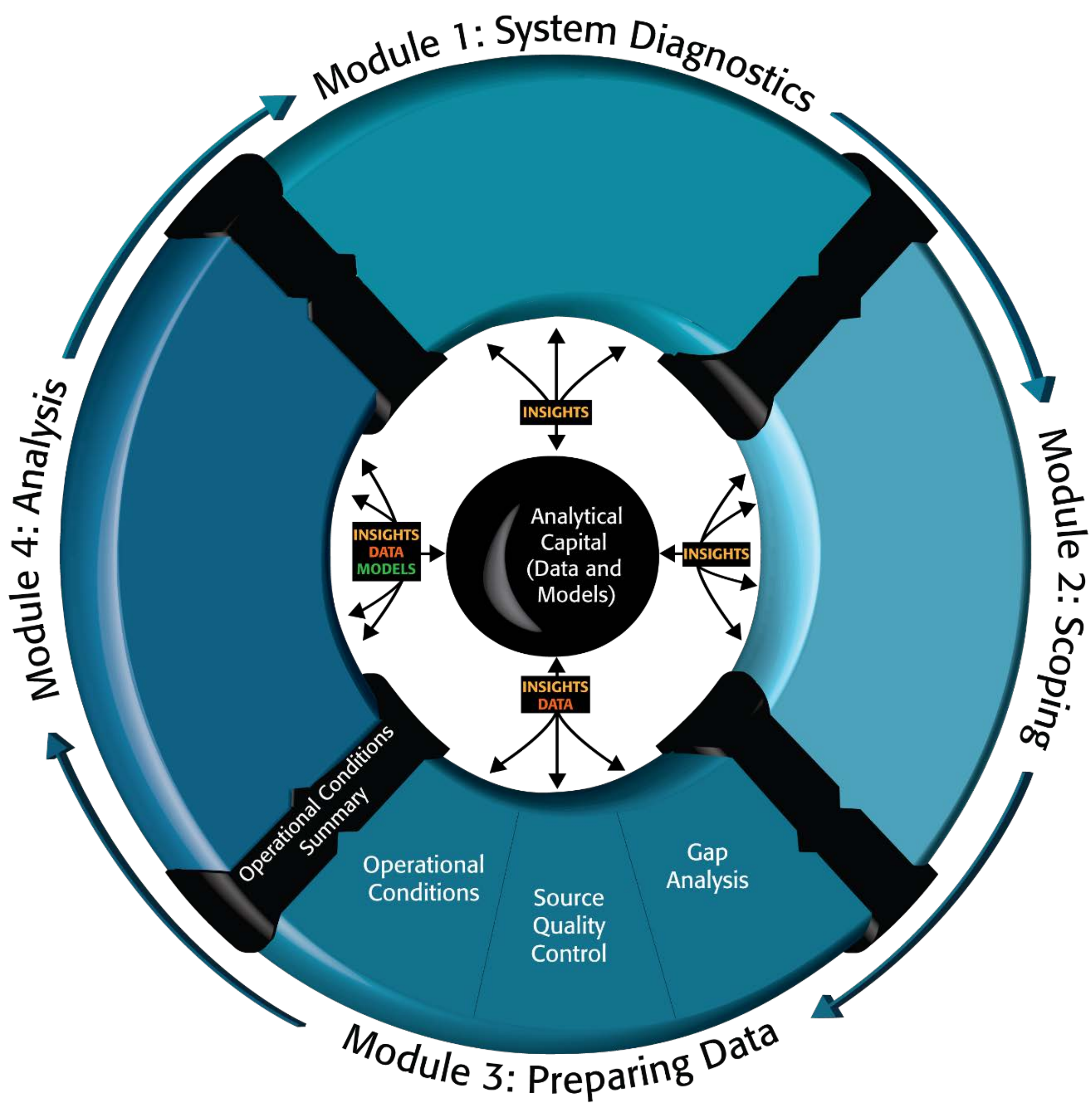

Figure 24. Diagram. Data preparation within the 21st Century analytic project scoping process. (Source: Federal Highway Administration.)

As shown in Figure 24, this module provides practical advice on assessing data gaps and needs, preparing and collecting data, and identifying operational conditions. In Module 2, the key performance measures and the analysis tool to solve the transportation problem were selected. Based on the selections, the required data to conduct the analysis is also identified. These data could be any key data whose absence would impair the ability to conduct the project. These key data may be collected by others and need only be obtained and integrated. In other cases, if targeted primary data collection is required, a data collection plan is developed and existing data are assessed and integrated. In this module, an analyst verifies the consistency and the quality of the data available and outlines the data collection plan to fill the gap between needs and availability. Depending on the nature and scope of the project, the analyst may review the available data to identify a representative set of operational conditions.

ASSESSING DATA GAPS

This section describes the types of data gaps and process for assessing these gaps to help an analyst determine if a new or additional data collection is needed to use resources as efficiently as possible. Concentrating on identified key measures, the analyst captures and assembles data that create, inform, or provide the context to understand these measures. In general, it is better to have a small set of measures and deep insight into their patterns and variations rather than a large number of measures with limited insight on their dynamics or interrelation. A data analysis also identifies the strengths and weaknesses of specific data (and data sources). For example, are the data overly smoothed or are they up-to date? If the data is not always accurate, analysis may reveal the conditions under which the data are more or less reliable.

Data gaps identification and potential primary data collection are pre-procurement activities, followed by operational conditions analysis, if required. Performance measures and analysis data requirements have been identified in Module 2, so Module 3 provides a more detailed assessment of data needs versus availability. The assessment result helps procure a properly resourced and focused analytical project in Module 4.

Data Sources and Limitations

Available data come from different sources, such as traffic data from sensors, incident data from enforcement or emergency response agencies, and weather data from weather data collection centers. Before integrating data from multiple databases, an analyst must understand how the data are collected and the source of the data from the existing studies in order to identify data gaps and needs. When preparing traffic data to conduct a microscopic simulation for AM peak period, a data analyst first identifies the AM peak hours based on the historical traffic data or previous studies because different locations or transportation modes have different peak periods. A simulation time horizon may last longer than the identified AM peaks when the projected performance for benefits last longer. If the existing data source cannot serve this need, a data analyst documents the gap. This gap may be resolved through an additional data collection activity or by integrating other data. Data resolution (e.g., 5 min., 15 min., or hourly) is another key component in this exercise. A higher data resolution is needed to evaluate an intersection traffic signal control performance compared to an annual freeway corridor performance study. The analyst must determine whether the data resolution is sufficient for the project, depending on the nature of the problem and the scope of the project. Figure 25 shows an example of using existing travel-time data to virtually identify the AM peak hours (from 6:00 a.m. to 10:00 a.m.).

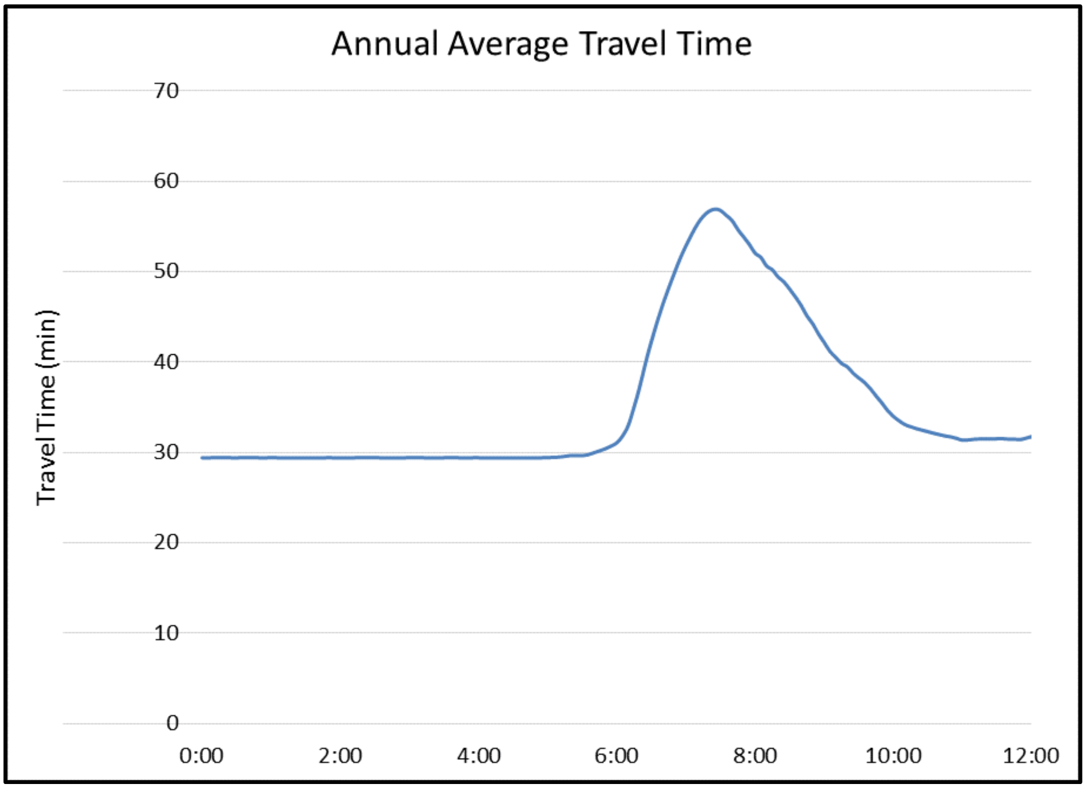

Figure 25. Chart. Annual average corridor travel time profile on Seattle I-405 south bound in 2012.
(Source: Federal Highway Administration.)

Temporal and Spatial Consistency

Consistency is one of the major concerns of using existing data. When data come from different sources and are collected in different time periods and different locations, are the data still useful and valid for an analytical project? To be able to use existing data, what kind of additional data collection is needed to adjust existing data? Detailed weather data may be available for the specific geographic locations (e.g., nearby airports); to be able to use this detailed weather data to describe conditions away from these locations on the surface transportation system, the analyst may choose to test weather conditions mid-way between these locations can be reliably predicted by combining airport observations by collecting a small sample of relevant weather and road surface data.

Not all inconsistency issues can be overcome easily through supplementary estimation by assessing congestion impacts due to incidents. Figure 26 provides a one-day travel time profile and an incident identified from the incident database within the same time horizon. The travel time profile shows a typical pattern through 7:15 AM with a dramatic increase around 8:30 AM due to an incident blocking a travel lane. This relationship between incident location and travel-time dynamics cannot be identified when using the traffic and incident data from different years—only contemporaneous data reveals the relationships.

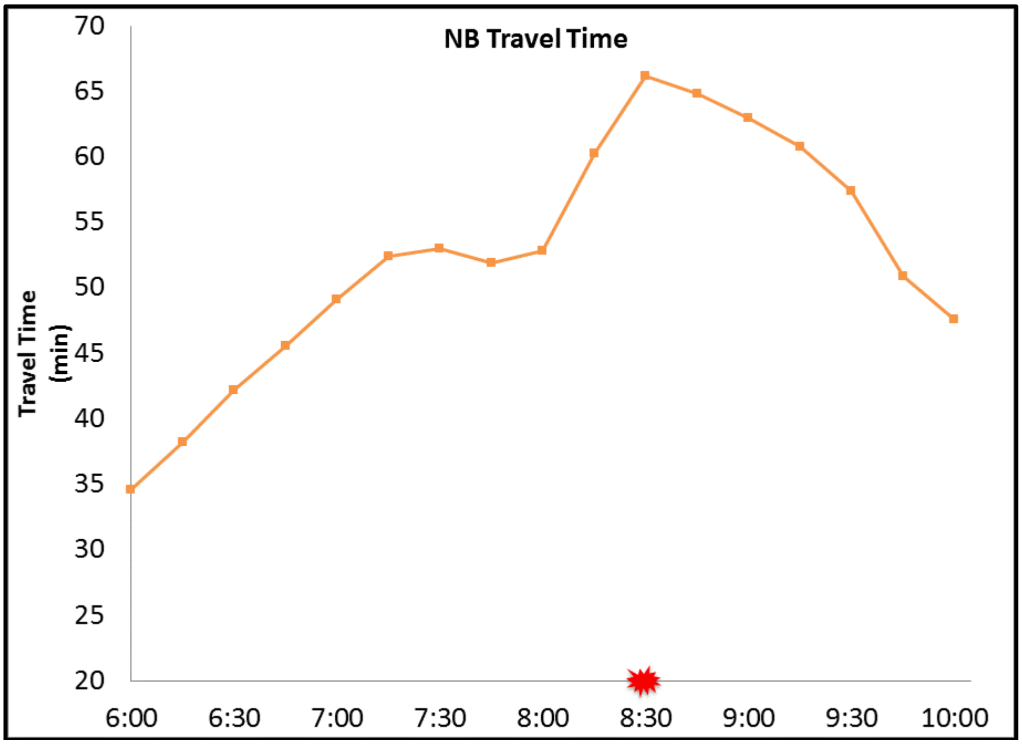

Figure 26. Chart. One-day travel time profile with an associated incident.
(Source: Federal Highway Administration.)

Emerging Trends in Operational Data

Traditionally, vehicle count is available through either a traffic count survey (manual or tube count) or traffic detectors (loop or mounted detector). Travel time is estimated by speed data from detectors or by recording vehicle trajectory from survey vehicles. With the 21st Century innovation technology, different types of datasets are available to provide more detailed and potentially more accurate information, such as probe data and connected vehicle data. Although the market penetration rate is not high enough to replace traditional data, this subset of data can be used for data verification or adjustment. Another set of data is crowd-sourced data from private sectors or individuals. Before using this set of data, the analyst must verify accuracy and legitimacy. This data set could be used as a supplement from a trusted source. One of the great opportunities for the 21st Century data analyst is to incorporate these emerging datasets into the analytic project development process. Some examples of emerging data sources are listed in the following subsections.

First-Generation Probe Data

Probe data are obtained by wireless communications with Global Positioning System (GPS)-equipped vehicles or mobile devices moving in the transportation system and post-processed to characterize current and historical patterns of congestion. These first generation systems primarily leverage vehicle or device position and current speed from many participants, fused with historical data and other sources to create comprehensive travel time and congestion data products.

Probe data providers. Some data providers collect trillions of bytes of information about vehicles on the roads from real-time anonymous mobile phones, connected cars, trucks, delivery vans, and other fleet vehicles equipped with GPS locator devices. The data collected is processed in real-time to provide historical, real-time traffic information, traffic forecasts, travel times, travel-time polygons, and traffic count to businesses and individuals. The capability of first-generation probe data product providers continues to grow in offerings and product detail.

Probe data technology solutions. Different from generalized data providers, probe data technology solutions facilitate targeted collection of vehicle position and location data on specific routes passively detecting and re-identifying vehicles moving in the transportation system. Several technologies can be considered; traditional License Plate Recognition (LPR) is a technology used to count vehicles and estimate arterial travel time. Using existing closed-circuit television (CCTV), road-rule enforcement cameras, or ones specifically designed for the task, this approach applies optical character recognition on images to read vehicle registration plates. The application includes police forces, electronic toll collection on pay-per-use roads and cataloging the movements of traffic or individuals, such as path travel time and the origin-destination (O-D) matrix. Similarly, toll tag technologies and Bluetooth reader technologies can be used to match vehicles in one location that appear later in other parts of the network.

These probe data are a readily available resource to the 21st Century data analyst and have some key features of interest to fill gaps or support travel time analyses, including continuous coverage over time, broad geographic coverage (beyond the facilities covered by fixed sensor deployments), and the ability to characterize travel times in multi-modal trip making.

Connected Vehicle Data

Efforts are under way to systematically augment position and speed data with other information. The U.S. Department of Transportation (USDOT), the Road Bureau of Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) of Japan, and the European Union's European Commission Directorate General for Communications Networks, Content & Technology (DG CONNECT) established a Trilateral Probe Data Working Group to coordinate research efforts on the three high-priority applications of connected vehicle data that were selected for joint study: Traffic Management Measures Estimation, Dynamic Speed Harmonization, and Operational Maintenance Decision Support Systems.[18] The Trilateral Probe Data Working Group defined probe data as data generated by vehicles about their current position, motion, and time stamp. Probe data includes additional data elements provided by vehicles that have added intelligence to detect traction information, brake status, hard braking, flat tire, activation of emergency lights, anti-lock brake status, air-bag deployment status, windshield wiper status, and so forth. Vehicle

[18] Assessment Report of US-Japan-Europe Collaborative Research on Probe Data, International Probe Data Work Group Phase 2, http://ntl.bts.gov/lib/60000/60100/60131/FHWA-JPO-16-356.pdf.

probe data may be generated by devices integrated with the vehicles' computers or nomadic devices brought into the vehicles.

Integrated Mobile Observation (IMO) project is sponsored by the USDOT Road Weather Management Program (RWMP) to demonstrate how weather, road condition, and related vehicle data can be collected, transmitted, processed, and used for decision-making. Data are collected from both vehicles and external sensors, including atmospheric pressure, steering angle, anti-lock braking system, brake status, stability control system, traction control status, differential wheel speed, and emission data. Based on a partnership between USDOT and state DOTs, Figure 27 provides an example of probe data collected by snow plow trucks.

