Digital Transportation Infrastructure Research Roadmap 2024

California Partners for Advanced Transportation Technology

Authors

Anthony Patire, Ph.D.
Research and Development Engineer
California Partners for Advanced Transportation Technology (PATH)
University of California, Berkeley

James Fishelson, Ph.D.
Executive Director
California Partners for Advanced Transportation Technology (PATH)
University of California, Berkeley

Justin Wiley
Strategy and Development Director
Institute of Transportation Studies
University of California, Berkeley

Executive Summary

The United States' transportation data ecosystem is famously fragmented, and balkanized. Different infrastructure owner-operators (IOOs) at the federal, state, county, regional, and local levels all collect their own data with their own methods, hindering our ability to address transportation issues in holistic, data-driven, and collaborative ways. As the entire world becomes rapidly digitized, and both vehicles and the roadway infrastructure produce increasingly massive amounts of data, it is critical that agencies like Caltrans now put as much emphasis on Digital Transportation Infrastructure (DTI) as they have in the past on traditional physical infrastructure like roads and bridges.

As defined by ITS America, DTI "represents the public and private technology assets that create, exchange, or use data to provide information and insights for transportation systems that advance safety and opportunity for all." By building and maintaining a robust DTI, Caltrans can and should be a leader in the nation in breaking down data silos, creating data platforms and APIs that are usable and accessible to all, facilitating cooperation, improving safety and efficiency, and accelerating innovation. California has done this before with the Performance Measurement System (PeMS) and the California Integrated Travel Project (Cal-ITP), and is now well-placed to be the nation's DTI leader.

This DTI Research Roadmap provides an overview of the emerging DTI space, recognizing the incredibly large scope and scale of the challenge, and proposes a path forward for Caltrans to demonstrate realistic success in the near term. While a holistic data dashboard that contains all of the transportation data in the US is obviously not achievable in the next 4 years, if it ever will be, this roadmap proposes a research direction that will:

- Identify low-hanging fruit and early wins
- Aim for real results within five years
- Discover opportunities to obtain high-quality, state-wide data
- Propose limited-scope efforts in high-value situations
- Define a clear customer for the research within Caltrans
- Identify opportunities that minimize organizational complexities and misaligned incentives, especially where Caltrans does not require external stakeholder involvement

It splits potential DTI Research Thrusts into 5 broad areas:

- Holistic Caltrans Data Platform(s)
- Statewide and Cross-Regional Traffic Modeling
- DTI-Enabled AI, Digital Twins, and Multi-Level Models
- Disruptions and Work Zones
- Connected and Digitized Data Platforms

Under these 5 research thrusts, 15 attractive use cases are detailed, including descriptions, value propositions, stakeholders, and more. Of these 15 different use cases, the roadmap selects 4 as being especially attractive for initial efforts, with one standing out above the rest:

- ***Caltrans-Only Data Platform:*** If the reader is left with only one takeaway from this roadmap, it should be that Caltrans must invest time, energy, and resources in creating and maintaining a data platform for all Caltrans data. While this is not achievable all at once, creating PeMS 2.0 is a reasonable goal, creating an accessible platform that contains a network of roads in map format, vehicle flow data, and infrastructure data for Caltrans roads and assets across the state. This may not be the most exciting work, but it would be truly unique and would be the foundation upon which Caltrans can build all other DTI initiatives.

- ***Statewide Mobilti Model as a Core Data Platform:*** Caltrans already maintains a version of a statewide travel-demand model, and a statewide Mobiliti model would represent a super-charged version of this. The speed and scale of Mobiliti make it uniquely suited for statewide analyses: everything from modeling evacuations during wildfires to mode-shift behaviors to travel pattern changes in response to road usage charges to inter-regional travel. This statewide mobility model would provide a simulation counterpart to the above Caltrans-only data platform, and the two would work synergistically with each other.

- ***AI-Enabled Work Zone and Incident Data Collection and Standardization:*** Safety data is one of the most attractive early applications of DTI, both because of the relative paucity of existing data and its extreme importance; safety has and always will be the highest Caltrans priority. This use case will include standardizing work zone and incident data, as well as collecting new data sets by integrating AI, such as by quantifying near-misses with camera-based sensors at intersections and high-risk areas. Ultimately, this should help reduce collisions and fatalities statewide, improving work zones and identifying the riskiest existing roads before a fatality occurs, instead of after.

- ***Long-Distance Digitized Freight Corridor:*** Along with safety, freight is another attractive early application of DTI, both because the outsize impact that heavy-duty vehicles can have on traffic flow and because freight operators are already incentivized to have more data and improve their services: lower fuel consumption, better on-time deliveries, etc. While unlike the other three use cases, this one will require more involvement from other stakeholders, such as the freight operators and other state DOTs, this effort is already ongoing along I-80. One of the first successful DTI applications is actually along this corridor, providing information on available truck parking to drivers looking to rest.

Introduction

The 21st century has seen the rapid digitization of many fields, but especially transportation. New cars can be described as “computers on wheels,” deployment of Connected Vehicle technologies continue apace, data generation and collection systems are orders of magnitude beyond where they were even a decade ago, and much more. Transportation infrastructure has always been more than just roads, bridges, and other physical assets; Digital Transportation Infrastructure (DTI) is becoming more and more important. DTI as a standalone term is fairly new, so even the definition is somewhat in flux. At a high level, it refers to the broader ecosystem of Infrastructure Owner-Operators (IOOs) storing and using various data sets, with focus on interoperability (e.g., making data sets available to other IOOs), increasing needs for [near] real-time availability, private sector involvement, and integrating new technologies (e.g., AI). DTI also has the potential to replace some physical assets, such as with virtual roadside units (RSUs).

Data will always be at the core of all DTI efforts; its greatest promise is to improve our existing transportation data ecosystem that is fragmented, siloed, incomplete, contradictory and impossible to validate. Transportation systems are poorly connected or not connected at all, and the digital infrastructure is fragmented across jurisdictions and across modes of travel. IOOs maintain independent data sets for their own business purposes. Private companies collect their own data and distribute it to the public in the context of their individual business interests. Note that California was an early DTI leader in building connected and integrated data ecosystems, helping to break down data silos and increase availability with early deployments like the Performance Measurement System (PeMS) and the Lane Closure System (LCS).

