Request for Information (RFI)

Title: Advancing data systems in support of innovative data analytics to improve safety-related policy, information, and decision making

1. **Introduction:**

This is a request for information (RFI) only, and is NOT a solicitation for proposals, proposal abstracts, or quotations. This RFI notice is for information purposes only and does not obligate the government in any way. The Government will not reimburse the respondents for any costs associated with the information submitted in response to this request. Information received in response to this RFI will be safeguarded adequately from unauthorized disclosure. The Government will treat each submission as confidential when requested by the submitter, and will not disclose such submission outside of the Government except as permitted by law.

Information gathered through public comment will inform decisions on whether and how to proceed with advancing the data systems and innovative data analysis the U.S. Department of Transportation (USDOT) needs to improve safety-related policy, information, and decision making. Respondents are invited to comment on any or all of the items described in this RFI.

To respond to any questions related to the RFI, USDOT is hosting a webinar for potential respondents on November 15th, 2018 from 2pm to 4pm Eastern Time. In order to register to attend the webinar, further information will be provided at the Safety Data Initiative website: https://www.transportation.gov/policy/transportation-policy/safety/safetydatainitiative.

2. **Background:**

Safety Data Initiative

At the 2018 Transportation Research Board Conference, USDOT announced the Safety Data Initiative (SDI)[1]. The SDI is led by offices within the Office of the Secretary of Transportation (OST) in close collaboration with the Federal Highway Administration (FHWA), the National Highway Traffic Safety Administration (NHTSA), and the other surface transportation subcomponents or Operating Administrations (OAs) within USDOT. The Safety Data Initiative has a vision, focus areas, strategies, and projects that will help build upon and enhance current surface transportation safety efforts related to data, analysis, and policy making.

Vision: Evolve from retrospective to predictive analysis that is transformed into compelling visualizations of insights to better target safety risk. Develop an integrated data eco-system for rapid, rigorous, and innovative safety data analysis and insights using new datasets and new analytic tools across surface transportation modes, accessible to decision-makers to support policy decisions. Be a leader in Safety Risk Management by working with transportation stakeholders and leveraging innovative data sources and practices.

[1] https://www.transportation.gov/briefing-room/dot0318

We will strategically prioritize and address transportation safety risks through data-informed decision making, with a focus on:

1) **Data visualization**: Make data analysis and insights accessible to policy-makers through clear, compelling data visualizations.

2) **Data integration**: Integrate existing USDOT databases and new private sector data sources to answer safety questions.

3) **Predictive insights**: Use advanced analytic techniques to identify risk patterns and develop insights that anticipate and mitigate safety risk to reduce injuries and fatalities.

In order to achieve this vision, we are pursuing three strategies:

I. Build USDOT's capacity to perform data analysis for policy and decision making based on risk and predictive insights;

II. Establish data integration inside USDOT and through collaboration with other agencies and entities to create data connections and integration; and

III. Promote the innovative use of safety data and visualization among traditional and non-traditional stakeholders to turn data into useful information for continuous safety improvement.

Safety Data Initiative Projects:

1. Solving for Safety: Visualization Challenge

DOT has launched a safety challenge asking participants to come up with innovative ways to visualize data that will reveal insights into serious crashes on our roads and rail systems while improving our understanding of transportation safety. For more information, visit the Solving for Safety page[2] or read the press release[3] announcing the challenge. In October five semi-finalists were chosen to develop a proof of concept visualization.

2. Waze Pilot

DOT's Volpe National Transportation Systems Center ("the Volpe Center") is leading a pilot project exploring the opportunity to estimate police-reported traffic crashes in near-real time by combining crowdsourced crash data from Waze with crash data provided by the State of Maryland via the National Highway Traffic Safety Administration's (NHTSA) Electronic Data

[2] US Department of Transportation website. https://www.transportation.gov/solve4safety.
[3] *"DOT Announces Competition on Advancing Innovative Ways to Analyze Crash Data."* US Department of Transportation website. https://www.transportation.gov/briefing-room/dot-announces-competition-advancing-innovative-ways-analyze-crash-data

Transfer pilot. The Volpe Center employed machine learning techniques with these datasets to train statistical models to predict crashes. In this pilot, DOT learned these models supported with Waze data produce reasonably good estimates of police-reported crashes. This pilot has laid the foundation needed for a future nationwide scale-up of a crash count tool, and was recently highlighted in the Washington Post[4].

3. Rural Speed Pilot

The rural speed pilot is an ongoing research effort to understand the contribution of prevailing speed, speed limit, and average travel speed to the prevalence and severity of crashes on rural highways. The pilot further seeks to understand the relationship roadway design and traffic volumes have with speed and crash outcomes. The Federal Highway Administration (FHWA) is using the National Performance Management Research Data Set (NPMRDS) – anonymized data from GPS-enabled devices – to learn about speed's role in crashes. The NPMRDS provides prevailing speeds at 5-minute intervals across the entire National Highway System. When combined with traditional datasets, the NPMRDS data can provide a closer look at speed's role in crashes.