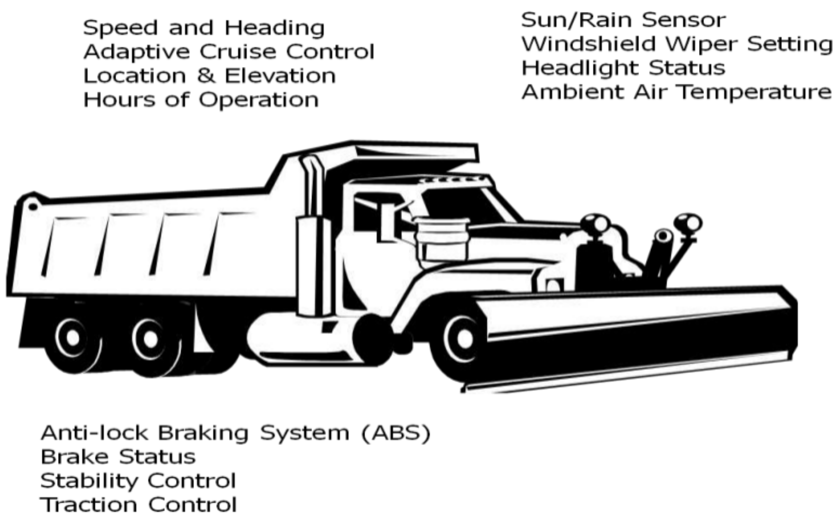

Figure 27. Illustration. Example probe data from snow plow trucks.
(Source: U.S. Department of Transportation.)

The USDOT initiated the connected vehicle research program to explore the potentially transformative capabilities of wireless technologies to make surface transportation safer, smarter, and greener and to enhance livability for Americans. The Society of Automotive Engineers (SAE) standard J2735, Dedicated Short Range Communications (DSRC) Message Set Dictionary defined the message sets, data frames, and data elements to produce interoperable DSRC applications. The message sets include a la carte message, basic safety message, emergency vehicle alert message, generic transfer message, a probe vehicle data message, and a common safety request message.[19] For example, connected vehicle safety applications rely on Basic Safety Message (BSM), which provides basic vehicle information, such as vehicle size, position, speed, heading acceleration, and brake system status. The vehicles equipped with connected vehicle onboard unit (OBU) will broadcast BSM. Figure 28 provides a graphical illustration of fully connected vehicle environment and the elements of vehicle data.

[19] SAE J2735—Dedicated Short Range Communications (DSRC) Message Set Dictionary, https://www.standards.its.dot.gov/Factsheets/Factsheet/71.

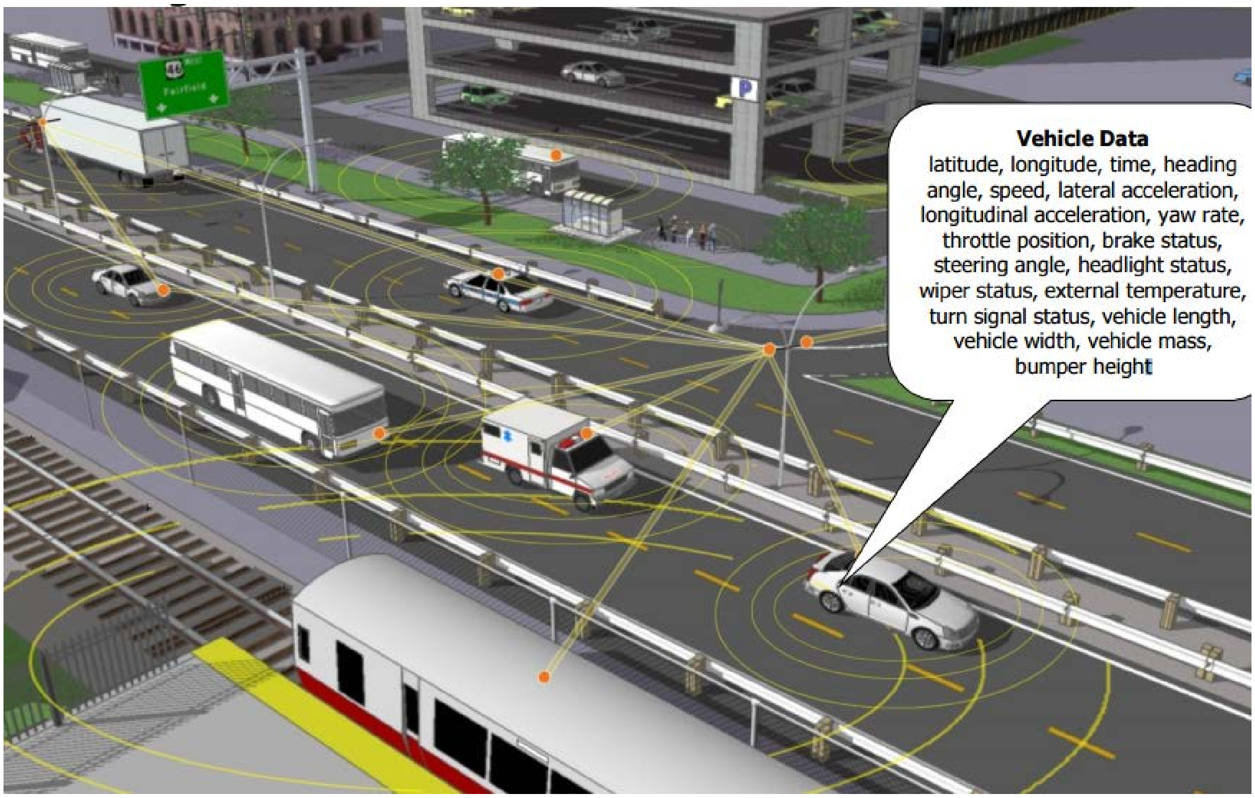

Figure 28. Illustration. A fully connected vehicle environment.
(Source: U.S. Department of Transportation.)

Data providers have begun to combine vehicle sensor data with other sources to provide new data feeds augmented with connected vehicle data. Vehicle temperature sensor and traction control data can be combined with traditional atmospheric weather information to give drivers advance warning of dangerous weather-related road conditions, keeping them safer on their route.

Crowd-Sourced Data

Crowdsourcing refers to "the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers."[20] One mobile navigation application relies on multiple forms of voluntary user input—crowdsourced data—to generate real-time traffic alerts, route suggestions, and estimated times of arrival. The USDOT Talking Technology and Transportation (T3) webinar provides some case studies of using crowdsourced data to enhance Transportation Management Center (TMC) operations.[21]

The nature and capability of crowdsourced data continues to develop. USDOT has sponsored the Enable Advanced Traveler Information System (ATIS 2.0) project to develop a smart phone

[20] Crowdsourcing definition, http://www.merriam-webster.com/dictionary/crowdsourcing.
[21] T3 Webinar, https://www.pcb.its.dot.gov/t3/s150311/s150311_crowdsourced_data_intro.asp.

application (shown in Figure 29) to collect traveler itinerary data and decision data, which can help refine near-term travel demand data. However, because these data rely on many individuals for content, one major concern of crowd-source data is the data quality and reliability. Before using it or committing to buy the data, a data analyst needs to do a quality check. Depending on the quality of the data, this type of data may only be used as supplemental data.

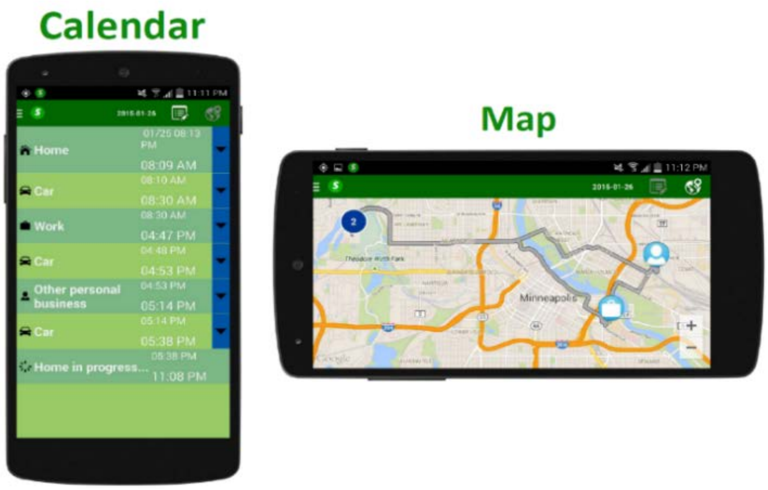

Figure 29. Screenshot. Daily detail views for displaying predicted daily activities and trips from cell phone.
(Source: SmarTrAC/University of Minnesota, 2015.)

MAKING DATA ANALYTICS-READY

In order to make the data ready and useful for conducting transportation analyses (Module 4), this section describes key components to consider when integrating data from multiple sources and controlling data quality. In some cases, if the needed data is not available, a data collection plan is prepared to describe the gap between available data and data needs, data resolution, how the data will be collected, and the final data format. Estimated cost is also included in the data collection plan. The analyst considers this process to be automatic if this is a recurrent effort.

Data Integration from Multiple Sources

A data analyst needs to ensure that each data set refers to the same clock time when combining data sets. For example, a TMC has a system that collects data from roadway sensors automatically through wireless or fiber. Due to the delay of transferring data from sensors to the database or the collection frequency, a data analyst might need to reconcile the clock time so the combination of, for instance, traffic volume and signal timing information makes sense. Similar to temporal issues, geographically, a data analyst needs to integrate data sets in a proper way. One of the steps in developing an analysis plan in Traffic Analysis Toolbox Volume XIII:

Integrated Corridor Management (ICM) Analysis, Modeling, and Simulation Guide[22] is to ensure data from multiple sources must also be for concurrent periods in order to neutralize seasonal and other travel pattern variances that can affect data. For example, data representing traffic conditions on the freeway during summer should not be compared with transit operating data collected during another time of the year.

Cross validation is a way to ensure the proper temporal and geographical integration of different data sets. When integrating data, a data analyst determines if the volume matches speed at the same time, if the left turn vehicles come from the most left two lanes or if the volume or speed data reflects the impact of an incident occurred at the same time on the same location. During a field test of Multi-modal Intelligent Traffic Signal Systems (MMITSS) applications, University of Arizona found that MAP distortion caused the vehicle to send a "cancel priority request" to the system. The issue was solved by including other information to correct vehicle position.[23]

When preparing data-related procurements, storage, licensing, and ownership issues are also critical. The analyst needs to figure out who owns the data and if the data can be used or manipulated. If an agency owns the data, does the agency have the right to release the license to the third party to conduct a procured project? If a database contains data from multiple sources, a data analyst needs to clarify the licensing and the ownership and how the data is stored and who has access to the database.

Metadata is a set of data that describes and gives information about other data. Having a common metadata framework across all the systems and using common controlled vocabularies are the keys to ensure the consistency and reliability of metadata applied to the information and data assets. For example, the USDOT Data Capture Management (DCM) Program developed a Research Data Exchange (RDE) platform to share archived and real-time data from multiple sources and multiple modes to better support the needs of Intelligent Transportation Systems (ITS) researchers and developers while reducing costs and encouraging innovation. The USDOT published Metadata Guidelines for the RDE to be adopted by public- and private-sector data providers to increase usability of their data.[24] Creation of metadata should be included in plans for the procurement of any data collection effort; otherwise, there is a risk that the data will be misinterpreted or abandoned as too arcane to support future analyses. A history of detector numbering should be included in the metadata so the analyst can link data sets from different years.

Quality Control and Missing Data Imputation

When conducting quality control of data or integrated data, a data analyst should avoid open-ended data quality control procedures and try to focus on the types of errors that are most likely to impact the results of the specific analytic project. The key notion is that there are many

---

[22] USDOT, ITS JPO, Traffic Analysis Toolbox Volume XIII: Integrated Corridor Management Analysis, Modeling, and Simulation Guide, May 5, 2012, http://ntl.bts.gov/lib/50000/50600/50615/30B00211.pdf.
[23] USDOT, ITS JPO, Multi-Modal Intelligent Traffic Signal Systems (MMITSS) Impacts Assessment, FHWA-JPO-15-238, August 19, 2015.
[24] USDOT, FHWA, Metadata Guidelines for the Research Data Exchange, 2012.

factors; since it isn't possible to control everything, the analyst focuses on controlling the factors that are important. In some cases, a data analyst needs to preserve outlier data in order to capture the time-variant traffic patterns (e.g., for a reliability analysis). Furthermore, certain types of analysis are more tolerant with respect to the error in the data, such as cluster analysis. The analyst needs to find a balance between setting quality control thresholds and preserving outliers while working around errors. An outlier is defined as an observation that lies an abnormal distance from other values in a random sample from a population. A variety of statistical tests are available to the analyst to identify and classify outlier data (including Scatter Plot, Box Plot, and Grubbs' Test).[25] A specific outlier may be the result of some sensor or processing error—or it could be an accurate reflection of variability in conditions. Before considering the possible elimination of these points from the data, the analyst should try to understand why they appeared and whether it is likely similar values will continue to appear. A key goal of the 21st Century data analyst is to preserve outliers not attributable to sensor and processing errors so that the full range of conditions can be characterized.

Quality control is usually set as an automatic process that looks at individual elements of the data. For example, the speed should be greater than 0 and less than 99 mph. However, some problems in the data that are not revealed by an elemental level of quality control might be related to relationships among data (e.g., temporal and geographical inconsistencies). These inconsistencies (e.g., widely inconsistent input/output counts for adjacent traffic count sensors) can be problematic, even if the data has passed multiple elemental-level checks. In our inconsistent count example, the analyst cannot calibrate a time-dynamic model of the system with this set of data. If vehicles enter a tunnel, they have to exit the tunnel. They cannot just disappear in the tunnel. To effectively use analysts-in-the-loop is the key to successfully developing a need-driven targeted quality control process. If the data is used for calibration purposes, the quality control focus is on the important geographic and temporal components. Critical geographic components include the location of bottlenecks, and temporal components include the time of day when the congestion states transition in/around these bottlenecks.

Once data is identified as missing during quality control, the analyst must decide whether to discard/disregard or impute missing data. Imputing data can be practical and realistic if the imputed data does not redefine or dominate the overall traffic pattern. Sometimes, the imputed data ends up introducing problems for calibration because the imputed data may introduce illogical relationships (e.g., unequal directional count data at a tunnel entrance and exit). When this happens, the modeled system cannot be calibrated. Another typical issue relates to the flagging of imputed data when passed from one step in the process to the next; one person assembles and imputes missing data, then hands off the data set to another person without pointing out which data has been imputed, causing calibration trouble because the comprehensive data set does not make sense anymore. Smoothing data (e.g., averaging values of certain time intervals) to minimize the impact of missing data is a common way to impute missing data. However, when the data are averaged several times during the integration and imputation process, the variation of data then is not significant. The analyst must avoid averaging data too much to capture the traffic patterns and impacts. Figure 30 through Figure 32 are

---

[25] Outlier definition, http://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm.

detector data from the Seattle I-405 corridor. Figure 30 provides an example with speed data on Lane 1 (left table) and Lane 2 (right table). On January 3rd (the green circle), Lane 2 has more missing values than Lane 1. Without doing quality control, the analyst may choose to average the two values, resulting in inaccurate speeds at that roadway segment, causing trouble for later calibration. Cross validation and quality control are foundational elements to ensure data quality. Once missing data is identified, the analyst must determine whether to impute the missing data.

Figure 31 shows an example of volume data at a detector station that all the values are missing and Figure 32 gives an example of volume data at another detector station where most values are available and a few are missing. The dataset in Figure 32 is the candidate for a data analyst to perform data imputation. Figure 33 illustrates an example of a travel time versus bottleneck duration graphic on the I-405 corridor. In this figure, the data with errors, such as zero travel time, are removed but the outliers, such as longer bottleneck duration are still preserved in the data set.

Element	**405es00171:_MN__T1**				
Data Content	Speed				
	1/3/2012	1/4/2012	1/5/2012	1/10/2012	1/11/2012
	Tue	Wed	Thu	Tue	Wed
6:30:00	56	51	57	53	57
6:45:00	56	0	57	57	0
7:00:00	56	0			
7:15:00	55	0			
7:30:00	56	55			
7:45:00	55	56			
8:00:00	56	55			
8:15:00	55	56			
8:30:00	56	58			
8:45:00	0	56			
9:00:00	56	56			
9:15:00	0	54			
9:30:00	53	54			

Element	**405es00171:_MN__T2**				
Data Content	Speed				
	1/3/2012	1/4/2012	1/5/2012	1/10/2012	1/11/2012
	Tue	Wed	Thu	Tue	Wed
6:30:00	62	60	60	58	58
6:45:00	61	60	59	63	63
7:00:00	0	0	60	61	0
7:15:00	58	0	60	58	0
7:30:00	62	66	60	0	58
7:45:00	0	60	61	0	0
8:00:00	63	0	62	0	0
8:15:00	61	0	61	0	60
8:30:00	0	0	59	0	64
8:45:00	0	0	62	0	60
9:00:00	64	60	0	0	0
9:15:00	0	0	61	0	0
9:30:00	0	63	64	0	0

Figure 30. Screenshot. Speed data on two lanes.
(Source: Federal Highway Administration.)

Element	005es15465:_MS_O_1						
Data Content	Volume						
	1/3/2012	1/4/2012	1/5/2012	1/10/2012	1/11/2012	1/12/2012	1/17/2012
	Tue	Wed	Thu	Tue	Wed	Thu	Tue
6:30:00	0	0	0	0	0	0	0
6:45:00	0	0	0	0	0	0	0
7:00:00	0	0	0	0	0	0	0
7:15:00	0	0	0	0	0	0	0
7:30:00	0	0	0	0	0	0	0
7:45:00	0	0	0	0	0	0	0
8:00:00	0	0	0	0	0	0	0
8:15:00	0	0	0	0	0	0	0
8:30:00	0	0	0	0	0	0	0
8:45:00	0	0	0	0	0	0	0
9:00:00	0	0	0	0	0	0	0
9:15:00	0	0	0	0	0	0	0
9:30:00	0	0	0	0	0	0	0

Figure 31. Screenshot. All values are missing.
(Source: Federal Highway Administration.)