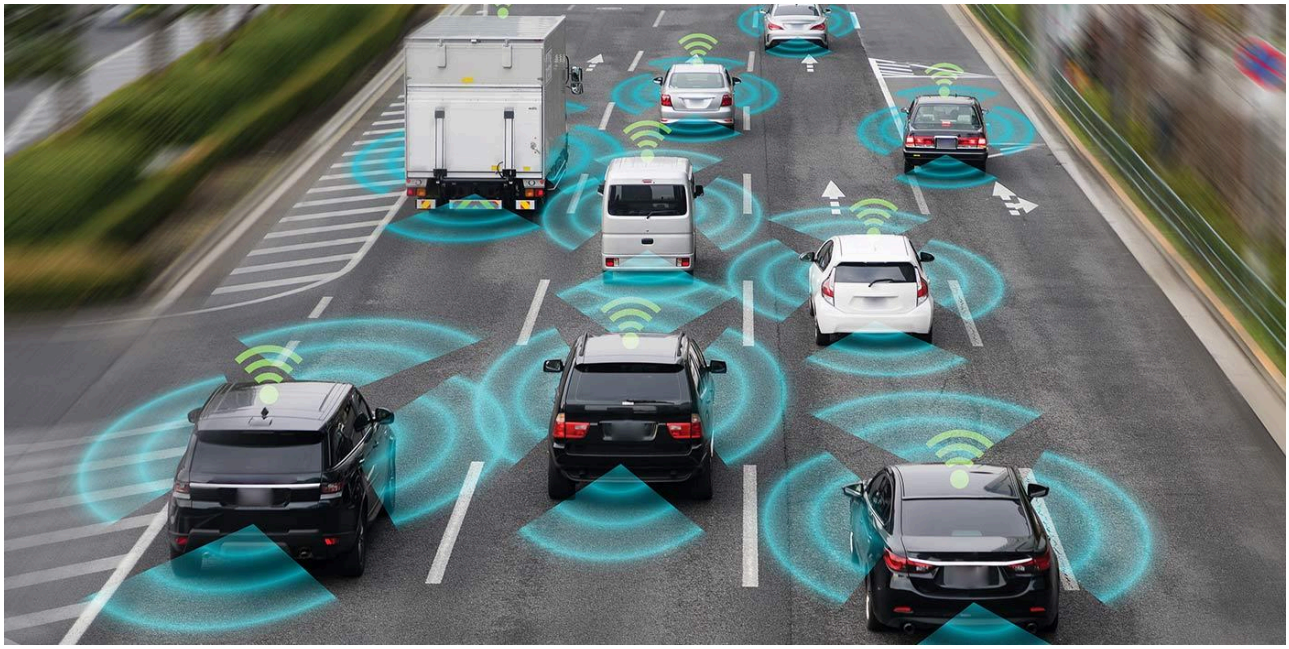
Due to the presence of so many actors with competing goals, it is difficult to find positive ways to establish and maintain alignment to foster effective collaboration. As a result, the ability to identify, understand, and solve today's transportation issues are negatively impacted. In particular, the multidimensionality of today's challenges needs to be reflected in the data, analysis, and measures used to assess progress toward strategic goals. This is a serious challenge and it is getting worse. Private companies such as Google work at global scales, and they are not currently able or willing to integrate hundreds of non-standardized data sources from thousands of individual public agencies. Moreover, many advanced transportation technologies and tech companies implicitly assume a certain maturity of digital infrastructure with high levels of connectivity, bandwidth, and computational power. As every IOO understands intuitively, the large number of different agencies, along with other factors, can make data sharing a huge hurdle, which is very different from the easy access of the internet writ-large, as evidenced by the ability of AI companies to scrape massive amounts of data. To be fair, where incentives align, it is possible to establish standards for public data to be ingested by private companies (e.g., GTFS), but this tends to be the exception rather than the rule.

As illustrated by recent events in Wyoming[1] (where drivers were mistakenly redirected off of one closed freeway onto other closed roads[2]), it is increasingly important for public and private parties to cooperate in the transportation data ecosystem. As new mobility solutions emerge and evolve, travelers become critically dependent on the reliability of the underlying information that supports it. A robust DTI will also support both public and private sectors in enabling extensive product testing and rigorous evaluation of new technologies in real-world settings, especially with an eye toward deployment at scale. A strong DTI system will support both real-world and simulation testing, and everything in between, such as digital twins where the technology can be tested in a simulated environment that is linked to real-time and real world data. DTI (especially when paired with robust data standards) will also support interoperability between systems and at scale. These existing hurdles to effective system integration impede experimentation and development, but also present a serious obstacle toward achieving strategic goals.

The purpose of this document is to present a near-term roadmap for advancing research in DTI. This roadmap aims to carry forward business needs identified in the Caltrans 2020-2024 Strategic Plan[3], but informed by recent advancements in the field as well as new perspectives offered in the more recent ITS America Digital Infrastructure Strategy Report[4]. It specifically focuses on the near-term (i.e., over the next 5 years). Though DTI offers great promise, trying to build a comprehensive and nationwide data ecosystem from our current balkanized state is too much to try to achieve all at once. Instead, this research roadmap focuses on "early wins," identifying what types of DTI Caltrans and her partners can focus on now that will lead from

[1] USDOT ITS Research - Connected Vehicle Pilot Deployment Program. (2023). Dot.gov. https://www.its.dot.gov/pilots/wydot_data.htm
[2] WYDOT Trying to stop Google Maps stranding motorists in frozen nowhere Wyoming, April 14, 2024.
[3] Caltrans 2020-2024 Strategic Plan
[4] Digital Infrastructure Strategy Report - ITS America

research to widespread adoption, including considerations on how it can be funded moving forward. This is especially important considering the size and complexity of California, which has contributed to occasionally lagging beyond other states with respect to DTI and data standards (e.g., with ATMS). Considering the current leadership and the fact that California acts as a microcosm of the broader country, now is a perfect opportunity to re-establish leadership.



The Promise of DTI

DTI is a newly emerging area of critical interest to major transportation stakeholders in the US and beyond. ITS America has a very active Digital Infrastructure working group that advises the federal DOT, and some version of digital infrastructure could be seen in over half of the recently awarded Stage 1 SMART grants. In many ways, the advent of DTI could address many of the challenges that have plagued IOOs over the past few years. Standardized and trustworthy data could facilitate better public-private cooperation, improve reliability and resilience of transportation systems, and accelerate innovation.

Challenges to Achieving that Promise

As an emerging field, specifics around DTI are in short supply. Even the name is not settled: digital infrastructure, digital transportation infrastructure, and roadway digital infrastructure are all used almost interchangeably. These variations in language increase chances for miscommunication and increase the difficulty of agreeing on commonly shared expectations.

Practical issues abound. The transportation ecosystem is complex, with many stakeholders. Responsibilities and data are fragmented and diffused among actors that might not have a history of close collaboration. Incentives are not always aligned.

There is a lack of a standardized geospatial frame of reference. Some states have multiple linear reference systems (LRS). Private competition exists for digital maps and mapping services. As a result, it can be very difficult to relate (or to join) diverse data sets that may have been originally created for vastly different purposes.

Much of the newest transportation data is being generated by the private sector, including vehicle OEMs, and mapping companies. However, there are no widely accepted standards or generalized methods for validation and acceptance for much of this new data.