Element	405es00171:_MNH__4						
Data Content	Volume						
	1/3/2012	1/4/2012	1/12/2012	1/17/2012	1/18/2012	1/19/2012	1/24/2012
	Tue	Wed	Thu	Tue	Wed	Thu	Tue
6:30:00	122	94	78	41	4	1	75
6:45:00	126	81	78	38	2	2	82
7:00:00	109	85	72	50	3	0	78
7:15:00	127	99	91	53	4	1	93
7:30:00	111	78	92	70	0	1	105
7:45:00	93	73	95	52	0	1	0
8:00:00	114	77	86	57	1	0	0
8:15:00	112	80	82	58	0	0	0
8:30:00	99	80	76	58	1	0	0
8:45:00	90	84	60	44	2	0	0
9:00:00	68	61	68	46	0	0	123
9:15:00	84	62	58	32	0	0	111
9:30:00	75	61	63	64	0	1	118

Figure 32. Screenshot. Few values are missing.
(Source: Federal Highway Administration.)

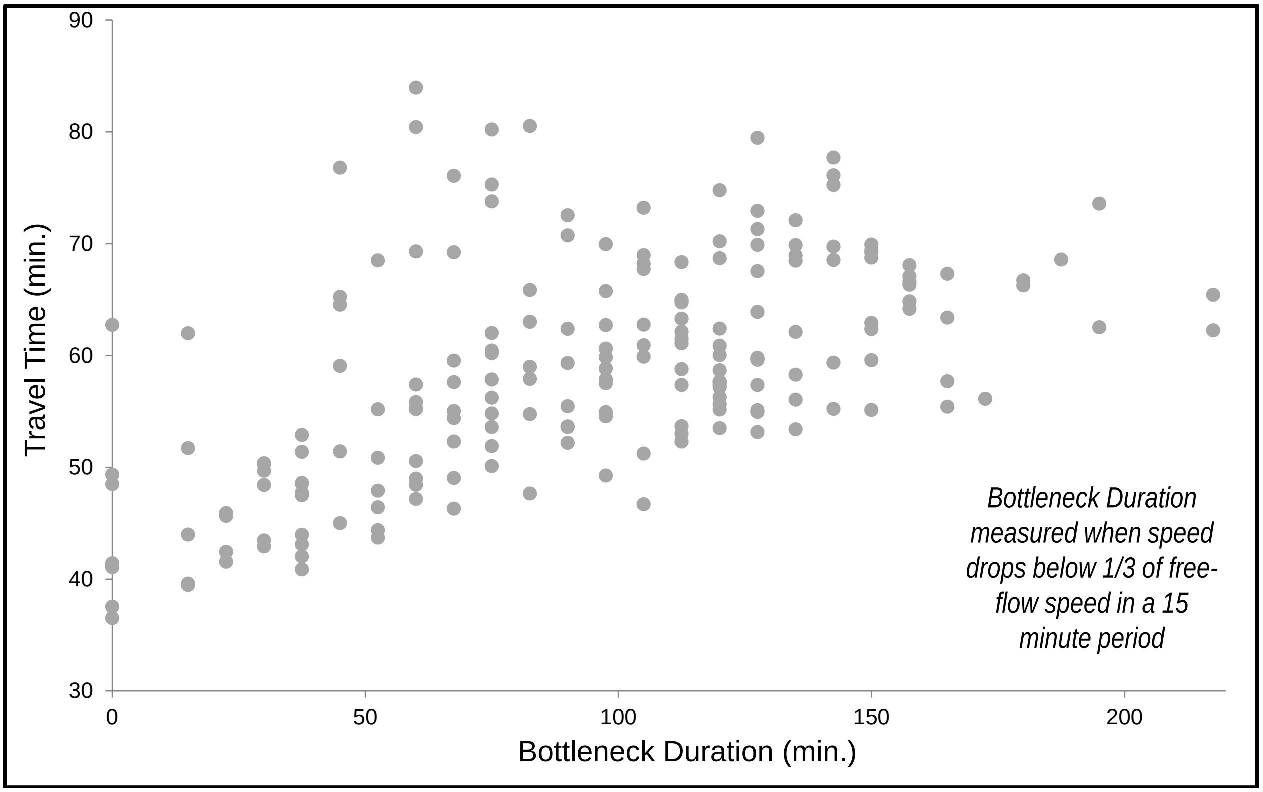

Figure 33. Chart. Natural Variation in Transportation System.
(Source: Federal Highway Administration.)

OPERATIONAL CONDITIONS

The availability of more continuous data improves the analyst's ability to characterize dynamic system performance. Analyses can be within a day (e.g., the rise and fall of travel times in a peak period) or over many days (e.g., the variation in travel times between a specific origin and destination departing at a specific time each day over a full year). Likewise, there is a corresponding pressure to assess more and more complex alternatives—intended to improve system performance—that are highly conditionally dependent (e.g., incident management and traveler information systems, road-weather technologies, and congestion pricing methods).

The focus of this guide is on the characterization of time-dynamic system performance and the use of data-driven analytics to improve performance of transportation systems. In the past, when system data were scarce and potentially rife with errors, analysts tended to fall back on trying to describe a single, nominal "normal" operational condition derived by taking the average of many different attributes. In a 21st Century context, where data are more broadly available and contain far fewer errors, such an approach is obsolete. The 21st Century analyst instead uses a data-driven method to identify multiple distinct operational conditions to better characterize transportation system dynamics. This set of operational conditions is a more effective and useful basis for comparing alternatives and is a foundational element of any analytical effort aimed at improving system performance. Figure 34 illustrates one of the key challenges for analytical projects—to fully leverage and use available data sources in the design and execution of meaningful analyses that properly represent and test the competing investment alternatives.

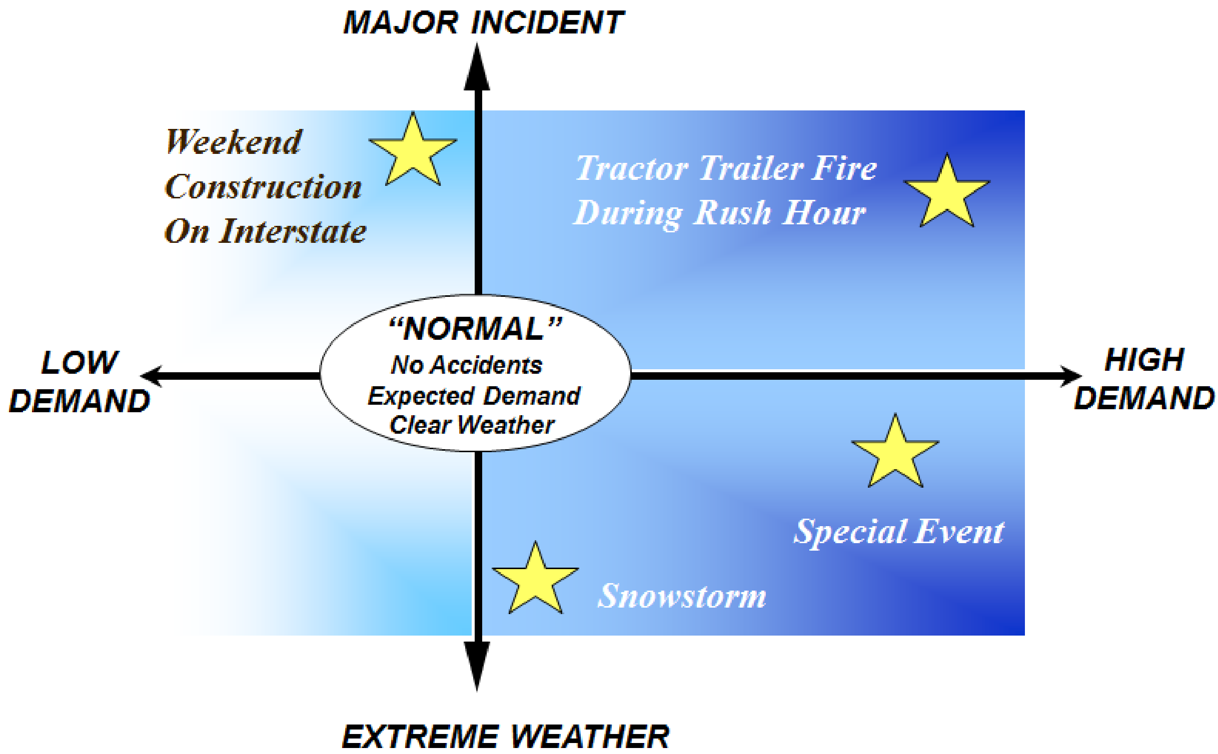

Figure 34. Diagram. Various operational conditions.
(Source: Federal Highway Administration.)

A simulation tool, for example, is intended to predict or represent conditions under very specific conditions—conditions with well-defined travel demand, weather, and incident patterns. If input representing the average of many days is created for a simulation model, one likely outcome is the creation of a strange hybrid "normal" condition: multiple muted bottlenecks, unnaturally smooth travel times and speed profiles, with partial incidents and vague weather conditions (e.g., neither rainy nor clear). Such a day cannot serve as a useful differentiator of complex alternatives nor can it alone reveal anything regarding day-to-day travel-time reliability. To support the analyses of complex, condition-dependent alternatives or to conduct a reliability study using a simulation tool, a systematic data analysis to identify a practical set of representative operational conditions is required.

Cluster Analysis. Cluster analysis is well-known and relatively simple statistical method that can be used to capture a variety of operation conditions that consider traffic, incident and weather impacts.[26] Cluster analysis has been used in numerous research efforts over the last

[26] FHWA, Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software, FHWA-HOP-16-070, 2016.

20 years to identify various traffic patterns and characterize operational conditions.[27,28] Other statistical techniques with similar goals (e.g., set partitioning methods) can also be employed to derive a set of operational conditions from underlying data.[29]

To conduct the underlying analyses to identify operational conditions, the analyst requires sufficient data. At a minimum, this includes time-variant traffic data (count and either speed or travel time data, incident data, and weather data). These data must be contemporaneous—all of the data must be from the same time period. The data required to characterize operational conditions is essentially the same data needed to create system performance profiles (Module 1). A minimum of 30 days of contemporaneous data is required to perform the simplest short-term analytic effort, generally associated with near-term operational analyses looking forward a few months. To compare alternatives expected to be in place for a year or more, an annual analysis based on as many days as possible (e.g., non-contemporaneous data removed) uniformly drawn from across a full calendar year is recommended. For analysts with good supporting data, it may be more practical to include all days from a calendar year or multiple years to characterize conditions rather than to randomly (or arbitrarily) reduce the sample set of data. The risk in random, arbitrary, or even systematic data reduction prior to analysis is that interesting outlier condition days (major incidents, bad weather, and special event days) may be underreported or unreported in a resulting analysis of these data.

Unit of Observation. One typical stumbling block in preparing data for analysis is a failure to select an appropriate unit of observation for operational conditions analysis. Operational conditions are intended to describe the holistic state and performance of the transportation system in periods of intense dynamic change. For example, consider an analysis of alternatives to reduce peak period congestion on weekdays. Assume the data analyst has a contemporaneous set of data covering 200 weekdays over a calendar year. If the transportation system experiences two peak periods per day, each lasting roughly 6 hours with (relatively) uncongested conditions in the midday and overnight hours, choosing two peak periods (AM and PM) for separate analysis is recommended. The unit of observation for this analysis is based on an individual day, broken into the two relevant AM and PM portions. The analyst conducts one analysis on 200 AM peak periods to derive a practical set of representative AM peak operational conditions. The analyst conducts a second analysis on the 200 PM peak periods to derive a practical set of PM peak operational conditions.

Operational conditions cannot be identified if the component elements (demand, incidents, and weather) are analyzed independently. First, the three component elements are not independent in reality. Poor weather suppresses travel demand and is correlated with higher incident rates and patterns. Second, such an independent analysis must eventually be merged together in order to conduct analyses—because analytic methods, particularly simulations, require the representation

[27] Chen, Y., J. Kim, and H. S. Mahmassani. Pattern Recognition Using Clustering Algorithm for Scenario Definition in Traffic Simulation-Based Decision Support Systems. In Transportation Research Board 93rd Annual Meeting, 2014.

[28] Azimi, M. and Y. Zhang. Categorizing freeway flow conditions by using clustering methods. Transportation Research Record, vol. 2173, pp.105 -114, 2010.

[29] Buxi G. and M. Hansen. Generating day-of-operation probabilistic capacity scenarios from weather forecasts. Transportation Research Part C: Emerging Technologies, vol. 33, pp.153 -166, 2013.

of a specific day, not a combination of days. The 21st Century analyst allows the dependencies among attributes of operational conditions to emerge from the data analysis.

Selecting attributes. Individual detector speed and count data, or single incident reports, or weather station temperature readings by themselves in raw form are not suitable as attributes to attach to the peak period unit of observation for cluster analysis. Each peak period should be characterized using a set of normalized attributes (see Traffic Analysis Toolbox Volume III) that describe the nature of the travel demand, incident number, intensity and pattern, and weather conditions.

Travel time and bottleneck throughput attributes. Two of the more typical attributes used in the characterization of operational conditions are travel times and bottleneck throughput rates. In each case, we are interested in the dynamics of these measures under congestion—how they rise and fall over time each day or peak period. The attribute of travel time for a route should reflect the travel time considering time of trip start from the origin and each intermediate node in the route, rather than an instantaneous addition of travel time from all links on the route at a specific time. This array of dynamic travel times can be used (when normalized) to characterize the timing and intensity of congestion on a route in the period. Note that several routes will likely be a part of a system characterization. Likewise, the flow rates at recurrent bottleneck locations is a critical determiner of system performance and the dynamics at these locations are critical for both calibration and system characterization. Bottleneck throughput tends to rise to some maximum, decline as the bottleneck becomes congested, then recover. The onset and dissipation times are keys to understanding the total system dynamics, as well as understanding analytical model calibration.

Enumerative or attribute stratification approaches. Enumerating or stratifying all the possible combinations of conditions is not recommended; such an approach is both impractical and unnecessary. A typical approach seen in the design of many analytical studies is to take each of the possible attributes of the individual day or peak period, characterize these attributes, and then create a large n-dimensional grid of all the possible combinations. This approach has several major weaknesses. First, the approach quickly becomes impractical. Consider an analyst who without a cluster analysis, arbitrarily defines four travel demand patterns (e.g., low, medium-low, medium-high, high), eight incident patterns, and five weather conditions (clear, light rain, rain, fog, snow/ice). This results in 4 x 8 x 5 = 160 potential operational conditions, each of which will require data to characterize and an accompanying analytical representation for calibration. This is an extensive and largely unnecessary expenditure of effort. Consider that a full year of weekday operations results in roughly 200 actual days. The stratification effort has reduced the complexity of the analysis by only 40 runs (or 20%) versus a pure enumerative approach. Even more telling is that if the 200 days were mapped into the 160 stratified conditions, the result would be that the majority of the cells in the grid would be empty, and there would be many singletons. So in an effort to reduce the analysis to something more practical than pure enumeration, the analyst can frequently find themselves creating a trap in which the level of effort is roughly the same as pure enumeration but actually a worse characterization of overall conditions (if all conditions are considered of equal weight).

Data-driven statistical methods. The overarching goal of these techniques is to find subsets of days (or peak periods) and to find a practical small set (for simulation studies, generally fewer than 20 operational conditions) of representative operational conditions. Further, a rule of thumb with respect to statistical methods like cluster analysis is that the resulting subgroups that characterize operational systems well lies somewhere around the square root of 50% of the number of days (or peak periods); for our 200-day analysis, this translates into Square Root (SQRT) of (50% x 200) = 10 most frequent conditions. This reduces the complexity of an annual analysis by 95% compared to a purely enumerative or stratifying approach and is more representative because the attributes characterizing each day are not wholly independent—in fact they are highly dependent. For example, snow/ice conditions are almost never associated with high demand. The days with high demand, specific incident patterns, and non-extreme weather tend to cluster together. Likewise lower demand days, with associated incident patterns and characterizing weather also group. One can continue this thought exercise from any of the three major attributes (e.g., extreme weather is associated with lower demand and so on). The result is the same—and the power of the statistical method is that the computer algorithms essentially conduct this kind of thought process from nearly every conceivable angle before suggesting specific subgroups (and related days).

Objective-focused operational conditions analyses. Note that operational conditions analysis should also be needs-driven. If the alternatives being compared only differ under specific weather conditions (e.g., icy/snow conditions) then clustering on a subset of days that have this attribute is a good way to focus project resources. One key is to use the full data set to characterize how frequent these conditions occur annually (e.g., these days represent 10% of all weekdays in the year) before using an analytic technique to characterize snowy/icy operational conditions in greater detail. A recent study[30] used this technique specifically for weather-related operational interventions.

Reliability analyses. Capturing various operational conditions can also help a data analyst conduct reliability analysis. Even without running cluster analysis, the analyst can still perform reliability analyses using observed data. Based on the system profile, a data analyst can identify operational conditions by simply filtering the data, such as low demand days without incidents, or the days with both incident and weather impacts. For example, Federal Highway Administration (FHWA) developed an approach to measure travel time reliability.[31] Transportation Research Board (TRB) Second Strategic Highway Research Program (SHRP2) also initiated several research related projects to incorporate reliability performance measures into the operation and planning process.[32]

Rare events. In some cases, an operational condition may only occur every few years or is unobserved, such as natural disasters (e.g., hurricane) or other special cases/events (e.g., Olympics). It is impossible to characterize the operational condition given the fact that data is not

[30] Chicago AMS Testbed report (to be completed).
[31] Travel Time Reliability Measures, http://ops.fhwa.dot.gov/perf_measurement/reliability_measures/.
[32] Strategic Highway Research Program 2 (SHRP 2) Reliability Research Reports, http://www.trb.org/Publications/PubsSHRP2ResearchReportsReliability.aspx.

available. A data analyst needs to figure out a way to best represent the condition, such as finding the maximum delay of the system, screening the data for the day that is close to it, or using data from other locations with the similar situation.