Overall Approach

The focus for this roadmap is to pick achievable wins that are realistic in the near-term, and can demonstrate measurable success. This roadmap seeks to:

- Identify low-hanging fruit and early wins
- Aim for real results within five years
- Discover opportunities to obtain high-quality, state-wide data
- Propose limited-scope efforts in high-value situations
- Define a clear customer for the research within Caltrans
- Identify opportunities that minimize organizational complexities and misaligned incentives, especially where Caltrans does not require external stakeholder involvement

The Role of Caltrans

Caltrans was among the first to deploy groundbreaking technologies such as PeMS and LCS, mentioned above. As California is a center of technological innovation, Caltrans is well positioned to be a leader, and a model, for other state DOTs. While the importance of interstate cooperation will grow over time, a short-term goal could be to achieve improved cooperation between districts. Moreover, California can act as a microcosm for the entire county: it has long-haul trucking, ports, dense cities, empty rural expanses, snowy mountain passes, hot deserts, and much more, all within a single state.

However, California's size and complexity also makes the deployment of DTI especially challenging, from both technical and organizational perspectives. These difficulties should not be ignored, and this roadmap urges two basic approaches to help ease the deployment pathways:

- ***Broad Use Cases:*** Here, Caltrans can focus on statewide data, but should minimize bureaucratic challenges whenever possible, such as encouraging better data sharing and standards between districts, rather than trying to incorporate data from other IOOs or the private sector. In general, such activities should seek to refine or link with systems that already exist. An example of this broad approach would PeMS 2.0.

- ***Narrow Use Cases:*** Here, Caltrans can focus on discrete use cases with clear customers, obvious value, and relatively straightforward deployment pathways (i.e., do not need the agreement of multiple external stakeholders). An example would be creating better digital infrastructure for work zones, which could help Caltrans both manage their infrastructure better and contribute toward improved safety for road crews.

Relationships Between DTI and Emerging Big Data Technologies

For a balanced view of DTI, it is crucial to consider multiple aspects of the data value chain and lifecycle, including:

- Metadata management
- Data collection and cleaning
- Data validation and quality assurance
- Data exchange and sharing
- Merging or fusion with other data sets
- Analytics or data integration with models
- Generation of metrics and KPIs
- Data storage and archive

Ultimately, the purpose for having data is to extract useful information from it in order to make better decisions. Better decisions will save lives, time, and money.

Emerging technologies such as high performance computing, large-scale traffic models, and artificial intelligence (AI) play an important role in this data value chain. It is worth noting that AI requires enormous amounts of high quality data for training. Until the state of DTI is improved, it will not be possible to obtain the needed data to realize the promise of AI in transportation. Another key consideration is that it is not practical, in general, to obtain anything close to perfect data coverage. At best, one can obtain a small sample and to extrapolate using some appropriate mathematical inference, model, or AI. Therefore, this research roadmap stresses both the importance of having quality data, making that data reusable, and tools to infer as much as possible from that data.

The Relationship Between DTI and AI

Recent advancements in AI have been made possible through the explosive growth of internet content. Analogous progress in transportation-related challenges will likewise require massive amounts of high quality data. Redoubled efforts to collect traffic and transportation data, make it usable through digitization, standards, and best practices are all needed to achieve the promise of DTI. Importantly, the rapid emergence of AI creates new possibilities for accelerating the rollout of DTI, and vice versa. First, a strong DTI can be critical in producing robust and reliable data for intake by data hungry AI models. Second, AI tools can help feed into DTI, especially where existing data is inaccurate and/or incomplete.

Where DTI will act as an input to AI, it will procure high quality data that will be crucial for training new AI models, such as by:

- Creating digital representations of non-standard and unusual traffic situations: work zones, emergency vehicles, extreme weather, etc.
- Leveraging high performance computing, and tools like Mobiliti, for large-scale regional traffic modeling and forecasting
- Synchronizing with the physical infrastructure: to enable Digital Twins

AI will also act as in input to improve DTI, such as by:

- Enabling holistic digitization of roads and factors that interact with roads
- Integrating language-based transportation data

Taken together in a virtuous circle, the combination of AI and DTI may:

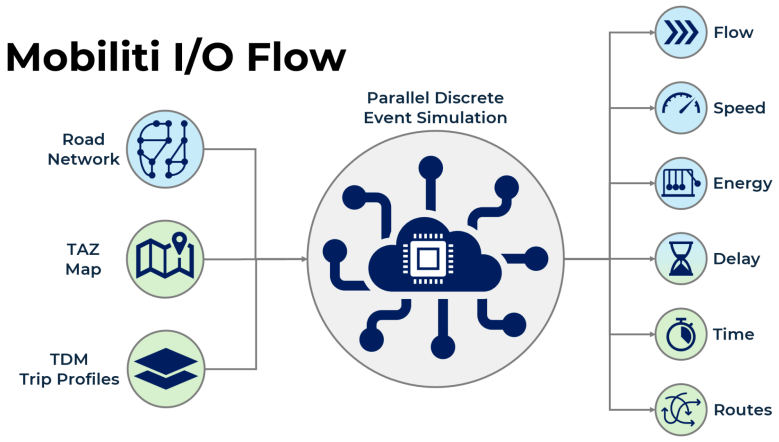
- Extend the benefits of data and automation to all: transportation justice, accessibility for disabled
- Lead towards the development of new AI-powered traffic models

In the near-term, there are several high-impact areas for machine learning and AI:

- Collect and extract data
- Generate new types of data
- Provide for more accurate filters for anomaly detection and data cleaning
- Facilitate joining data across traditionally non-compatible data sets
- Leverage LLMs to accelerate data analytics

High-Performance Computing and DTI: Using Mobiliti as an Example

Beyond AI, a robust DTI will also support and be supported by other big data and computational emerging technologies that are data hungry. Put bluntly, Excel spreadsheets may be sufficient for storing the traffic signal timing of a small city with a single traffic engineer. However, if we are to realize the promise of storing statewide signal data in a single place, something more rigorous will be required, especially for any tools to intake and process said data.



Mobiliti is one example of such a tool: it is a dynamic, massively parallel, discrete event simulation platform that can holistically reveal large-scale cascading effects of dynamic rerouting across a metropolitan region or state. Developed by UC Berkeley under funding from the Department of Energy in collaboration with Lawrence Berkeley National Lab, It is a first-of-a-kind traffic simulator that emerged from a DOE project to understand the energy footprint of transportation, but its utility is much broader.

Mobiliti is a unique platform that provides new capabilities and significantly extends insights from existing regional traffic demand models (TDMs). With continued development, the vision is to use Mobiliti, along with real-time sensor data, to provide digital twinning[5] services to manage transportation infrastructure and support scenario planning. Example uses of this dynamic analytic approach are:

- Understanding network effects and non-local ripple effects
- Characterizing the energy footprint of transportation
- Simulating large-scale transportation disruptions due to weather, earthquake, fire, etc.
- Informing emergency response and evacuation efforts

Mobiliti was originally developed in a high performance computing (HPC) environment at LBNL[6]. It was designed to leverage new parallel processing frameworks being developed on LBNL supercomputers and has since been ported to a cloud environment in order to broaden its availability to the public/private sector.

DTI Priorities

Thematic priorities for DTI can be organized into the following categories reflecting balanced needs for both data and modeling:

- Data Digitization
    - Tools and methods to facilitate digitization of physical assets and situational information
    - Automation of data quality verification
    - Processes for synchronization between digital representations and physical reality

- Data Standards
    - Emerging specifications and standards for data exchange in open formats

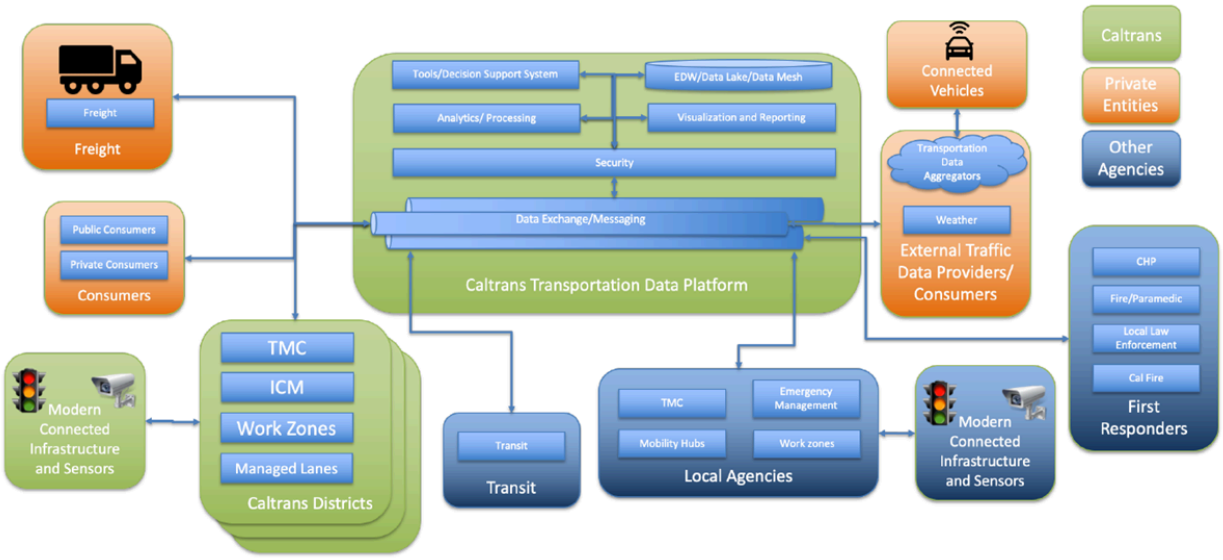
[5] "A digital twin is a digital model of an intended or actual real-world physical product, system, or process that serves as the effectively indistinguishable digital counterpart of it for practical purposes, such as simulation, integration, testing, monitoring, and maintenance." Wikipedia: https://en.wikipedia.org/wiki/Digital_twin. [Accessed 11 19 2024].

[6] Cy Chan, Anu Kuncheria, and Jane Macfarlane. 2023. Simulating the Impact of Dynamic Rerouting on Metropolitan-scale Traffic Systems. ACM Trans. Model. Comput. Simul. 33, 1–2, Article 7 (April 2023), 29 pages. https://doi.org/10.1145/3579842.

    - Collaboration with partners that are already developing DTI standards, such as MobilityData which maintains GTFS, GTFS-rt, and GBFS
    - Building a shared data layer
    - Metadata for interfaces and integrations with multiple applications
    - Security and permissions for “need-to-know” data

- Data Fusion
    - Portals for updated and authorized data sets
    - Metadata to enable combining public and private data
    - Provenance and stability of underlying data sources for data fusion
    - Archival of history and state to enable the generation of consistent KPIs over time

- Modeling and Inference
    - Simulation of large-scale transportation disruptions due to weather, earthquake, fire, etc.
    - Understanding of network effects and non-local ripple effects
    - Emergency response and evacuation efforts both during planning and in near real-time
    - Baseline data and structures for multi-level modeling and integration with transportation-adjacent models involving weather, energy footprint, electrical grid, emissions, air quality, etc.



Potential Future DTI Schematic with Caltrans as a Central Data Broker

DTI Research Thrusts

The DTI research thrusts in this roadmap can be organized into five broad areas plus one additional category of special use cases:

- Holistic Caltrans Data Platform(s)
- Statewide and Cross-Regional Traffic Modeling
- DTI-Enabled AI, Digital Twins, and Multi-Level Models
- Disruptions and Work Zones
- Connected and Digitized Data Platforms
- Special Use Cases (high level overview only)

These thrusts are not mutually exclusive, and many potential projects will naturally overlap onto multiple areas. Still, DTI use cases are grouped below under each thrust area, summarized below in Table 1. The following section recommends next steps, including the four use cases for Caltrans to prioritize. Note that all of the four (with one exception) build on existing efforts and do not require external (i.e., non-Caltrans stakeholders) to be successful. The exception, a long-distance digitized freight-corridor, is included because of the extensive work already performed, including stakeholder engagement along I-80.

The remainder of this section provides descriptions for each of the thrust areas and their attendant use cases, including whether it is a broad or narrow use case, who the relevant stakeholders are, and what the value proposition is. Note that some use cases are broader and

bigger than others, and would require different levels of effort to implement. Both the thrusts and the use cases were selected in line with the “Overall Approach” detailed earlier.

Table 1: Overview of DTI Research Thrusts and Use Cases

Research Thrust	Use Case	Existing Efforts	External Stakeholders
Holistic Caltrans Data Platform(s)	***Caltrans-Only Data Platform***	Yes	No
	Data Fusion with Non-Caltrans Data	Limited	Yes
	ITS Data Standardization: C-V2X, Advanced Sensing, Etc.	Yes	Yes
Statewide and Cross Regional Traffic Modeling	***Developing a California Statewide Mobiliti Model***	Yes	No (initially)
	Statewide Mobiliti Model as a Core Data Platform	Yes	No (initially)
	Large-Scale Traffic Disturbances and Response Planning Modeling	Limited	Yes
	Detailed Statewide Modeling of Traveler Behavior in Response to New Infrastructure and Policies	No	Yes
DTI-Enabled AI, Digital Twins, and Multi-Level Models	***DTI-Enabled AI and Big Data Analytics***	Limited	Yes
	DTI-Enabled Digital Twins and Other Models	Limited	Yes
Disruptions and Work Zones	***AI-Enabled Work Zone and Incident Data Collection and Standardization***	Yes	No (initially)
	Development of Real-Time Safety Measures/Tools	Limited	No (initially)
	Integration of Safety Measures into Public and Private Fleets	No	Yes
Connected and Digitized Data Platforms	***Long-Distance Digitized Freight Corridor***	Yes	Yes
	Cloud-Based Digitized Corridor Data Platform	Limited	Application Dependent
	Connectivity-Enabled Partial Automation	Yes	Application Dependent

Holistic Caltrans Data Platform(s)

Like with any “hot topic” in transportation, there is a tendency for practitioners to imagine running before they can walk. It is exciting to talk about digital twins, automation, and all the things that could be possible if all transportation data were in the cloud and accessible to everyone. However, any successful DTI will fail unless it is built on a strong base, and that is the focus of this thrust area. More important than anything else, the first step in creating a Caltrans-led DTI is to create data platforms for existing Caltrans data, so that as much data as possible is stored and labeled so that it is accessible to anyone in Caltrans. This is a largely unappreciated challenge that also relies on the development of processes to create a data-informed culture[7]:

- Documenting domain specific business processes and their associated data use
- Building domain ontologies and data dictionaries for the datasets used in the business processes
- Establishing a Caltrans data registry with links to enterprise data catalogs

Once the existing data is “digitized” then the platform can be used to integrate new data sets, as well as creating mechanisms that said data could be accessed by outsiders (e.g., researchers).