MOVING TO MODULE 4: OPERATIONAL CONDITIONS SUMMARY

In order to move forward towards a more complete analytical design, the analyst that assembles, assesses, and analyzes data to characterize operational conditions should capture the main elements of that effort in an Operational Conditions Summary. The simple template offered in Table 5 can be tailored for the specific type of analysis to be performed. Note that the Summary is not used to explain how the analysis was conducted; it shows the results of the effort that will impact the proposed analysis, namely the identified operational conditions.

Table 5. An operational conditions summary template.

Data Summary	**All**	**Op. Con. 1**	**Op. Con. 2**	**Op. Con. 3**	**Op. Con. 4**	**Op. Con. 5**	**Op. Con. 6**
Periods/Days	196	40 (20%)	25 (13%)	6 (3%)	41 (21%)	28 (14%)	56 (29%)
Operational Condition Characterization		Low Demand	Low Visibility	Weather + Incidents	Many Incidents	Bottleneck Trouble	Few Incidents
Representative Day		9/6/2014	7/18/2014	2/15/2014	8/19/2014	11/1/2014	9/15/2014
Attributes	**Avg.**	**Op. Con. 1**	**Op. Con. 2**	**Op. Con. 3**	**Op. Con. 4**	**Op. Con. 5**	**Op. Con. 6**
North Bound Bottleneck Duration (minutes)	74.46	21.0	71.4	55.0	69.1	128.0	93.2
South Bound Bottleneck Duration (minutes)	113.6	39.4	127.2	112.5	149.3	190.7	95.9
North Bound Maximal Travel Time (minutes)	54.9	48.8	57.0	69.2	58.7	57.5	52.6
South Bound Maximal Travel Time (minutes)	63.2	45.5	69.7	90.3	67.6	74.7	61.0
Number of Incidents (count)	1.64	1.63	1.60	2.67	2.98	1.21	0.79
Maximal Incident Duration (minutes)	22.8	27.7	21.1	62.3	28.5	20.0	13.2
Visibility (miles)	8.45	9.53	2.25	3.33	9.48	9.03	9.96

Note: Operational Condition is denoted as “Op. Con.”

The periods/days identifies the full set of days or peak periods used in the analysis (under the header “All”) and the number of days associated with each operational condition. In each of the operational conditions columns, the percent of annual occurrence or frequency is expressed as a percentage of the total number of days analyzed.

The operational conditions characterization describes the nature of the operational condition providing context for the analyst. Characterizing some operational conditions will be clear while the underlying root cause may be difficult to discern for others. In this sample, the condition associated with "bottleneck trouble" was a set of conditions under which one or more of the two recurrent bottlenecks had poor performance over the period, although the underlying cause was not evident from the data assembled. One conjecture is that for these days, local visibility (glare) issues or road surface conditions may have played a role in impeding bottleneck operations.

The representative day identifies a single day from the operational condition to be used within the analysis plan. This day is generally near the center of the cluster but should also have good time-dynamic data for the key performance and calibration measures.

The attributes table shows the set of attributes that the algorithmic approach used to differentiate the operational conditions. The value of each attribute is shown, as well as the aggregate annual average for comparison.

Summarizing the operational conditions is a critical resource in order to complete the analysis plan. The number of operational conditions is critical to understanding the calibration and analytical requirements for the overall effort. Identifying the representative day in each operational condition is critical to providing the detailed time-variant data required for calibration. Note that the summary alone is not enough to conduct the analysis. Detailed time-variant data for each of the key performance measures must be organized and made available for calibration for each of the identified representative days.

MODULE 4. CONDUCTING AND DOCUMENTING TRANSPORTATION ANALYSES

Following the data preparation guidelines presented in Module 3, Module 4 provides guidance on how to conduct and document a transportation analysis project. Figure 35 describes the contents of this Analysis Module 4, including developing the Analysis Plan containing the experimental design for the analysis, calibrating and validating the analysis tools, conducting sensitivity analysis, and producing and presenting the analysis results.

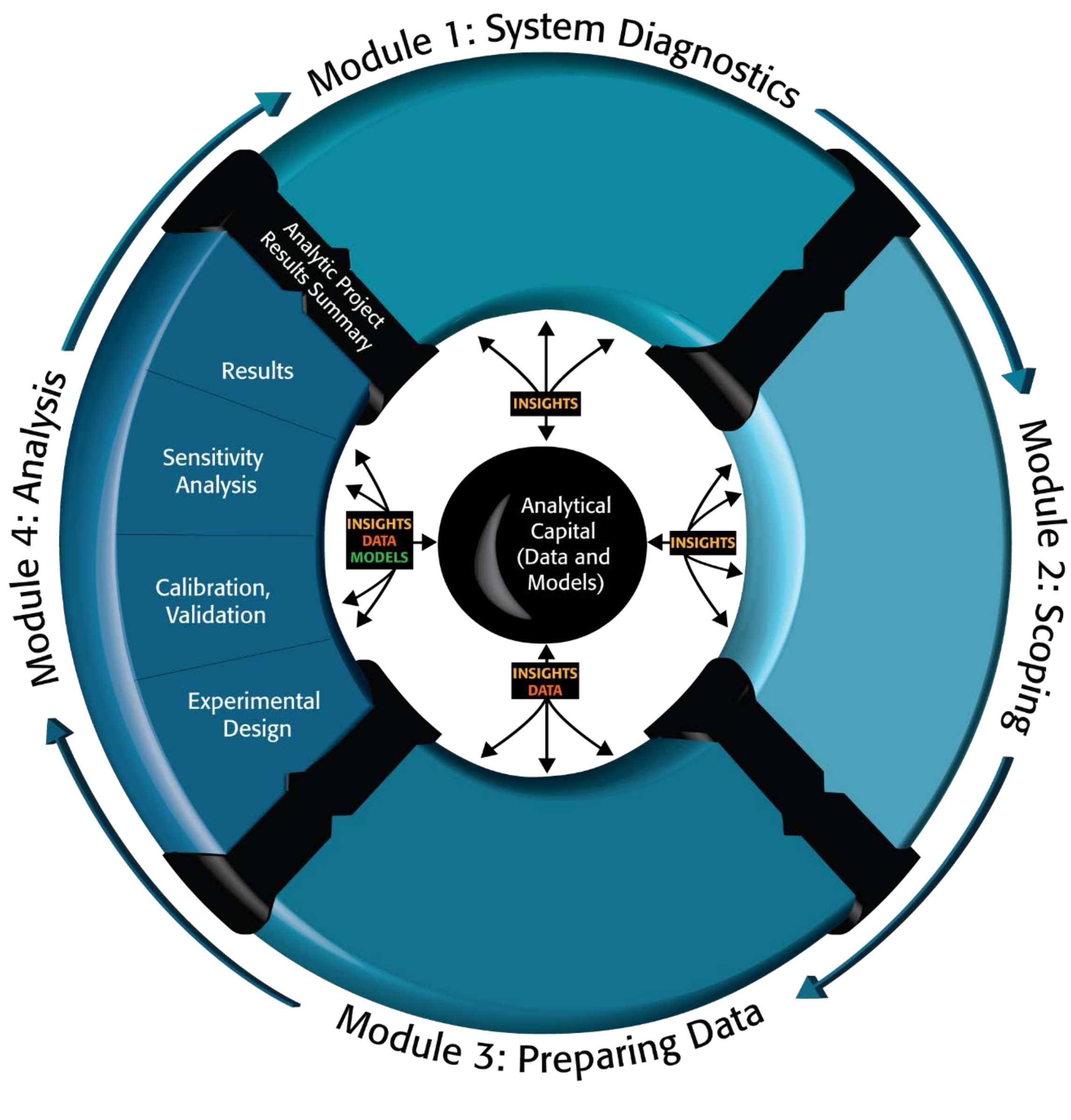

Figure 35. Diagram. Projects execution and documentation within the 21st Century analytic project scoping process.
(Source: Federal Highway Administration.)

THE ANALYSIS PLAN

The analysis plan is closely related to the scoping plan; it includes planning for the analysis work and then iterative updates to assumptions, scope, and agreements as the project moves forward. The analysis plan is more detailed than a scoping plan and must be adapted as the strengths and limitations of the data and tools organized for the analysis become better understood. The analysis plan can also help maintain a clear and mutual understanding among stakeholders of expectations and assumptions and can help identify potential flaws or technical issues in the project's expected operation.

The analysis plan confirms the analysis approach and methodology, as well as project objectives, the study area conditions, performance measures, the strategies being implemented, and the tools and data to be used in the analysis. The analysis plan needs to be sufficiently detailed to provide practical guidance on the actual conduct of analysis, yet it should also retain some flexibility to adapt to project contingencies as they are encountered.

Contents

Improving specificity on the contents of a scoping plan, an analysis plan would typically include the following interrelated components—the presentation order of these components may vary depending on project characteristics:

- **Project definition**. This component includes problem statement and project goals and objectives, existing operational conditions, problem diagnosis, affected stakeholders and project partners, analysis time horizon, analysis periods, study area and geographic scope for the analysis, and modes and facilities affected.

- **Geographic and temporal scope**. The spatial extent of the study area include intersections, highways, and other facilities to be analyzed. The study area must cover beyond the end of the full spatial extent of queues and congestion in the baseline and future years of analysis. The project analysis time period, (AM/PM/midday peak hour and/or peak period, off-peak period, etc.) must cover the beginning and end of full temporal extent of queues and congestion in the baseline and future years of analysis.

- **Selecting the appropriate analysis tool**. Analysis tool selection is based on an assessment of several key factors, including the geographic and temporal scope of the analysis, travel modes, proposed mitigation strategies, performance measures, scenarios and operational conditions of interest for the analysis, data to become available, and cost-effectiveness for the analysis tool(s) to be used in the analysis. If partners want to model pricing strategies, it is important to select a tool that can accommodate this. The analysis team may need to consider integrating the analysis capabilities from multiple tools to achieve the necessary capabilities.

- **Performance measures to be used in the analysis**. Performance measures should provide an understanding of travel conditions in the study area, including both localized and system-wide metrics and recurrent and non-recurrent congestion and impacts. These measures should be able to map back to the needs and issues the analyst is trying to solve (identified in Module 1).

Analysis tools need to be selected based on how they address required performance measures. Key performance measures include mobility (travel time, delay and throughput), reliability of travel time, transportation safety, emissions and fuel consumption, and costs to deploy, operate, and maintain the transportation improvements and replace equipment as they reach the end of their useful life.

- **Analysis data requirements**. These requirements for developing and calibrating a transportation analysis tool generally include the following types of data: roadway geometry, traffic control, travel demand and volume, performance (such as queue locations, queue lengths, travel times, and speeds); and vehicle characteristics, such as vehicle classifications or vehicle mix. Typical data challenges include comprehensiveness, reliability, and accuracy.

- **Analysis tool calibration criteria and expectations**. Guidance on the overall tool calibration process is presented in Volume III of the Traffic Analysis Tools suite. The suite update allows projects to be analyzed under different operational conditions with a better expected statistical fit than a “typical day” trying to represent a reliability space. These operational conditions include different incidents, work zones, weather events, and special events.

- **Alternatives to be studied, including analysis scenarios and transportation mitigation strategies**. Analysis scenarios should be developed for the range of operational conditions of greatest interest to the site in light of its analysis objectives. Based on the frequency and severity of conditions linked with elevated congestion levels, “operational clusters” can be defined and used in the analysis. Transportation mitigation strategies typically include geometric and operational alternatives to be analyzed and compared to the baselines. On the following page, Figure 36 shows a comparison of two alternatives using schematic drawings that represent different geometric configurations for the alternatives.

- **Expected cost, schedule and responsibilities for the analysis**. This section of the Analysis Plan specifies the responsibilities of different parties involved in the analysis (analysis team, project management team, client, partners, and stakeholders), presents a schedule with specific milestones for completion of analysis activities, and provides a cost estimate associated with different analysis tasks.

Some transportation agencies have developed formalized procedures on how to document analytical procedures; this documentation is sometimes referred to as the Methods and Assumptions document, which typically requires stakeholder concurrence before proceeding with the analysis. This document keeps track of all assumptions made as the analysis progresses. In the Methods and Assumptions document, the analyst keeps track of all of the decisions they made dealing with uncertainty, poor data quality, and other factors that may influence the analysis downstream. The outline for the Analysis Plan presented here is consistent with the contents of the Methods and Assumptions document.

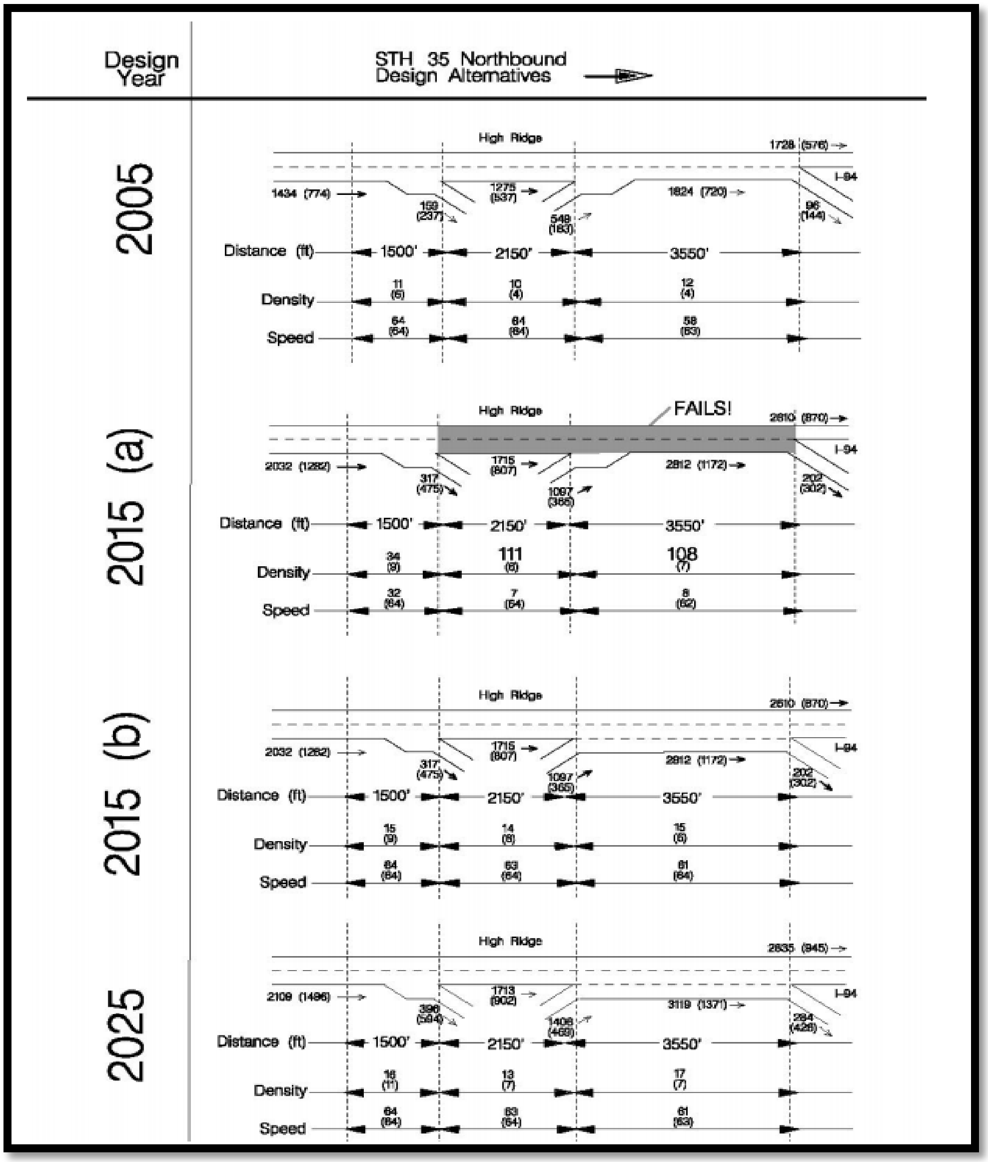

Figure 36. Chart. Sample comparison of project alternatives using schematic drawing. (Source: Minnesota Department of Transportation/Federal Highway Administration Traffic Analysis Toolbox Volume IV.)

Iterative Development Process

Just as the scoping plan (described in Module 2) was developed in an iterative fashion, the analysis plan requires adaptation and revision over the course of the effort. Adaptations can be related to insights gained from analyzing the underlying data, construction and testing of the analytical tools, and preliminary results obtained from the project. In some cases, initial estimates of potential impact can be examined to see if differences in system performance hypothesized among scenarios are realized when the analysis is conducted.

STAKEHOLDER INVOLVEMENT AND REVIEW

Stakeholders who participated in the development of candidate project concepts and scoping activity should also be engaged in the development and iterative adaptation of the analysis plan. All issues regarding the nature of system performance and the alternatives to be tested (described in Module 2) are still in play. The analysis plan is the document that keeps track of stakeholder insights and concerns, how they are addressed, and the impact on the overall analysis.

Most critically, the analysis plan should document the harmonization of conflicting or antithetical objectives and constraints as they relate to different performance measures. Overall, the analyst should strive to strike a balance between multiple performance measures to tell a potentially complex story about system performance and localized impacts—optimized in different ways for different operational conditions. As examples of this balancing process, Figure 37 and Figure 38 present evaluations of the performance of different market segments against their expected transit market share.