- ***Caltrans-Only Data Platform:***
    - **Description:** This use case represents a broad category to aggregate and standardize Caltrans data, and not everything can be done at once. At a high-level, Caltrans should work to aggregate and fuse all of its existing data, storing it on a standard platform that can be accessed on demand. Many questions exist here, such as the data storage mechanism (cloud vs. server), mechanism for discovery and access, how long to store historical data, etc. As such, the first step should be to create an integrated vision for linkages between newly collected (including real-time) and legacy traffic data (HPMS, Census, PeMS, WIM, classification, LCS, etc.,) prioritizing which data sets should be added to the platform first. PeMS stands out as an attractive option, if only because it was such a notable early digital infrastructure win for Caltrans. This effort will include digitization of infrastructure assets (e.g., roads, curbs, bridges, signage, controllers, sensors, cameras, etc.,) as well as mobile assets (e.g., fleet vehicles such as snowplows, maintenance vehicles, etc.)
    - **Deployment:** Broad
    - **Stakeholders:** Caltrans HQ and districts
    - **Value Proposition:** Perhaps most importantly, every one of the following use cases will rely on a viable Caltrans data platform to some extent. However, the platform, even if incomplete, will still have value on its own. It will ease Caltrans decision-making, enable better comparisons between districts, and help guide

[7]Jane Macfarlane, Anthony Patire, Rick Wagner, Viswanath Nandigam, Caltrans Strategic Plan for Data Collection, Final Report, Caltrans, June 2022. https://dot.ca.gov/-/media/dot-media/programs/research-innovation-system-information/documents/final-reports/ca22-task3834-dor-guidance.pdf

the development of transportation digital infrastructure in California to facilitate data discovery and data exchange.

- ***Data Fusion with Non-Caltrans Data***
    - **Description:** For any Caltrans data that is currently stored on a platform, or for any more comprehensive platform in the future, such as described above, the opportunity exists to fuse with external data sets. Obviously, this would require the buy-in of the relevant stakeholders (i.e., external data owners). Starting with other public agencies is likely the easiest first step, such as creating a statewide database of all traffic signal timing. Fusing private data is also possible, such as combining INRIX and Caltrans traffic flow data.
    - **Deployment:** Broad
    - **Stakeholders:** Caltrans HQ, Caltrans districts, and public/private data owners
    - **Value Proposition:** There are two broad values for data fusion. For the public data fusion, the value is similar to the Caltrans-only data platform. Just as it would be better for all district data to be stored on the same platform and be able to be considered together, it would benefit both Caltrans and other users to have relevant data shared between it and other IOOs. The other key value comes from fusing heterogeneous data sets, such as from the private sector. This fusion could allow for more accurate and reliable information. For example, such private sector data could help fill in for gaps in existing Caltrans data, where sensors are unreliable or non-existent.

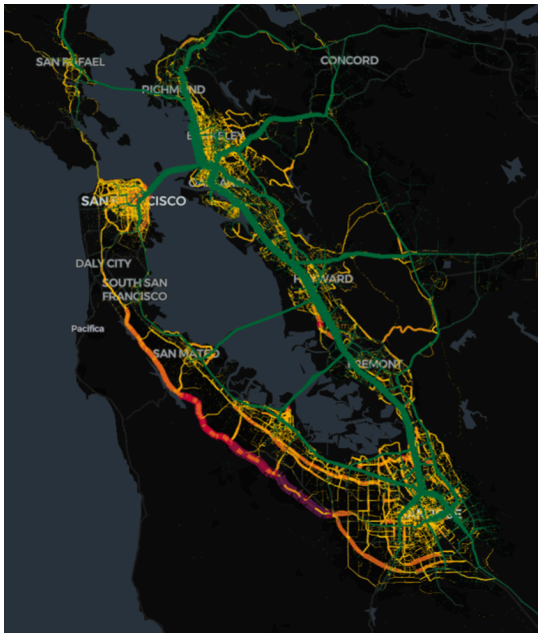
- ***ITS Data Standardization: C-V2X, Advanced Sensing, Etc.***
    - **Description:** As described earlier, Caltrans has an opportunity to be a leader in the ITS space, ensuring that new and emerging data collection and generation technologies can be integrated into a holistic platform. One key part of this is the broader field of connectivity, where such a platform will help promote C-V2X interoperability, especially across multiple vendors and among multiple states. First and foremost, this platform should coordinate efforts across the state, so that the same data platform is used in both D4 and in D12. Standardization will be especially helpful with advanced sensing technologies, identifying how to combine overlapping technologies (e.g., cameras, lidar, vibration detectors, etc.) where there are many different vendors and approaches.
    - **Deployment:** Narrow (to start)
    - **Stakeholders:** Caltrans HQ, Caltrans districts, ITS vendors/providers, other state DOTs, and other key ITS leaders (e.g., ITS America)
    - **Value Proposition:** This ITS data standardization will provide a number of benefits. First, it will reduce costs and difficulties as California continues to adopt ITS technologies, promoting interoperability and preventing the need to build multiple, redundant systems. Similarly, this standardization will show how Caltrans can best invest in future sensing technology resources (e.g., do we need to replace ALL loop detectors). Second, it will help promote California as a leader in the space and enable it to coordinate with other states who are leaders

(Michigan and Texas stand out for CAV work). Third, it will help support a robust research infrastructure for ITS, allowing Caltrans and partners to assess technologies and methods for a range of situations and use cases (e.g., where virtual RSUs vs. physical RSUs are merited). It will similarly support the testing of new mobility solutions (e.g., more data to assist in the training of AV systems).

Statewide and Cross-Regional Traffic Modeling

All of the below use cases will rely on Mobiliti, with some explicitly integrating AI as well (generative and otherwise). The base for all use cases is a statewide or cross-regional model, which will ingest regional travel demand models from individual MPOs, refining the state-wide model maintained by Caltrans as appropriate. This will enable contextual refinement of regional TDMs. The Mobiliti models (either statewide or cross-regional) can be run quickly and accurately. Developing this statewide model can be considered a use case itself, and it is followed by more specific use case descriptions.

- ***Developing a California Statewide Mobiliti Model***
    - **Description:** This statewide Mobiliti model will allow for the modeling of travel throughout all of California, including mode choice components, route choice, congestion and flow estimates, and much more. See the above section on Mobiliti for more details.
    - **Deployment:** Broad
    - **Stakeholders*:*** Caltrans HQ, Caltrans districts, and potentially MPOs (though Mobiliti has already ingested TDMs from many California MPOs: SFCTA, SCAG, SACOG, etc.)
    - **Value Proposition:** Gain holistic insights into freight and long-distance travel with a detailed state-wide transportation model, so that state goals for sustainability targets can be achieved. Similarly, a subset of the statewide model can be used for cross-regional TDM integration (e.g., SANDAG and SCAG).