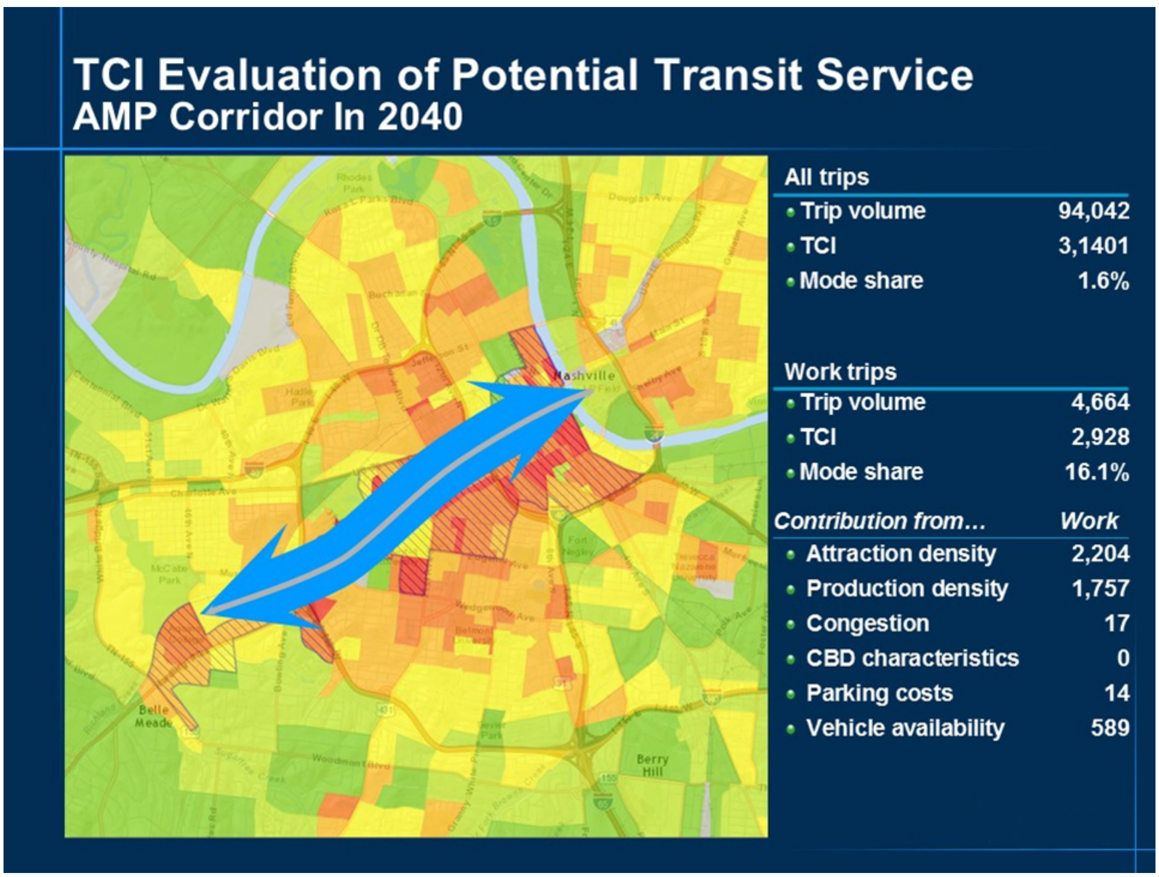

Figure 37. Snapshot. Evaluation of potential transit service.
(Source: Federal Highway Administration.)

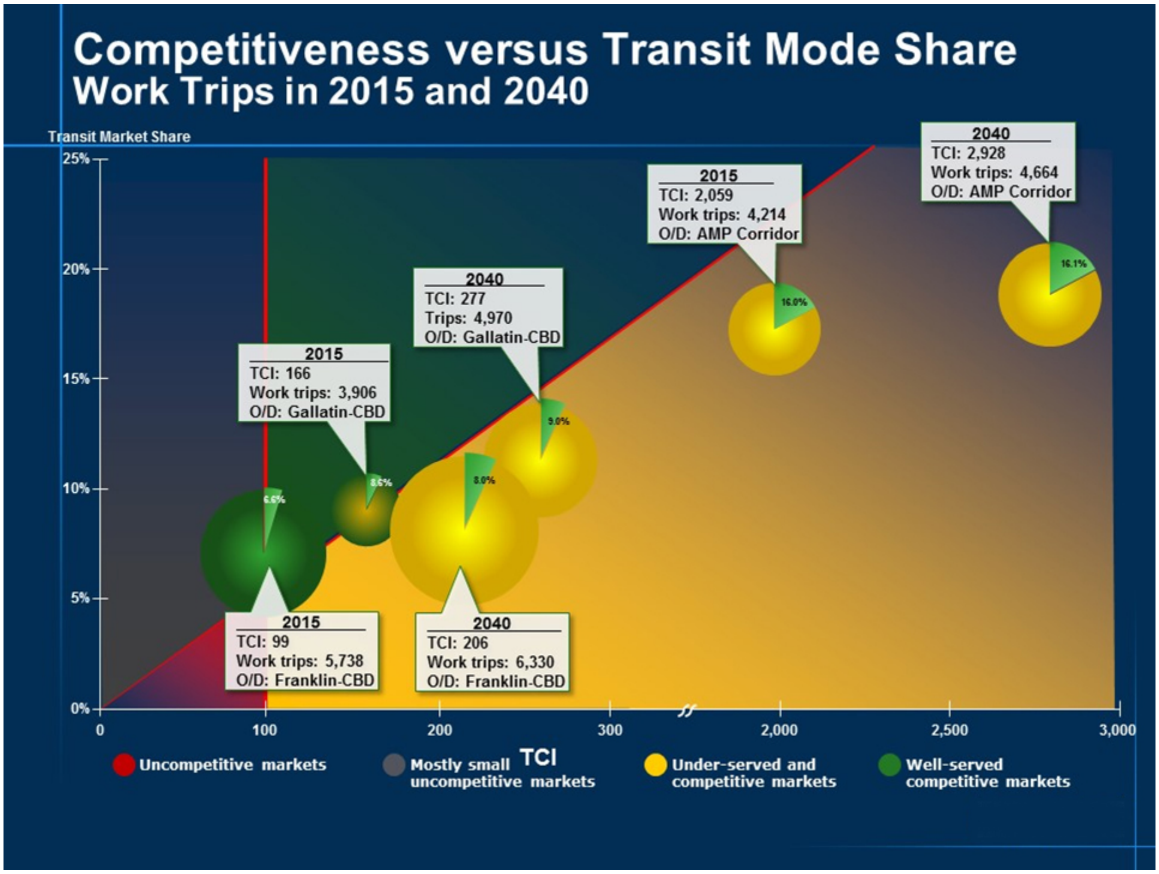

Figure 38. Snapshot. Summary comparison of transit competitiveness.
(Source: Federal Highway Administration.)

RE-EXAMINATION OF PROBLEM IDENTIFICATION AND DIAGNOSIS

The identified problem, hypotheses, and underlying diagnostics should be checked again before proceeding with detailed development of the analysis plan; time may have passed since the project was initially conceived. The analyst should reexamine the diagnostics performed when the project was initially created (see Module 1) and the assumptions about the project scope (see Module 2.)

EXPERIMENTAL DESIGN FOR ANALYSIS OF DIFFERENT OPERATIONAL CONDITIONS

Scenarios should be developed for the range of operational conditions of greatest interest to the site in light of its analysis objectives. For example, while traffic incidents are the single largest cause of non-recurrent congestion, stakeholders are encouraged to investigate and understand other influences, including special events, weather, fluctuations in demand, and work zones. This initial analysis also includes developing a preliminary understanding of both supply side (infrastructure/capacity) and demand-side influences on the study area across all modes (including underlying causes of demand, such as directionality or day of week), with a goal of identifying potential issues (where demand exceeds supply to an extent believed to interfere with study area performance) and opportunities (underutilized capacity/supply that could potentially help absorb demand).

Practitioners should compile data on the frequency and severity of conditions linked with elevated congestion levels to create the foundation for identifying "operational clusters" that characterize the operational conditions found in the study area and can later be used to organize the analysis. Comparisons or distributions of various sources of delay should be assembled and evaluated to identify the relative frequency of events/conditions related to congestion; this is often referred to as "reliability space." From this assessment, practitioners should critically assess the potential impact of various scenarios. Scenarios identified as having a low frequency of occurrence or likelihood and low expected impact should be assigned a low priority, since the impact of their inclusion—which should be weighted by their low likelihood of occurrence in the analysis—likely provide much less impact on the final analysis outcomes than scenarios with greater frequency of occurrence or higher expected impact. This problem diagnosis may reveal needs previously unknown to the practitioners, help to weed out inconsequential scenarios, and greatly assist in targeting resources to provide the greatest expected value from the analysis.

The data analysis required in this step includes identifying the frequency and likely impact of the scenario. Those scenarios recording the greatest frequency and the greatest impacts should be given the highest analysis priority. Scenarios with a low likelihood but major impact (e.g., major snowstorms) or scenarios with a frequent occurrence but limited impact (e.g., minor incidents occurring on otherwise normal days) should be provided slightly less priority. Scenarios with low frequency and low impacts should be considered for deletion from consideration.

EXPERIMENTAL AND CONTROL CASES (WITH AND WITHOUT)

Once the analysis scenarios have been identified, the analyst develops improvement strategies and determines under which analysis scenarios the strategies will be activated. It is critical to understand when (under what conditions) the strategies will be applied and how their application may vary under different conditions. To better understand what factors influence congestion and the frequency in which these factors occur, the analysis team and the stakeholders need to identify the combinations of travel demand, incidents, special events, and weather events that affect the study area operations. A transportation improvement project is likely concerned with nonrecurring congestion on a level equal to or greater than typical recurring congestion levels. Therefore, it is critical that the analysis team recognize the non-typical factors that affect nonrecurring congestion.

The analysis team explores preliminary analysis hypotheses and assumptions and identifies possible opportunities and constraints associated with the application of improvement strategies identified under specific operational conditions. Freeway managers may be interested in opportunities to divert drivers from the freeway to arterials as an incident management strategy. They engage local arterial managers to understand whether the local jurisdiction can accommodate this diversion and activate these strategies to accommodate the desired performance in a sufficiently timely manner. Discussions also take place about the potential to avoid problems—such as changing signal timing or Dynamic Message Signs (DMS) to avoid queues that may lead to collisions at critical locations. If there are contributing circumstances that can be avoided that lead to greater non-recurrent congestion, the team must devise strategies to minimize these circumstances. They also seek to understand and address any concerns or constraints the local jurisdiction may have: Is the technology in place to accommodate the needed signal timing? Can the timing strategies be activated in a timely fashion to make the strategy feasible?

Likewise, if local agencies are interested in exploring opportunities to divert freeway or arterial traffic to transit, they will engage in collaborative dialogue with the regional transit managers to understand possibilities for creating available transit capacity, including parking facilities, to accommodate the possible influx of demand under certain scenarios. Figure 39 and Figure 40 present example comparisons of expected changes in speeds and delays, respectively, in a transportation network that contains both freeway and arterial segments; such comparisons help stakeholders visually assess expected improvements and reductions in service in different parts of the network.

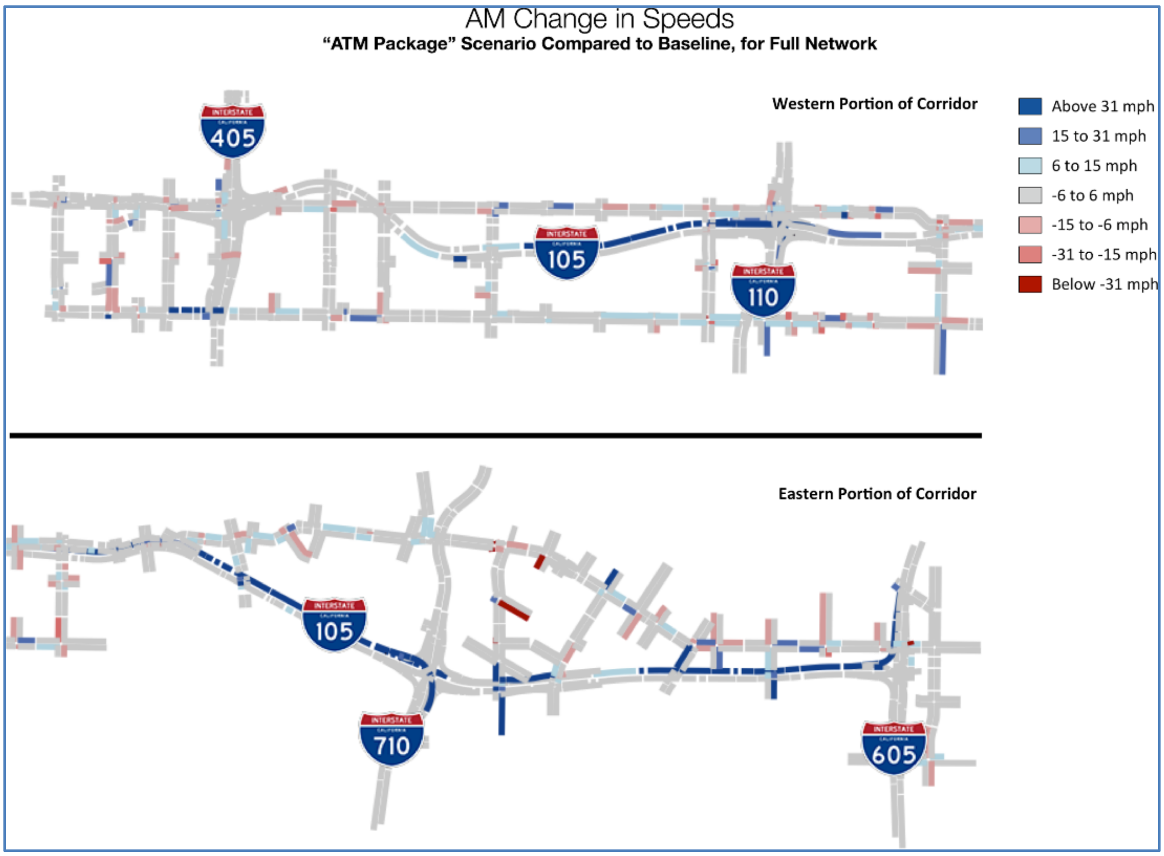

Figure 39. Map. Change in speeds comparison.
(Source: Federal Highway Administration.)

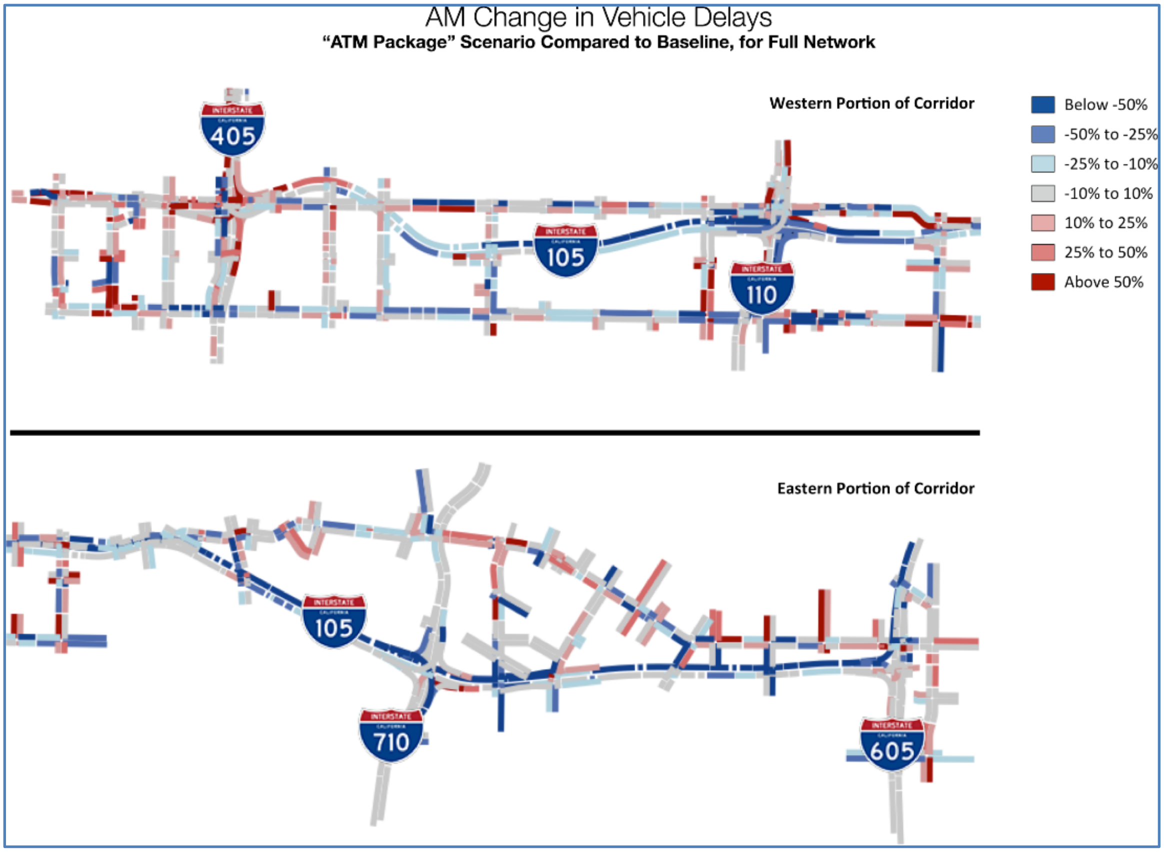

Figure 40. Map. Comparison of change in delays.
(Source: Federal Highway Administration.)

The identification and refinement of improvement strategies is not typically a linear process. Initial identification takes place in close concert with the concept of operations and design phases—formulating likely strategies and combinations of strategies, then mapping the strategies to the existing conditions output generated in the previous stage. As the analysis continues and the initial results are reviewed and shared with the design team, some modifications or new alternatives may be proposed as some initial alternatives are found to be impractical or result in unforeseen negative impacts. The design team must also have a good understanding of the latency of implementation for the proposed strategies. For example, activating and implementing "flush" signal timing plans requires time to implement in the field because of operational or technological constraints. The design team must be aware of these constraints and communicate these limitations to the analysis team so they can be properly represented in the analysis.

As these discussions progress and more quantitative data becomes known, stakeholders update and refine assumptions and hypotheses, bringing further clarity and detail to the hypotheses and assumptions associated with the operational conditions and strategies to be analyzed. The analysis team must identify which parts of the envisioned strategies they may want to make dynamic—strategies that could be manipulated in response to changing operational conditions, such as ramp metering, High-Occupancy Vehicle (HOV) 2+ vs 3+, pricing strategies, or traveler behavior such as mode choice. This exercise results in the definition of experimental cases (potential improvement strategies or "with" improvements) that are analyzed and then compared to the control case (baseline, or do-nothing, or "without" improvements). Figure 41 shows a representation of a dynamic mitigation strategy; illustrations such as this help stakeholders visualize the implementation of the strategy, enabling them to make better-educated decisions about how this strategy should be implemented.

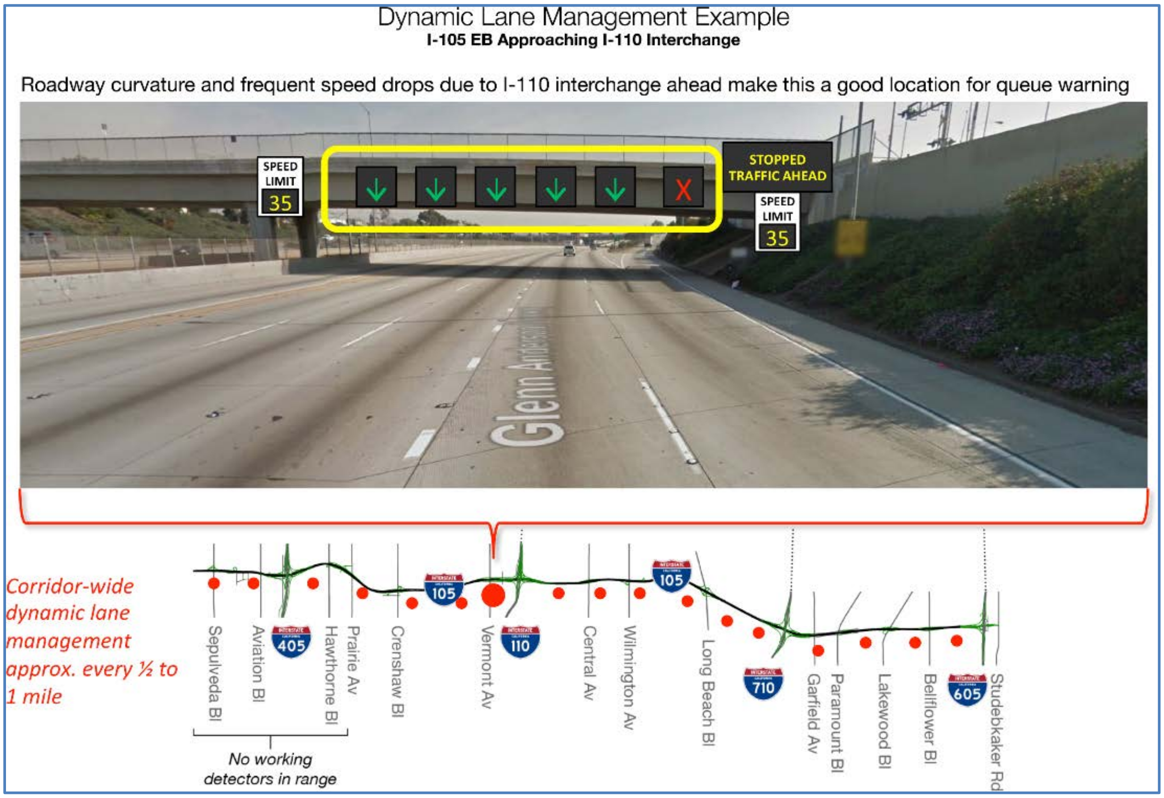

Figure 41. Snapshot. Illustration of an alternative.
(Source: Federal Highway Administration.)

Lastly, in any analysis there is uncertainty associated with various system components that may not have been defined in a robust way. The analyst has little control over "unknowables", but techniques used to minimize or bound uncertainty typically involve sensitivity or risk analysis implemented by varying analysis parameters and assumptions using probabilistic distributions of these parameters based on "back-casting" methods. Expectations on traveler responses to congestion can be bound by ranges of traveler behaviors observed in previous studies; if the analysis estimates that 4-6 percent of travelers will divert to another route in response to congestion, it is advisable to compare this diversion against previously observed traveler behavior in the same area. These methods result in a probabilistic representation of impacts that represent a statistical likelihood that a project will have a certain expected impact. In any case, as long as analytical assumptions and parameters do not vary across different alternatives the analysis can produce fairly valid conclusions on the relative effectiveness of different alternatives on, say, improved mobility.

ANALYSIS METHODOLOGY

A systematic process is required to develop a successful base transportation analysis tool. The process should include clear documentation of the analysis tool structure, assumptions, and calibration criteria. Documenting and following the process ensures that the actual development of the analysis tool is done in orderly way, thereby minimizing coding errors and mistakes. The base analysis tool development process can be broken down into the following areas:

- Network (nodes, links, and link connectors).
- Lane geometry (number of lanes, length of turn lanes, etc.).
- Traffic control information.
- Travel demand.
- Entering volumes.
- Turning percentages.
- Origin-Destination (O-D) information.
- Modes being considered in the analysis and any specialized transit links.
- Improvement strategies being considered and their likely impacts.
- Likely diversion routes within the study area.
- Location of major multimodal transfer locations.
- Jurisdictional boundaries and the need to segment out the performance measures according to these boundaries.
- Special Generators—known locations that create or attract large amounts of trips on regular or irregular schedules (e.g., factories with shift workers, schools, stadiums).
- Error checking for all of the above.

Once the initial baseline network development is complete, the next step is to develop and deploy the trip demand data. Depending on the results of this process, it may be necessary to return to the network development process to adjust parameters within the network.

It is important that quality control is performed at this stage of the process. Because of the detail involved in specifying network and traffic signal parameters, it is fairly easy to miss some of the critical details. A separate team (internal or external) should be assigned to conduct a quality control exercise of the baseline analysis tools developed in this step.

The next step in the analysis setup process is determining travel demand for the baseline period. This process includes identifying study area travel demand and disaggregating this demand into more discrete time-period trip tables. The primary tasks involved in determining travel demand for the baseline year are identified below:

- **Develop trip tables for detailed subareas**. Origin-destination trip tables are established for the subarea model. This requires the aggregating and disaggregating zones into the identified traffic analysis zone structure.
- **Develop time-of-day distribution**. Peak-period trip tables are disaggregated into more discrete time slices. Archived data from automated traffic surveillance monitors is useful in this step to identify the average proportion of travel in the network at any given time.

- **Conduct origin-destination matrix estimation**. This estimation is used to develop a balanced trip table for study area.

In order for the analysis tool to correctly represent not only the traffic volume but the patterns of movement, some type of origin and destination information is required. Using O-D matrices is particularly helpful in complex networks, especially when parallel facilities are included in the network. If the project purpose involves evaluating operational strategies and determining alternate paths based on congestion, then O-D matrices make the use of the model more efficient, especially when analyzing alternate geometric configurations in future networks. Once the O-Ds are established, it becomes easier to test design alternatives and other more complex strategies.

There are multiple methods for developing O-D inputs into transportation analysis models. This is largely dependent on the software, which may have no, partial, or full O-D inputs. Depending on the size, complexity, available data, and software platform selected for the analysis, the O-D estimation technique may include one or a combination of traffic counts, surveys (license plate or roadside), and travel demand models.

A future baseline model (or future no-build alternative) is an essential part of the analysis process. It is the basis for comparison between alternatives but in a future time horizon as required by analysis needs. For the purposes of transportation analysis, a common methodology for developing future demand forecasts is to take into account regional growth due to land use, demographics, and socioeconomic activity. In cases where this information does not exist, it is acceptable to utilize a trend projection of travel demand.

ANALYSIS TOOL CALIBRATION

Analysis tool calibration is the process of systematically adjusting tool parameters so that the tool is able to reproduce the observed traffic conditions. The process is continued until the error between the performance measures taken from the field data and the performance measures calculated in the analysis is less than a predetermined margin of error. Once the analyst verifies that the tool does reproduce observed conditions, calibration can focus on specific performance measures, such as volumes, speeds, travel times, and bottlenecks. It is important to note that in the calibration process, there needs to be a tradeoff between the required precision and the available resources to collect data and conduct the analysis. Figure 42 presents an example of freeway speeds in space and in time for one analysis scenario—this type of diagram helps the analyst compare observed and modeled speeds in space and in time, so an assessment can be made about whether the model can adequately replicate existing conditions.

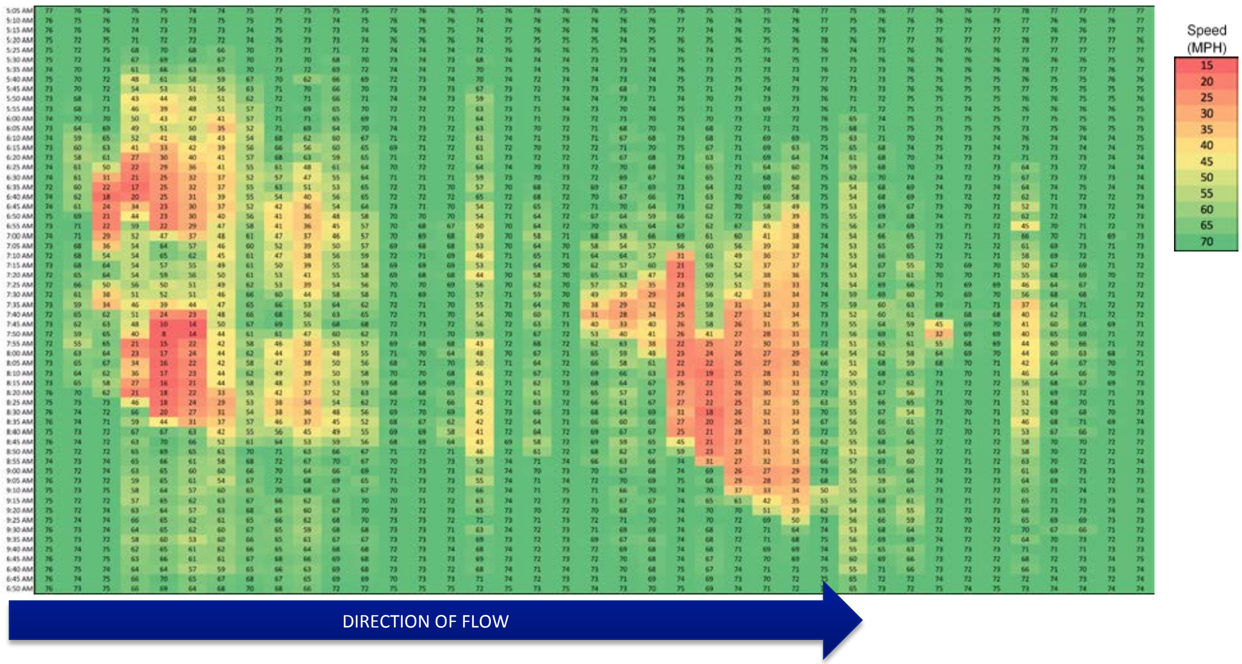

Figure 42. Heat map. Speed diagram for one analysis scenario—I-15 simulation.
(Source: Federal Highway Administration.)

The goal of calibration is to make the tool represent local, observed traffic conditions. However, since traffic may vary greatly from day to day, it is not possible for one tool to accurately represent all possible traffic conditions. Driver behavior differs by region, and it may differ significantly from normal to non-typical days. For example, poor visibility, severe weather, incidents, presence of trucks, and pavement conditions all impact driver behavior. We do not recommend using a tool developed with data from one region to represent traffic conditions in another region. Investment decisions made using a tool that has not been calibrated to local field conditions will be flawed. Guidance on the overall analysis tool calibration process is presented in Volume III of the Traffic Analysis Tools suite. The overall tool calibration process can be divided into three main steps, as described below.

- Identify the performance measures related to the analysis objectives and critical locations for the tools to be calibrated against.
- Determine the strategy for calibration, consistent with Volume III of the Traffic Analysis Tools suite.
- Conduct model calibration runs following the strategy and conduct statistical checks; when statistical analysis falls within acceptable ranges, then the model is calibrated.

The Volume III of the Traffic Analysis Tools suite update allows for the analysis of projects under different operational conditions with a better-expected statistical fit than a “typical day” trying to represent a reliability space, including operational conditions (different incidents, work zones, weather events, special events, etc.).

During the analysis process and the development of the data collection plan, the types of required field measurements are determined, including speeds, volumes, queuing, and other congestion observations at different locations in the network. Since traffic conditions fluctuate daily, it is important to obtain field data from multiple days. The multiple days of data serve as a database in which field variations are used to determine the tolerance of error in the simulated results. During model development, it is recommended that the spatial and temporal analysis limits extend beyond where and when the congestion in the field occurs. The statistical calibration should not necessarily include every link in the network, but should be focused on the critical design elements within the primary study area.

The number of locations selected for comparing the performance measures in field data against analysis outputs needs to be balanced against the quality and location of the available data, the desired level of statistical confidence, and the availability of resources. The analysis output and reporting for these statistical tests should be similar to the performance measures that will be used later in the analysis of the study area. Selecting the number of data days to be used should be based on an analysis of available data in terms of what data is available and cost-effective to collect. In an urban area where freeway sensor data are archived and readily available, more days of data can be used. In areas where there is no surveillance and where manual or temporary data collection devices are used, collecting a significant amount would be more resource-intensive.

Each transportation analysis software program has a set of user-adjustable parameters that enable the practitioner to calibrate the tool to better match specific local conditions. These parameter adjustments are necessary because no analysis tool can include all of the possible factors (both on- and off-street) that might affect capacity and transportation operations. The calibration process accounts for the impact of these “unmodeled” site-specific factors through the adjustment of the calibration parameters, which is included in the software for this specific purpose. Therefore, calibration involves selecting a few parameters for calibration and the repeated operation of the tool to identify the best values for those parameters. Calibration improves the ability of the tool to reproduce local travel conditions accurately. The key activities in calibration are listed below:

- Identify necessary calibration targets.
- Select the appropriate calibration parameter values to best match locally measured street, highway, freeway, and intersection capacities.
- Select the calibration parameter values that best reproduce current route choice patterns.
- Calibrate the overall analysis tool against overall system performance measures, such as travel time, delay, and queues.
- Document the above four activities.

The final step is to document the approach to the calibration process, including criteria and acceptance targets used, output results, any unresolved issues, and lessons learned. These items should be detailed in the calibration/validation report. A Methods and Assumptions section typically is part the analysis tool calibration documentation. In this section, the analyst keeps track of all assumptions and decisions made in the analysis dealing with uncertainty, poor data quality, and other factors that may influence the analysis downstream.

Developing and calibrating the analysis tools is often the riskiest task in the analysis; it requires the greatest investment in time and resources. The analysis team must take special care to calibrate the tool so that it replicates existing conditions (recurrent and non-recurrent traffic congestion, as well as transit system performance) as closely as possible. Errors resulting from misspecification (e.g., not gathering concurrent data) or incorrect expectations (e.g., by relying on anecdotal perceptions of corridor problems rather than archived data) can have a significant impact on project budget and timeframe. Failure to suitably invest resources in this task can result in models that are incapable of providing the correct assessment of impacts, which could require repeating this effort at significant cost. Finally, an analyst should take care not to fall into a trap of "calibrating everything"—a quixotic quest to develop models calibrated against all possible measures, all possible locations, for all possible days. Calibration should focus on a small and meaningful set of alternative-differentiating measures for locations and conditions most critical in separating the potential approaches under evaluation.

The need to calibrate and validate the developed analysis tools correctly cannot be understated. The correct calibration of the tools will influence the accuracy of the analysis outputs and animation and will ultimately determine the success of the analysis approach.

ALTERNATIVES ANALYSIS

This activity develops alternative scenarios within the analysis tools developed and calibrated in previous tasks. These alternative scenarios are analyzed and the results documented according to guidelines provided in the analysis plan.

The analysis provides critical feedback to system designers and operators to allow them to steer their investment into the right strategies. This objective of the analysis not only includes the major investment decisions (prioritizing and selecting the right mix of strategies to deploy), it also includes the ability to assist planners and operators in devising appropriate operating parameters and concepts of operation to optimize the impacts of the selected strategies.

Analysis allows planners and operators to maximize their investments in their improvement strategies by allowing for the analysis of various "what if" scenarios to test different operational schemes, parameters, and concepts to optimize the efficiency of the proposed systems. Analysis allows for these "what if" scenarios to be tested, modified, and refined without having to resort to real-world experimentation, where any mistakes would have high costs.