- ***Statewide Mobiliti Model as a Core Data Platform***
    - ***Description:*** A validated and calibrated statewide Mobiliti model represents a unique transportation data set, providing a reasonable representation of all trips in the state. Beyond running simulations, as described in the below use cases (e.g., showing what happens during evacuation events), the baseline data

contained in this model can be incredibly useful for a variety of stakeholders, providing a detailed overview of transportation in the state. This use case focuses on maintaining, improving, and making accessible said data in a platform.
    - ***Deployment:*** Broad
    - ***Stakeholders:*** Caltrans HQ and districts, and potentially more, including the private sector.
    - ***Value Proposition:*** Such a core data platform has a wide variety of potential uses, from more traditional to more exploratory. On the traditional side, a statewide view of all trips and incidents could be used to easily create a road risk score, characterizing road segments for risk using historical data while integrating weather, hard braking, and pedestrian activity. Also on the traditional side, this data platform could be of use to non-transportation professionals who are still interested in how people travel through the state. For example, real estate developers would be able to identify what areas have a high share of desired trips to help them plan both commercial and residential projects. On the more exploratory side, the data platform could leverage AI and LLM tools to accelerate exploration and analysis of traffic model outputs.

- ***Large-Scale Traffic Disturbances and Response Planning Modeling***
    - **Description:** Evaluate potential impacts of large-scale traffic disruptions and identify cost-effective means for mitigation.
    - **Deployment:** Broad
    - **Stakeholders:** Caltrans Districts and local IOOs, especially in areas at risk from wildfires, earthquakes, etc.
    - **Value Proposition:** This modeling can identify weaknesses in existing evacuation plans and build improved plans and contingencies so that more lives can be saved during future emergencies. It will help understand non-local ripple effects and unintended consequences, and we could explore possibilities for decision support at scale by reviewing past large-scale disruptions. By simulating in Mobiliti, and comparing with historical data, we can both operate at broader scales (e.g., beyond individual MPO boundaries) and run simulations much more quickly than traditional TDMs.

- ***Detailed Statewide Modeling of Traveler Behavior in Response to New Infrastructure and Policies***
    - **Description:** Mobiliti would be able to quantify the trade-offs and impacts of roadway network changes (e.g., new bike lanes, road diets, contra-flow lanes, etc.), as well as other policies that could impact travel behavior (e.g., road usage charges, various road tolling, transit subsidies, etc.). Traditional TDMs are also capable of doing this, but not at the speed of Mobiliti, and most importantly, not at the state-wide or trans-regional level.
    - **Deployment:** Broad
    - **Stakeholders:** Caltrans Districts, local IOOs, and state government in helping evaluate statewide policies and laws.

- **Value Proposition:** This statewide modeling will allow the relevant stakeholders to better plan for large-scale transportation policies and network changes, quantifying and exploring the tradeoffs of different policies and decisions. One example is exploring trade-offs among alternatives for roadway network changes such as modal prioritization (for transit, bike, etc.) so that realistic goals can be set with confidence. Another is testing the impacts of different pricing/taxing initiatives, such as partly replacing the gas tax with different types of road usage taxes and/or implementing higher registration fees for electric vehicles. A third would be helping to coordinate HOV and equivalent infrastructure solutions at a regional/statewide level and identifying locations in the existing HOV network where enforcement activities and/or policy changes could help meet performance objectives. By understanding possible outcomes in context of the traffic dynamics, key stakeholders can make sound financial investments and avoid unintended consequences.

DTI-Enabled AI, Digital Twins, and Multi-Level Models

A robust DTI, complete with data platforms as described above, are necessary precursors to AI and other modeling work, especially those that are data hungry. The traditional modeling adage of “garbage in - garbage out” applies to AI as much as to traditional models, and all require validated and accurate data. Some tools and models, such as many digital twins, even require real-time data.

- ***DTI-Enabled AI and Big Data Analytics***
  - **Description:** Generative AI in particular requires massive amounts of training data. California is already a home to much of the world-leading AI research institutions and companies. Caltrans itself has also prioritized generative AI as a powerful tool, and robust DTI will enable it to reach these goals. The ways that DTI will enable AI or other big data analytics will depend on the use cases. For example, better prediction of traffic flow and congestion would obviously require statewide traffic flow data.
  - **Deployment:** Broad
  - **Stakeholders:** Dependent on use case
  - **Value Proposition:** The end value depends on the AI or analytics tool. The value proposition of this use case is to avoid a negative outcome; it will be difficult to build a successful AI tool on top of Caltrans data WITHOUT a robust DTI.

- ***DTI-Enabled Digital Twins and Other Models***
  - **Description:** Digital twins are effectively simulations that integrate real-word environments and data. For example, a digital twin of a C-V2X corridor could include a traffic simulator (such as in AIMSUN) that is calibrated to the real corridor (such as the El Camino Real Testbed). By sending [near] real-time data from the corridor to the simulation tool, this digital twin will allow the testing of a

variety of scenarios. This is especially useful for testing edge cases of safety-critical events (such as near misses), combining the use of real-world data without risking actual collisions. DTI will also help support other modeling tools, including multi-level and cross-disciplinary models. For example, a statewide Mobiliti model could be integrated with transportation-adjacent models involving weather, energy footprint, electrical grid, emissions, air quality, etc.
    - **Deployment:** Narrow
    - **Stakeholders:** Dependent on use case
    - **Value Proposition:** As with the above, a robust DTI will act as a necessary precursor to the development of digital twins, and will make them easier to develop from a standardized platform. It can support other modeling efforts as well, increasing the predictive capabilities of multi-level and cross-disciplinary models.

Disruptions and Work Zones

Disruptions such as incidents, work zones, and weather events are examples of exceptions (sometimes called “edge cases” in software engineering) where typical road markings and lane designations may be overridden by other signage or by personnel to facilitate emergency response, construction or maintenance activities. Not only are disruptions and work zones difficult for AVs to navigate, but they are also associated with higher incidents and crash rates among human drivers. Therefore, efforts to digitize, monitor, and connect these events using DTI to increase safety can directly benefit Caltrans and its work crews: they are practical, achievable in the near-term, and likely to yield substantial benefits. As such, they represent a good category for early wins. Note that unlike the above group of use cases, these can initially be developed and deployed in a narrow context, such as a certain set of work zones in a given area. However, for all such narrow deployments, it will be important to develop a platform and data standards to enable broader deployment.