The ability to analyze various conditions and test different operational parameters under these varying conditions helps operators identify deficiencies in their operational plans that would result in inefficiencies in operating the transportation system. Different modifications and refinements can be tested and compared using analysis and benefit/cost analysis to develop optimal plans to maximize the efficiency of the transportation investment. The typical alternatives analysis process involves testing different operating parameters and the refining of different scenarios through several iterations of analysis, comparison, and adjustment until the resulting operational assumptions and parameters produce the most optimal level of performance, thus maximizing the effectiveness of the transportation investment.

Successful completion of this activity results in a prioritization of potential transportation investments and a clear communication of the potential project benefits. Alternatives analysis represents the culmination of all of the previous analysis tasks; the results can be used to shape investment decisions and to secure deeper and broader support for the project among stakeholder organizations (executive agency leaders and managers, planners, operators, analysts and engineers in these organizations) and elected officials. Analysis managers are encouraged to work with communications professionals to translate technical results into accessible visual and "soundbite" messages that can be shared broadly. The following work steps comprise the alternatives analysis.

Develop Future Baseline Networks and Trip Tables for All Operational Conditions

Once the existing baseline tools have been calibrated, the analysis team can then proceed to develop the future baseline analysis tools. The analysis plan defines all the alternative scenarios that need to be analyzed in this task. Analysis tool networks and trip tables should be modified according to the analysis plan guidelines to represent the scenarios and the impact of the improvement strategies.

Conduct Analysis of Improvement Strategies for All Operational Conditions

The analysis should be conducted using the modified networks and trip tables developed in the previous step. This may include multiple tool runs for each scenario, depending on whether analysts want to conduct additional verification of results—assure results generally are within expected realms (e.g., that Vehicle-Miles Traveled [VMT] results are approximately the same as in the baseline)—and investigate counterintuitive results. All analysis tool runs should be adequately documented to ensure the application of the correct inputs and assumptions according to the analysis plan. Supporting steps for conducting this analysis include:

- Evaluate the initial operational analysis assumptions, scrutinizing the results for any underperforming or counterintuitive metrics.
- Brainstorm a number of causes for the underperformance and a potential set of "what if" adjustments that might be made to alleviate the deficiencies.
- Formulate a set of scenarios that may be evaluated in the analysis structure to assess the impacts and benefits of adjustments to the operational assumptions.
- Analyze, compare, and refine—and re-run through the analysis procedures as necessary—to identify the optimal operating parameters.
- Document the tested scenarios and results for potential future use.
- Re-conduct the refinement process in a continual feedback loop as future conditions change or encountered deficiencies are warranted.

Following these steps helps ensure that the analysis identifies the appropriate technologies and strategies to deploy and that the strategies are operated in the manner that best optimizes the transportation investment.

Assess Performance Measures

In this step, the analysts assess the results of the previous step in light of the performance measures defined in the analysis plan. Below is a summary of how the basic types of performance measures defined in the analysis plan can be calculated to gain insight into overall benefits to system performance:

- **Mobility**. Three primary types of measures were used in the analysis plan to quantify mobility: travel time, delay, and throughput. Travel time and delay are fairly straightforward to calculate using model outputs. Throughput is calculated by comparing travel times under the incident scenarios to those under no incident. By comparing the percentage of trips under the same threshold travel time in the with- and without-improvement scenarios, the relative influence of transportation improvements on reducing extreme travel times can be estimated.

- **Reliability and Variability of Travel Time**. Travel-time reliability/variability is reported in terms of changes in the Planning Index or changes in the standard deviation of travel time.

- **Emissions and Fuel Consumption**. Estimates can be produced by using emissions and fuel consumption rates based on factors such as facility type and vehicle mix, combined with model output, such as travel speed.

- **Safety**. Safety is typically measured in terms of accidents or crashes in the study area, including fatalities, injuries, and property-damage-only accidents. Currently available safety analysis and prediction methodologies are not sensitive to transportation improvement strategies. At best, available safety analysis methods rely on crude measures such as Volume-to-Capacity ratio (V/C), or empirical comparison methods, such as identifying safety benefits that result from implementing a certain type of mitigation strategy and applying the same expected improvement rate to a future implementation of the same or similar strategy. Clearly, this is an area deserving new research. Figure 43 and Figure 44 show how existing accident rates can be presented in space and in time and how this depiction can help analysts determine problematic locations and time spans when accident rates are greater than average.

- **Cost Estimation**. Planning-level cost estimates can be prepared based on life-cycle costs (capital, operating, and maintenance costs). These costs can be expressed in terms of the net present value of various components. Annualized costs represent the average annual expenditure expected to deploy, operate, and maintain the improvement and replace equipment as it reaches the end of its useful life.

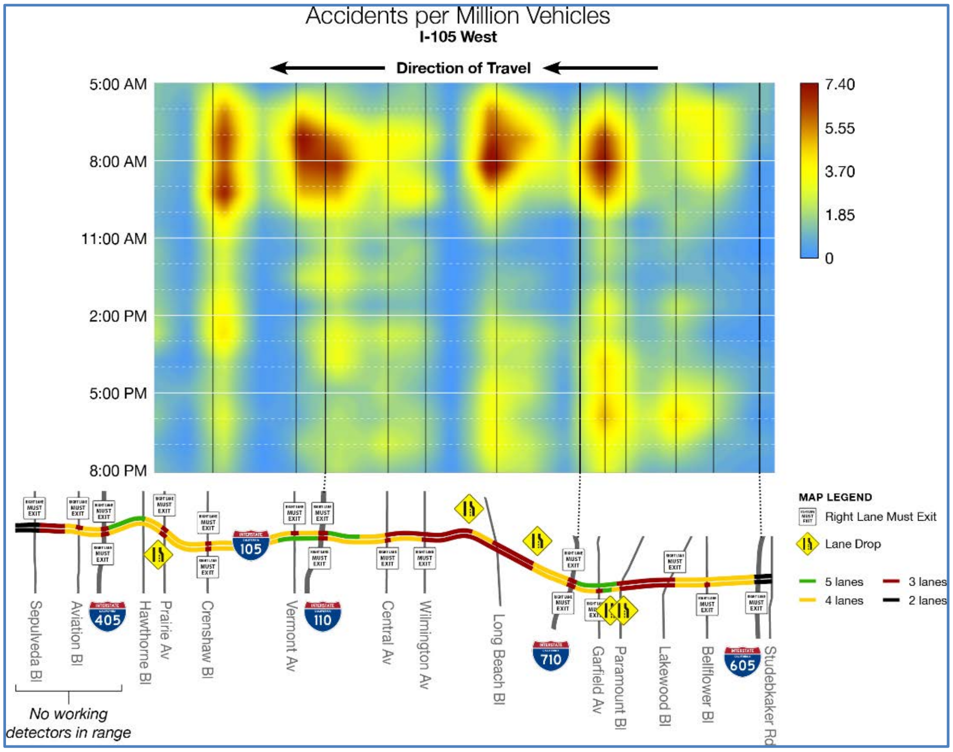

Figure 43. Snapshot. Accident rates in space and in time.
(Source: Federal Highway Administration.)

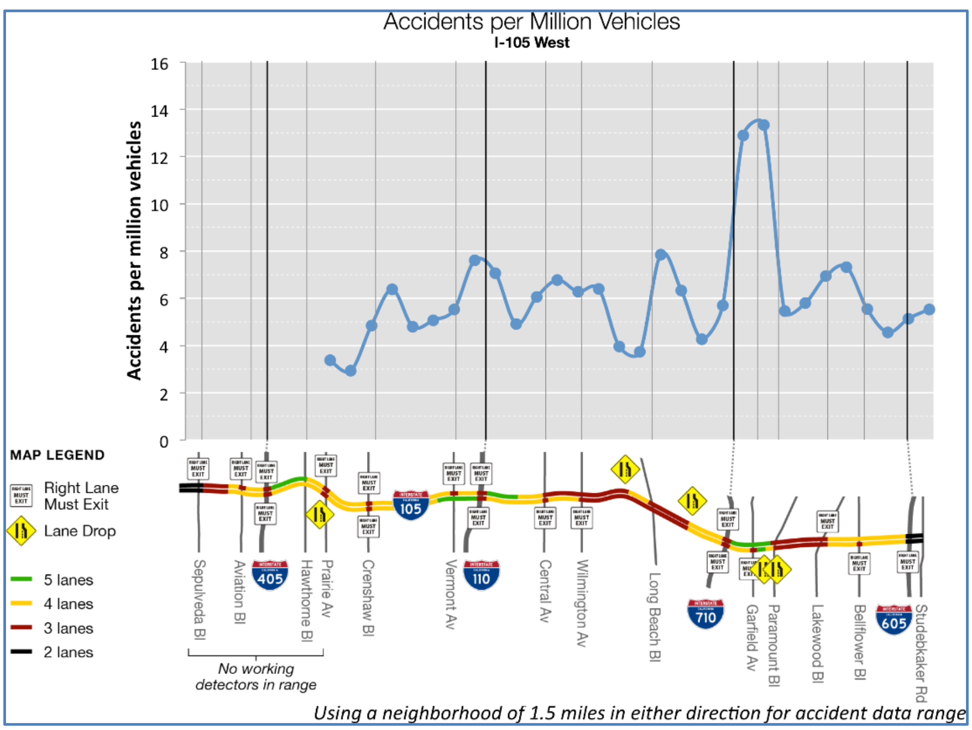

Figure 44. Snapshot. Accident rates by location.
(Source: Federal Highway Administration.)

Conduct Benefit-Cost Evaluation for all Performance Measures

Benefits should be estimated for the improvements by monetizing the incremental change in performance measures associated with the strategies and scenarios analyzed. The incremental change in the performance measure should reflect the weighted sum of changes for all analysis scenarios. The results of this analysis should enable the analysis team to assess the optimal strategies—or combinations of strategies—that can deliver the greatest impact on the corridor's transportation objectives for the cost.

A benefit/cost analysis—a rich analytic method in its own right with its own supporting literature—should be undertaken carefully. A properly calculated benefit/cost analysis monetizes metrics that are comprehensive, mutually exclusive, and designed to render all effects to the appropriate side of the ledger as either a cost or a benefit. There are tools available that provide costs for capital, operating, and maintenance for developing benefit/cost analysis, such as "Operations Benefit/Cost Analysis TOPS-BC User's Manual: Providing Guidance to Practitioners in the Analysis of Benefits and Costs of Management and Operations Projects (June 2013)."

To estimate the benefits in annual dollar values, the annual incremental change in the various performance measures should be multiplied with an estimate of the monetary value of benefits (e.g., the value of an hour of travel time saved). Monetary values of benefits (e.g., value of time, value of accident reduction) should be consistent with those values typically applied in the region. For those performance measures with no established local value, national benefit valuations may be applied.

For the identified improvement strategies, planning-level cost estimates need to be prepared, including life-cycle costs (capital, operating, and maintenance costs). Costs, expressed in terms of an annualized cost or the net present value of various components over a given time horizon (e.g., 20 years), are defined as follows:

- **Capital Costs**. Include up-front costs necessary to procure and install equipment. These costs are shown as a total (one-time) expenditure and include the capital equipment costs, as well as the soft costs required for design and installation of the equipment.

- **Operations and Maintenance (O&M) Costs**. Include continuing costs necessary to operate and maintain the deployed equipment, including labor costs. These costs are presented as annual estimates.

- **Annualized Costs**. Represent the average annual expenditure expected in order to deploy, operate, and maintain the transportation improvement; and replace (or redeploy) the equipment as they reach the end of their useful life. Within this cost figure, the capital cost of the equipment is amortized over the anticipated life of each individual piece of equipment. This annualized figure is added with the reoccurring annual O&M cost to produce the annualized cost figure.

Figure 45 shows an example comparison of two alternatives across various performance measures. Figure 46 illustrates monetized expected benefits for a certain alternative across the same performance measures. The combination of these two exhibits provide the analysts with a comparison of expected improvements resulting from implementation of alternative mitigation strategies based on absolute values of the expected improvements in different performance measures, and monetization of these benefits for different parts of the transportation network.

AM Performance Measures

System-wide, for a typical day

		Percent Change from Baseline	
Metric	**Without ATM**	*ATM Package*	*HSR Only*
VMT (vehicle miles)	1,400,468	1.51%	0.70%
VHT (vehicle hours)	44,832	-26.74%	-22.13%
Vehicle Hours of Delay (vehicle hours)	20,451	-59.59%	-48.64%
Person-Miles Traveled (passenger miles)	1,953,167	1.35%	0.50%
Person-Hours Traveled (passenger hours)	60,453	-25.06%	-20.87%
Person-Hours of Delay (passenger hours)	26,677	-57.68%	-47.30%
Average Travel Time (seconds per mile)	115.24	-27.83%	-22.67%
Average Trip Time (minutes per trip)	6.34	-27.31%	-22.47%

Figure 45. Chart. Comparison of alternatives across performance measures.
(Source: Federal Highway Administration.)

AM Monetized Benefits Per Year

"ATM Package" Scenario

Benefit Category	Freeways Only	Arterials Only	System-Wide
Person-Hours of Delay Saved (recurrent congestion)	$37,997,253	$9,707,588	$47,704,841
Emissions Reductions	$801,901	$220,067	$1,021,968
Fuel Savings	$2,241,780	$845,969	$3,087,749
Collision Reductions	$5,350,762		$5,350,762
Improved Travel Time Reliability	$8,123,244		$8,123,244
Total Monetary Benefits	$54,514,940	$10,773,624	$65,288,564

Figure 46. Chart. Summary of monetary benefits across performance measures.
(Source: Federal Highway Administration.)

DOCUMENT ANALYSIS RESULTS

Upon completion of the alternatives analysis, the results should be documented in an Analysis Report. The Analysis Report presents performance measures for all alternatives; benefit/cost analysis for each alternative; and a prioritized list of improvement strategies for each scenario. This document should build upon data described in the analysis plan, the data collection plan, and the calibration/validation report. The Analysis Report should function as a stand-alone document that fully encapsulates the process and the results of the analysis.

In creating the Analysis Report, the team should refer back to the analysis plan documentation to make sure that all anticipated analyses have been successfully performed and document any deviations from this plan. The Analysis Report should also document lessons learned through the completion of the alternatives analysis.

In assessing the analysis results, the analysts need to weigh the model outputs carefully against the expected outcomes identified in the analysis plan. Where discrepancies exist, further scrutiny is required to assess whether the unexpected outcomes are a result of discrepancies in the analysis method or whether the expected outcomes were not realistic. If the analyst determines that strange analysis results are a result of analysis discrepancies, modifications to input parameters may be considered and the alternative re-analyzed; however, it is critical that any modifications to the analysis input be carefully documented and presented in the Analysis Report.

Examples on Reporting Results of the Alternatives Analysis

Transportation analysis tools produce many performance measures and data that help provide a quantitative assessment of alternatives. Reducing these data down to a few core tables and visualization of essential information is needed for an effective decision-making process. Figure 47 through Figure 49 are examples of reporting analysis results across different alternatives and network segments, including speeds, densities, and delays.

Analysis Segment	Design Year							
	2005		2015 (Alternative A)		2015 (Alternative B)		2025	
	Speed[1]	Density[2]	Speed[1]	Density[2]	Speed[1]	Density[2]	Speed[1]	Density[2]
I-94 Merge to High Ridge Exit	64 (64)	11 (6)	32 (64)	34 (9)	64 (64)	15 (9)	64 (64)	16 (11)
High Ridge Exit to High Ridge Entrance	64 (64)	10 (4)	7 (64)	111 (6)	63 (64)	14 (6)	63 (64)	13 (7)
High Ridge Entrance to I-94 Diverge	58 (63)	12 (4)	8 (62)	108 (7)	61 (64)	15 (6)	61 (63)	17 (7)

Notes: 1. Speed is expressed as mi/h.
2. Density is expressed as veh/lane/mi.
Table values are listed as XX (YY), where XX represents the a.m. peak average and YY represents the p.m. peak average.
Shaded cells represent where the average speed drops below 30 mi/h for any peak period for that alternative.

Figure 47. Chart. Sample measures of effectiveness summary table, Minnesota Department of Transportation/ Federal Highway Administration Traffic Analysis Toolbox Volume IV. (Source: Federal Highway Administration.)

Direction	Scenario 13	Scenario 14	Difference	% Difference
WB	14,109	10,757	-3,351	-23.8%
EB	971	920	-51	-5.3%
Total	**15,080**	**11,677**	**-3,403**	**-22.6%**

Figure 48. Chart. Delay Comparison between two scenarios: In vehicle-hours; I-210 simulation. (Source: Federal Highway Administration.)