- ***AI-Enabled Work Zone and Incident Data Collection and Standardization***
    - **Description:** As with many other use cases, this one has two components: data collection and standardization in a platform. On the data collection side, Caltrans would start by using cameras and equivalent sensors (e.g., radar/lidar) that are trained on high-priority areas, such as work-zones or high-risk intersections. It would then evaluate the sensor outputs using AI tools (especially those that are commercially available) to create and standardize work zone and incident data. This will include using AI-tools to identify not just collisions from the sensor data, but also near-misses and other safety-risks (e.g., harsh braking). In the near-term, Caltrans can focus on only Caltrans collected data, but we will want the platform and standards to be able to integrate third party data as well, such as OEM analytics, collected by companies like INRIX, and others.
    - **Deployment:** Narrow

- **Stakeholders:** Caltrans HQ, Caltrans districts, emergency vehicle operators, and other large fleet operators (especially in the public sector).
- **Value Proposition:** Simply standardizing and improving the collection and accessibility of safety data in high-risk areas like work zones has the potential to save lives, especially those at risk such as Caltrans crews. As such, this use case has the benefit of promising real-value even for early stage deployments, such as a handful of initial work zones and/or intersections with known high incidence of collisions and existing sensor availability. As we collect more data, from both the public and private sector, the benefits will increase. We can more and more accurately identify safety/risk patterns, helping to synthesize guidance and best practices.

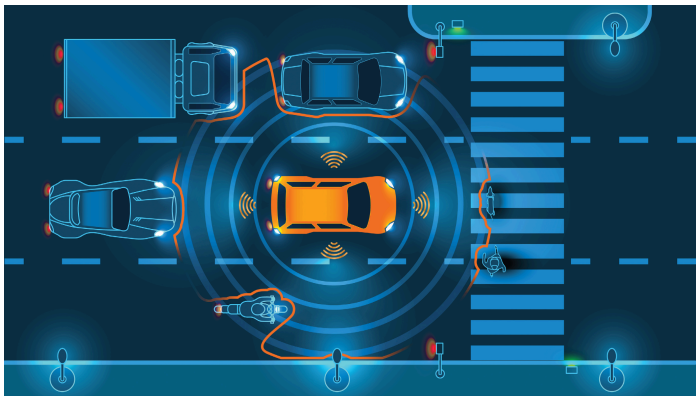
- ***Development of Real-Time Safety Measures/Tools***
  - **Description:** This use case will focus on using existing data to provide real-time safety measures and messages in high risk locations. As with the above use case, we believe that an initial focus on work zones is a good starting point. Such tool(s) would allow for the sending and receiving of messages between Caltrans fleet vehicles, equipment, and variable message signs to maintain and communicate situational awareness. Of special interest here is the ability for automatic WZDx message generation. The data could also be made available to third parties, such as municipal fleets or other data companies.
  - **Deployment:** Narrow
  - **Stakeholders:** Caltrans HQ, Caltrans districts, and potentially public fleets and third party data aggregators
  - **Value Proposition:** Real-time messages can help prevent incidents at work zones and other high risk locations, which is the highest priority. This use case will also help digitize the work-force, standardizing some tasks and enabling better analytics, especially in helping to better understand incidents when they occur. For example, it would ensure that a newly arrived crew at a work zone got a standard message before arriving each time. In concert with the data platform described in the above use case, we could also quantitatively evaluate the measures and messages that are most effective. For example, if we saw a given work zone had too many high-speed vehicles passing by, a new measure might be to recommend that cones be set out further upstream of the work zone for equivalent cases.

- ***Integration of Safety Measures into Public and Private Fleets***
  - **Description:** While getting individual drivers to commit to safety initiatives can be a challenge, private fleets can be more amenable (e.g., they can accept the argument if it reduces long term business costs), public fleets more so, and Caltrans fleets most of all. Therefore, this use case addresses how safety measures can be better integrated into fleets, both for data collection and data use. For example, emergency and other public fleet vehicles can have enhanced sensors (lidar, cameras) installed, if they are not there already, to automatically

log traffic incident location, provide contextual information to 911, ER doctors, etc.
    - **Deployment:** Narrow
    - **Stakeholders:** Caltrans HQ, public and private fleets
    - **Value Proposition:** Vehicle-based data collection represents one of the biggest opportunities for DTI. While we strongly believe that safety should be the initial focus, a fleet collecting data on public roads could create a wide variety of benefits, such as roadway condition and infrastructure monitoring, better estimates of traffic flow and counts, etc. Moving from fixed sensors (e.g., cameras mounted at intersections) to movable sensors (i.e., on a vehicle) open up a world of possibility to bring advanced data collection to places that lack it, such as rural communities. Long-term, coordination with the private sector may make sense (e.g., OEMs and AV companies are already collecting massive amounts of data). However, by using their own fleets, Caltrans can get a head start on vehicles as a sensor.

Connected and Digitized Data Platforms

The below use could be used by all vehicles, but they will prioritize freight, transit, and other fleet operations. The first two focus on crowded corridors, where freight, transit, and fleet operators face similar challenges: tight schedules, hours-of-service regulations, etc. For all potential use cases, the data platforms can provide real time information on road hazards, weather, and unanticipated traffic congestion, which are examples of challenges that may be mitigated through better traffic predictions, dynamic routing, or optimized departure timing. Public-private partnerships and DTI could enable new solutions to improve traffic for the traveling public, and boost profitability and economic output.



- ***Long-Distance Digitized Freight Corridor***
    - **Description:** This would be an example of standardized federation of data, potentially across state boundaries, that would be used by freight operators and related actors. Important in creating this platform would be developing guidance for standards, specifications, and data exchanges. Building such a platform would be greatly assisted by getting the feedback from various stakeholders, identifying their key needs and barriers.
    - **Deployment:** Narrow (but long distance)

    - **Stakeholders:** Caltrans HQ, Caltrans districts, neighboring state DOTs, logistics/trucking companies and related actors
    - **Value Proposition:** A digitized freight corridor could build off of existing Caltrans work (such as along I-80 or along the C-V2X deployment in D12), helping to realize the safety, productivity, and economic benefit of better data accessibility. Users would have access to information such as weather conditions, available parking and charging, current traffic conditions, better estimates of arrival time, alternate suggested routes, work zones or other incidents, and more. In addition to building off existing efforts, other attractive locations could include places with high levels of intermodal transfers, such as near ports and in the Inland Empire.

- ***Cloud-Based Digitized Corridor Data Platform***
    - **Description:** Such a platform could be used by both transit and freight, although the initial development would likely prioritize one or the other. It could even support the above long-distance corridor. What makes this use case unique is the focus on cloud-based data, such as that collected by C-V2X technologies, virtual RSUs, cell phones, and more. Note that this use case would likely build off of existing data platforms as described in earlier sections, that utilize traditional sensor data.
    - **Deployment:** Broad or narrow (depending on application)
    - **Stakeholders:** TBD (based on applications)
    - **Value Proposition:** This use case is more exploratory than the above one, but the core value is finding early applications for integrating cloud data into existing Caltrans data platforms.

- ***Connectivity-Enabled Partial Automation***
    - **Description:** Medium-latency (1-5 seconds) C-V2X can enable new AV solutions by providing granular and timely information about vehicle speeds or traffic signal timing. By linking connected and [partly-] automated vehicles with DTI, we can find benefits from such vehicles at relatively low penetration rates (e.g., <2% of vehicles on the road). For such applications, we strongly recommend focusing on fleet operators (convincing individual travelers on a 1:1 basis is hard) and areas where the fleet represents relatively high percentages of the vehicles on the road, such as a freight fleet along a drayage route. Working with Caltrans fleets makes it even easier, reducing the stakeholder count. We recommend two specific applications that could meet these standards, but there are others to explore as well:
        - *CIRCLES 2.0: Collaborative Traffic Smoothing:* This application uses partially automated vehicles that demonstrate traffic smoothing via longitudinal control using real-time traffic flow data from sensors and participating vehicles. Individual efficiency benefits occur at 0% vehicle penetration rate, and system-level efficiency and congestion benefits at 3-5% vehicle penetration rates.

        - *Eco-Driving Along Arterials:* Like CIRCLES 2.0, this application smooths traffic via longitudinal control, but along signalized arterials. Vehicles adjust their speed based on real-time traffic signal timing (e.g., slow down if the light will turn red before they arrive. Individual efficiency benefits occur at 0% vehicle penetration rate, and system benefits at 5-10% penetration rates. Freight and other heavy-duty vehicles are especially attractive considering the high energy costs of acceleration.
    - **Deployment:** Narrow
    - **Stakeholders:** TBD (based on applications)
    - **Value Proposition:** This use case can provide congestion mitigation and energy savings for both the vehicles themselves and the system overall, even at relatively low CAV penetration rates.

Special Use Cases

Additional opportunities abound, and may be further developed or combined with the ideas listed above. These areas are less developed, and the below three are just a sample. These are intended to show the broad potential reach of DTI, but PATH is not making any concrete recommendation on next steps to Caltrans in these areas.

- *New Mobility and Land Use*
    - Very broad area, but basically aims to leverage new transportation technologies with land use.
    - For example, a mobility hub could create non-private vehicle options, so it could be paired with areas that have fewer parking minimums and denser zoning allowances. There is some overlap here with transit-oriented development.
    - We could build the transportation infrastructure and land use infrastructure at the same time.
    - Real estate firms (and related interests) could be key stakeholders and enablers, especially from a monetary perspective.
    - Public-private partnerships are options
- *EV Conversion of Caltrans Fleets*
    - Build upon existing efforts in progress
    - Understanding how Caltrans can electrify/ZEV their fleet will provide guidance for others.
- *AI Assistant for Paratransit*
    - Expand ongoing SMART project with CCTA, using AI to support paratransit decision making and booking, which is currently done only via inefficient, book-in-advance phone calls.

Key Recommendations

To reiterate, the overall approach taken by this research roadmap has the following goals:

- Identify low-hanging fruit and early wins
- Aim for real results within five years
- Discover opportunities to obtain high-quality, state-wide data
- Propose limited-scope efforts in high-value situations
- Define a clear customer for the research within Caltrans
- Identify opportunities that minimize organizational complexities and misaligned incentives, especially where Caltrans does not require external stakeholder involvement

A few areas stand out. Freight is increasingly seen as one high-value opportunity area, where application of recent technological improvements may yield substantial impacts. The same is true for fleet operations, especially for Caltrans fleets. The balkanization of data is another critical problem, in multiple dimensions: within Caltrans, between different public agencies and Caltrans, and between Caltrans and the private sector

More important than anything else is to start by building a solid DTI base within Caltrans. With that in mind, while all of the above research thrusts and use cases were selected with these objectives in mind, a few specific use cases stand out as what Caltrans should focus on now. We recommend one use case as the absolute highest priority, and then the following three as falling one tier below.

Caltrans-Only Data Platform

This is the most important step that Caltrans can possibly take. As described earlier, building such a platform is a large, fundamental effort. As it stands now, much Caltrans data is stored in different formats, held by different divisions and districts, so that simply accessing and validating data can easily turn into a research project itself. For example, districts have ATMS data, which is within their scope of operations; they do not have data access problems themselves. However, the districts don't naturally think about integrating this data with other districts, states, etc. A Caltrans-Only Data platform is needed to integrate and standardize this and other data. Within Caltrans, we can point to Cal-ITP as doing good work in publishing and sharing standardized transit across California; we can follow this example in building a larger-scale Caltrans Data Platform that will be the envy of other states. Just as transportation professionals throughout the world once flocked to PeMS data since it was such a unique and valuable statewide data set when it first came out, we expect that a new Caltrans Data Platform will have a similar impact.

While we may not see a fully-finished platform soon, not everything has to happen all at once, and we can see real benefits in the near term. A reasonable, organizational-wide approach would be to simply define what data can and should be included in this data platform, and when. PeMS 2.0 is an attractive first step, including a network of roads and key assets in a map format

and vehicle data (e.g., counts) across time (including daily, weekly, monthly, and yearly variations). Infrastructure data, such as traffic signal timing, may come next, but this might be a somewhat more intense effort**.** Critically, this should be a standardized, statewide effort, just like PeMS, that is accessible to anyone within Caltrans and can be shared with other partners and stakeholders easily and as appropriate; the outward facing data platform might be a narrower-version of the Caltrans-facing version.

Statewide Mobiliti Model as a Core Data Platform

The fast, regional-scale, dynamic traffic simulation capabilities of Mobiliti are new and exceptional. Its sheer speed and scale make it uniquely suited for a new level of large-scale disturbance/evacuation events modeling and mode-shift analysis. However, none of these benefits can be realized unless a statewide model is finished, and the work is mostly completed (i.e., approximately half the TDMs have already been ingested). We recommend the first step of using the statewide model as a core data platform, and working with various Caltrans stakeholders to identify the most impactful future steps and research directions. Just this data platform itself would be a unique tool for researchers the world over to help solve California's transportation challenges.

AI-Enabled Work Zone and Incident Data Collection and Standardization

This is the narrowest of the four priority use cases recommended here, but it is also the one with the highest potential impact. Safety is always the highest priority for Caltrans, and anything that we can do to save lives, the lives of travelers and Caltrans employees must be considered. Moreover, traffic safety is a famously data deficient area; actual collisions on an intersection-by-intersection basis are fairly rare and fatalities even rarer. The baseline data set are police reports, which can be incomplete or inaccurate. By both standardizing work zone and incident data, as well as collecting more data from novel means (e.g., setting up cameras at work zones to detect near-misses), Caltrans has the ability to save lives and provide an invaluable traffic safety platform for transportation professionals across the world.

Long-Distance Digitized Freight Corridor

Data about truck parking availability, weather conditions, traffic delays, work zones, etc. are not uniformly and consistently kept up-to-date. A digitized freight corridor will lay the groundwork for efforts to improve the data quality, make it usable for logistics companies, and to facilitate future interstate collaboration. This basic groundwork must be implemented before it will be possible to achieve state and national goals for improved safety, and efficiency. This digitized corridor can also build off of existing Caltrans efforts, so I-80 is an attractive initial option.