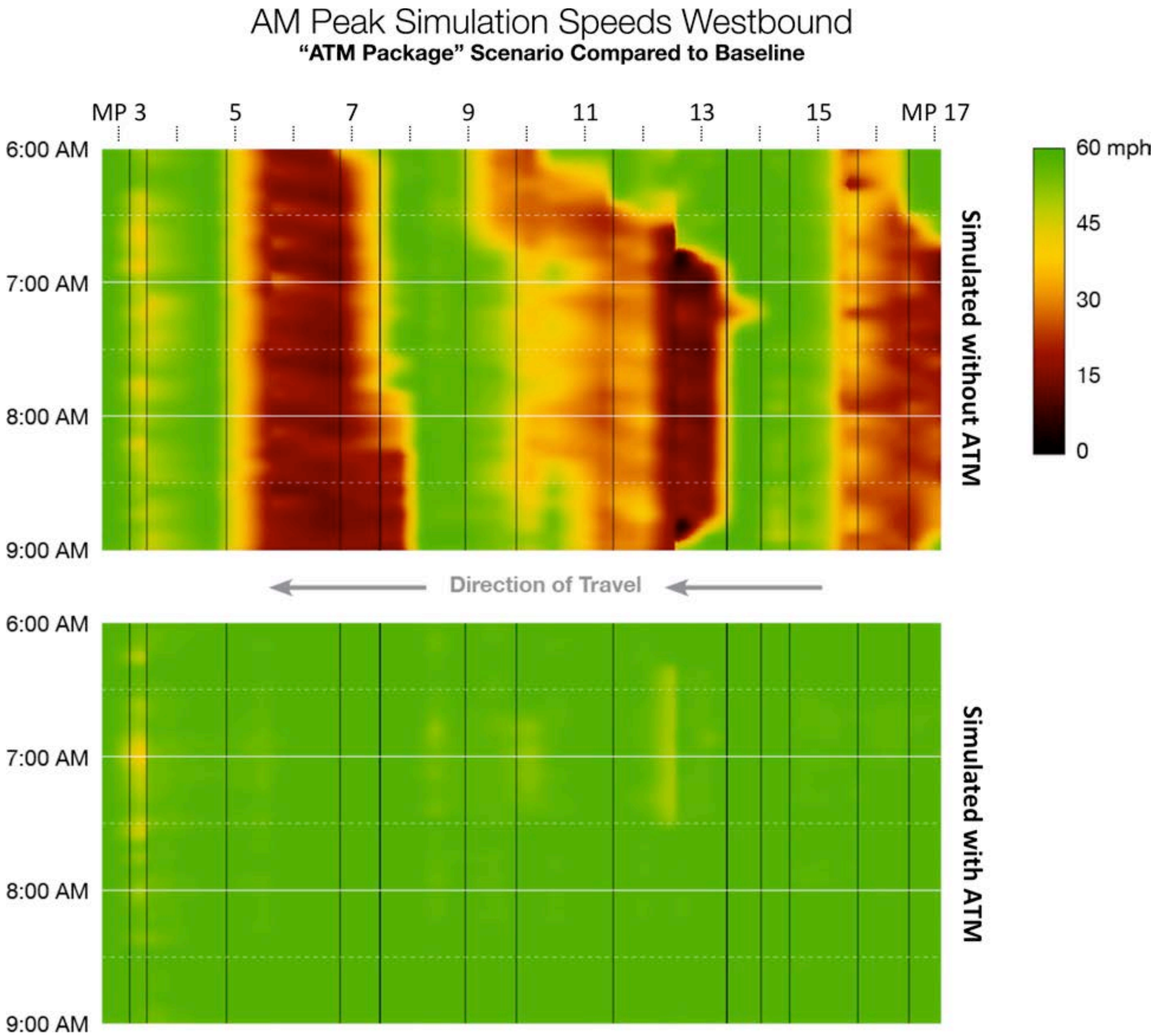

Figure 49. Heat map. Comparison of freeway speeds between alternatives. (Source: Federal Highway Administration.)

DATA, ANALYSIS TOOLS, AND MODEL SUSTAINABILITY

Given the dynamic nature of traffic congestion, it is likely that the analysis process will need to be repeated in the future—perhaps as often as every five years—in order to adjust the transportation improvement components to current conditions. The agencies in charge of the analysis must ensure the maintenance of the analysis tools, models, and datasets, greatly reducing the costs, enhancing the ease with which future analyses may be performed, and improving the effectiveness in which future investment decisions are made. The analytical capital accumulated through this continual improvement process serves (depicted as a central component of the 21st Century Analytic Project Scoping Process in Figure 2 in the Introduction of this Guide) not only to improve the analysis that is currently being conducted immediately, but also to enhance analytical capabilities for future analysis of strategies and investments.

Transportation analysis represents a significant investment by the implementing agency. The processes and tools developed for the analysis have numerous potential future applications. Therefore, the datasets and tools developed for the analysis should be carefully archived. Infrastructure must be in place to accommodate analysis tools reuse, and data dictionaries and user guides should be developed in parallel to assist in the future use of the analysis outputs. If the transportation improvements are proven to have significant benefits, there are likely to be calls to expand the study area or apply the strategies to additional areas in the region. Therefore, proper maintenance of the models, tools, and datasets will ensure that these future analyses can be performed at a greatly reduced cost and with improved ease of application.

While maintaining the analysis tools, models, and datasets may require a mindset change for some agencies unaccustomed to these activities, the investment has significant benefits. Continuous improvement may require changes to agency policies, work habits, and data processes and systems. There is a tendency to want to forego this feedback task once the major analysis tasks have been completed. However, this task is absolutely critical to improving the analysis. Therefore, the resources necessary to complete this ongoing task should be planned for in the analysis plan, and analysis managers should devote adequate effort to ensure its full and successful completion.

Lastly, it is important to document lessons learned from the scoping and analysis process to benefit and lower the risk of future scoping and analysis efforts. The lessons learned can relate to the analysis methodology, data sources, the data used to represent different operational conditions, the model's representation of the transportation network, the estimated versus actual cost of the analysis effort, difficulties encountered in the model calibration effort, and skill gaps identified that are needed to improve the effectiveness of the analysis team. Proper documentation of these lessons learned can help future analysis efforts avoid the same mistakes and more effectively conduct future transportation analyses.

To capture these lessons learned and preserve developed analytical capital (models and data), a Project Results Summary document should be assembled covering the following topics (see Table 6).

Table 6. Project results summary elements.

Project Definition	A concise statement of the overall system problem including cross-validation and other insights from stakeholders on the nature of the issue and potential solutions.
Geographic Scope	The geographic area covered by the analytical project, including a statement of the required detail of representation within this geographical area.
Temporal Scope	The times of day, days of week, seasonality, and years of operation assessed in the analytical effort. This includes an assessment of the simulation horizon.
Hypotheses	The *hypotheses* represents the leading underlying cause of the system performance issue.
Results Summary	A text description summarizing the analytical results of the effort. This section should reference the final report that details project findings.
Analytical Approach	A description of the method used for evaluating the effectiveness of the mitigating strategies in resolving the system performance issue.
Developed Models	The one or more tool types used in the analytical approach, and the models developed to represent the system. This section should identify where these data are archived and documented.
Data Resources	A summary of data used to characterize operational conditions, represent alternatives, and model the geographic and temporal aspects of the system. This section should identify where these data are archived and documented.
Alternatives Modeled	Detailed description of the alternative solutions and/or operational practices assessed within the analytical project.
Key Operational Conditions	The set of travel demand, incident, and weather conditions under which a meaningful examination of alternative impacts were conducted.
Selected Performance Measures	The measures of system performance used in the effort.
Actual and Expected Costs	The actual and projected cost of the analytical project, including data collection.
Actual and Expected Schedule	The actual and projected time to conduct the analysis, including data collection.
Lessons Learned	An assessment of lessons learned regarding technical and non-technical issues.
Risks	A summary of risks comprising risks in data collection, technical risks, and non-technical risks—and how they were overcome or mitigated in the effort.

GLOSSARY

ID	Term	Definition	Source of Definition
1	Alternative	A specific set of specific operational practices or transportation system improvements represented and evaluated within a microsimulation analysis. The impact of each alternative is compared to other alternatives in the analysis based on key performance measures to inform transportation management decisions.	TAT Volume III Update
2	Computational Platforms	A computing platform is, in the most general sense, whatever a pre-existing piece of computer software or code object is designed to run within, obeying its constraints, and making use of its facilities. The term computing platform can refer to different abstraction levels, including a certain hardware architecture, an operating system (OS), and runtime libraries. In this document, the term computational platform is intended to describe the computing system used by a transportation analyst when conducting an analytic project—the platform on which the transportation analysis software runs.	Wikipedia
3	Connected vehicles	Connected vehicles interact with each other (V2V), the roadside (V2I), and beyond (V2X) via wireless communications.	ITS Strategic Plan
4	Crowdsourcing	The practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers.	Merriam Webster
5	Dedicated Short-Range Communication (DSRC)	DSRC is a two-way short- to- medium-range wireless communications capability that permits very high data transmission critical in communications-based active safety applications.	ITS Strategic Plan
6	Intermodal	Involving more than one mode of transport during a single journey, e.g., a commuter drives to a bus stop; the bus transports him/her to a subway station for the remainder of the trip; a freight container travels by ship to a port where it is transferred to rail, and then to truck at the destination area.	its.dot.gov
7	Internet of Things	A concept that built infrastructure is linked and dynamic, often displaying intelligence via integrated sensors and adaptive controls.	its.dot.gov

ID	Term	Definition	Source of Definition
8	Interoperability	The ability of two or more systems or components to exchange information and to use the information that has been exchanged.	ITS Strategic Plan
9	ITS Project	ITS Project means any project that in whole or in part funds the acquisition of technologies or systems of technologies that provide or significantly contribute to the provision of one or more ITS user services as defined in the National ITS Architecture.	Rule 940
10	Macroscopic Simulation	Simulation based on deterministic relationships of the flow, speed, and density of the traffic stream; simulation takes place on a section-by-section basis.	its.dot.gov
11	Metadata	A set of data that describes and gives information about other data.	Oxford Dictionary
12	Metropolitan Planning Organization (MPO)	A federally mandated and federally funded transportation policy-making organization in the U.S. that is made up of representatives from local government and governmental transportation authorities	ITS Strategic Plan
13	Microscopic Simulation	Simulate movement of individual vehicles based on car-following and lane-changing theories; vehicle enter a network using statistical distribution of arrivals and are tracked through the network.	its.dot.gov
14	Mobility	The ease with which members of a society move using transportation infrastructure such as roadways, bridges, pedestrian walkways, bike lanes, rail, etc.	ITS Strategic Plan
15	Model Calibration (also called Analysis Tool Calibration)	Model calibration is the process of systematically adjusting model parameters so that the model is able to reproduce the observed traffic conditions.	Module 2
16	Multimodal	Describes more than one mode of transport, or a technology application that affects more than one mode of transportation.	its.dot.gov
17	Operational Conditions	A set of analogous conditions within the transportation system based on variation in travel demand, incident patterns, prevailing weather, and other factors that create similar system performance.	TAT Volume III Update
18	Outlier Data	An observation that lies outside the overall pattern of a distribution.	
19	Performance Measures	Performance measures support objective and serve as a basis for comparing alternative improvement strategies (investment and policy approaches) and for tracking results over time.	ITS JPO Fact Sheet

ID	Term	Definition	Source of Definition
20	Probe Data	Data generated by vehicles about their current position, motion, and time stamp. Vehicle probe data may be generated by devices integrated into the vehicle itself or provided by nomadic devices brought into the vehicles.	ITS Probe Data Fact Sheet
21	Roadside Unit (RSU)	A connected device that is only allowed to operate from a fixed position (which may in fact be a permanent installation or from temporary equipment brought on-site for a period of time associated with an incident, road construction, or other event). Some RSUs may have connectivity to other nodes or the Internet. Also referred to as Roadside Equipment (RSE).	ITS Strategic Plan
22	Scenario	A scenario describes one way that a system is or is envisaged to be used in the context of activity in a defined timeframe. The timeframe for a scenario could be (for example) a single transaction; a business operation; a day or other period; or the whole operational life of a system. Similarly, the scope of a scenario could be (for example) a single system or piece of equipment; an equipped team or department; or an entire organization. In this document, a scenario refers to a specific combination of a distinct operational condition and a particular alternative operational strategy that generates a resulting pattern of system performance (either in reality or in a modeled environment).	Wikipedia
23	System	A combination of interacting elements organized to achieve one or more stated purposes.	INCOSE
24	Systems Engineering	Systems Engineering is an interdisciplinary approach and means to enable the realization of successful systems. It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, then proceeding with design synthesis and system validation while considering the complete problem.	INCOSE

REFERENCES

1. Sneed, Annie. 2015. *Moore's Law Keeps Going, Denying Expectations*. Available at: http://www.scientificamerican.com/article/moore-s-law-keeps-going-defying-expectations/.
2. Federal Highway Administration (FHWA). 2014. MODEL LONG-RANGE TRANSPORTATION PLANS: A Guide for Incorporating Performance-Based Planning. FHWA-HEP-14-046. August.
3. U.S. Department of Transportation (USDOT) 2015. *Transportation Performance Management*. Available at: http://www.fhwa.dot.gov/tpm/.
4. Florida Department of Transportation (FDOT). 2015. *The Florida Transportation Indicators Web site*. Available at: http://www.floridatransportationindicators.org/.
5. American Association of State Highway and Transportation Officials (AASHTO). 2011. *The New Language of Mobility: Talking Transportation in a Post-Recession World*. Available at: http://downloads.transportation.org/ANewWayToTalkAboutTransportation/NewLanguageofMobility.pdf.
6. Federal Highway Administration (FHWA). 2011. *Guide on the Consistent Application of Traffic Analysis Tools and Methods*. FHWA-HRT-11-064. November. http://www.fhwa.dot.gov/publications/research/operations/11064/11064.pdf.
7. Washington State Department of Transportation (WSDOT). 2015. *Traffic Map Archive*. Available at: http://www.wsdot.wa.gov/data/tools/traffic/maps/archive/?MapName=SysVert.
8. Kansas City Scout. 2015. *Traffic Dashboard*. Available at: http://www.kcscout.net/.
9. U.S. Department of Transportation (USDOT). 2006. *Travel Time Reliability: Making It There On Time, All The Time*. Prepared by Texas Transportation Institute with Cambridge Systematics for FHWA. Available at: http://ops.fhwa.dot.gov/publications/tt_reliability/.
10. Kansas City Department of Transportation. 2013. *KC Scout FY 2013 Annual Report*. Available at: http://www.kcscout.net/downloads/Reports/Annual/AnnualReport2013.pdf.
11. University of Maryland CATT Lab. 2016. *The Regional Integrated Transportation Information System (RITIS)*. Available at: http://www.cattlab.umd.edu/.
12. Metropolitan Washington Council of Governments. 2015. *National Capital Region Congestion Report*. Available at: https://www.mwcog.org/documents/2016/07/04/national-capital-region-congestion-report/.
13. Federal Highway Administration (FHWA). 2015. *Planning and Preliminary Applications Guide to the HCM*. NCHRP 07-22. December.
14. Federal Highway Administration (FHWA). 2015. *Traffic Analysis Tools Volumes I-XIV*. Available at: http://ops.fhwa.dot.gov/trafficanalysistools/index.htm.
15. Federal Highway Administration (FHWA). 2014. *Guidance on the Level of Effort Required to Conduct Traffic Analysis Using Microsimulation*. Available at: http://www.fhwa.dot.gov/publications/research/operations/13026/13026.pdf.
16. U.S. Department of Transportation (USDOT).2016. *Assessment Report of US-Japan-Europe Collaborative Research on Probe Data, International Probe Data Work Group Phase 2*. http://ntl.bts.gov/lib/60000/60100/60131/FHWA-JPO-16-356.pdf.

17. U.S. Department of Transportation (USDOT). 2013. *US-Japan Collaborative Research on Probe Data: Assessment Report*. FHWA-JPO-13-091. Available at: http://ntl.bts.gov/lib/51000/51100/51168/US-JP_Probe_Data_Final_Report_v8_2__FHWA-JPO-13-091_.pdf.
18. Federal Highway Administration (FHWA). 2011. *Integrating Mobility Observations in Maintenance and Operations Management Tools*. Available at: http://www.its.dot.gov/presentations/roadweather/pdf/Guevara%20-%20IMO.pdf.
19. Society of Automotive Engineers (SAE) standard J2735. 2009. *Dedicated Short Range Communications (DSRC) Message Set Dictionary*. Available at: https://www.standards.its.dot.gov/Factsheets/Factsheet/71.
20. U.S. Department of Transportation (USDOT). 2012. Presentation on Vehicle Based Data and Availability. Available at: http://www.its.dot.gov/itspac/october2012/PDF/data_availability.pdf.
21. U.S. Department of Transportation (USDOT). 2015. *T3-Webinars: Using Crowdsourced Data from Social Media to Enhance TMC Operations*. Available at: https://www.pcb.its.dot.gov/t3/s150311/s150311_crowdsourced_data_intro.asp.
22. U.S. Department of Transportation (USDOT). 2012. *Traffic Analysis Toolbox Volume XIII: Integrated Corridor Management Analysis, Modeling, and Simulation Guide*. Available at: http://ntl.bts.gov/lib/50000/50600/50615/30B00211.pdf.
23. U.S. Department of Transportation (USDOT). 2015. *Multi-Modal Intelligent Traffic Signal Systems (MMITSS) Impacts Assessment*. FHWA-JPO-15-238. August.
24. U.S. Department of Transportation (USDOT). 2012. *Metadata Guidelines for the Research Data Exchange*.
25. National Institute of Standards and Technology. *Engineering Statistics Handbook. Section 7.1.6: What are outliers in the data?* Available at: http://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm.
26. Chen, Y., J. Kim, and H. S. Mahmassani. 2014. Pattern Recognition Using Clustering Algorithm for Scenario Definition in Traffic Simulation-Based Decision Support Systems. Proceeding in Transportation Research Board 93rd Annual Meeting.
27. M. Azimi and Y. Zhang. 2010. *Categorizing freeway flow conditions by using clustering methods*. Transportation Research Record Vol. 2173.
28. G. Buxi and M. Hansen. 2013. *Generating day-of-operation probabilistic capacity scenarios from weather forecasts*. Transportation Research Part C: Emerging Technologies Vol. 33.

U.S. Department of Transportation
Federal Highway Administration
Office of Operations
1200 New Jersey Avenue, SE
Washington, DC 20590

www.ops.fhwa.dot.gov

January 2017

Publication Number: FHWA-HOP-16-072

This material is based upon work supported by the FHWA under contract number DTFH61-12-D-00048-T-5008.