4. Pedestrian Fatalities Pilot

The pedestrian fatalities pilot sought to understand the relationship pedestrian fatalities may have with transportation system and built environment characteristics. Two key takeaways were discovered through analysis of data from FHWA, NHTSA, the Environmental Protection Agency (EPA), and the U.S. Census Bureau. In urban areas, traffic on non-access controlled arterials was found to significantly increase pedestrian fatality risk. Traffic on other urban roadways and all roadway types in rural areas also contributed to pedestrian fatality risk, but with weaker effects. Additionally, employment density in the retail sector was strongly associated with increased pedestrian fatality risk in both urban and rural areas. Lessons learned from this pilot may be used to understand place-specific risks. The work from this project was published in the December 2018 edition of *Accident Analysis & Prevention*.

5. Fatality Analysis Reporting System Data Visualization

NHTSA is experimenting with the presentation of its Fatality Analysis Reporting System (FARS) data – a nationwide census of fatal injuries suffered in motor vehicle crashes – to supplement existing data summaries on specific topical areas. NHTSA is in the process of beta testing an interactive visualization of the 2016 Traffic Fact Sheet focused on speeding using visualization software. By creating more interactive information, the hope is to present the data in a new way that may be helpful to policy-makers and the general public. The prototype visual is located at https://www.nhtsa.gov/nhtsa/fars/speeding_data_visualization.

[4] *"Federal researchers are using data from Waze and Maryland to try to predict road dangers."* Michael Laris, Washington Post, July 7 2018. https://www.washingtonpost.com/local/trafficandcommuting/federal-researchers-are-using-data-from-waze-and-maryland-to-try-to-predict-road-dangers/2018/07/07/bb40b51c-7ec9-11e8-bb6b-c1cb691f1402_story.html?utm_term=.30a1678a3e92

Current State of Transportation Safety and USDOT Safety Authority

In 2016, we suffered 37,806 traffic fatalities on U.S. roadways with around 2.5 million estimated nonfatal injuries. There were another 846 fatalities in rail, vessel, and pipeline modes of transportation. Traffic fatalities increased in both 2015 and 2016, erasing years of traffic safety gains. Traffic fatalities for 2017 declined to 37,133 but remain a leading cause of unintentional injury death in the United States. USDOT is strongly committed to continuing to reduce the incidence of serious and fatal injuries across our surface transportation network. To do this, the Department is taking a systematic approach to safety that uses data to identify risk patterns, prioritize risks, and develop effective solutions. As part of a systemic safety approach USDOT is working to improve the collection, management, integration, analysis, and use of data within and across the Department's OAs, and we are developing advanced analytic tools to better quantitatively identify the risk factors that contribute to transportation-related fatalities and serious injuries.

Within USDOT, multiple OAs have regulatory safety jurisdiction and defined safety responsibilities that require the collection, management, integration, analysis, and visualization of data. Within OAs, different offices execute safety activities to further USDOT's commitment to a safe and efficient transportation system. Beyond USDOT, a number of stakeholders play a role in the collection, management, integration, analysis, and visualization of transportation and safety data within the surface transportation ecosystem. This includes other Federal government agencies, State, regional and local governments, researchers and universities, operators in the transportation system, transportation-focused organizations and associations, vendors involved in transportation data contracts and activities, and the public more generally. Overall transportation safety has a multitude of stakeholders with varied focus areas, and transportation risk management is increasingly a multi-modal challenge that requires action across traditional organizational boundaries.

Safety data are commonly used to obtain a better understanding of safety issues and risks, inform policy and decision making, and execute a data-driven approach using evidence derived from data. Traditionally, available data sets included a mix of administrative data and surveys, of which many data are de-identified or aggregated and require sophisticated techniques for data integration/fusion. Analysis of data for safety purposes in the Department has been primarily retrospective, of high quality, and usually issued annually.

The rise of new technologies, capabilities, and tools over the past few years has both changed the ways the transportation community can use data, and enabled new data opportunities that were unavailable just a few years ago. The proliferation of data from new sources, and more timely contemporaneous, "big data" provide new opportunities for analysis and insights and more rapid feedback from the operating environment, but also pose new challenges in terms of computing power, statistical and analytical capabilities, and data access. The future of vehicle automation presents a potential new source of safety data; however, this initiative is focused on leveraging existing data sources and processes. The USDOT's work related to automation can be found in

Automated Vehicles 3.0: Preparing for the Future and on our Automated Vehicle Research Web page.

Current State of the USDOT Information Technology (IT) Environment for Safety Data Analysis

The USDOT IT environment is de-centralized and, as a result, data sources and methods that support innovative analyses are compartmentalized across the Department. USDOT's Operating Administrations employ a wide range of technologies for the collection and aggregation of surface transportation data related to roadways, trucks and passenger vehicles, public transportation, non-motorized transportation, rail transportation (freight and passenger), and pipelines. Data are collected to advance safety, monitor system performance, track funding and grants, and other purposes.

Systems are distributed among on-site data centers as well as the Microsoft Azure and Amazon Web Services clouds. Furthermore, the Department employs a range of database technologies (though these are primarily Oracle and Microsoft SQL server) and data warehousing/business intelligence technologies (including, but not limited to: SAP Business Objects, IBM Cognos, Oracle Business Intelligence, SAS Business Intelligence, Microstrategy, Tableau). Data and relevant contextual data are broad in scope and examples include safety outcome data from the FARS (https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars), transportation infrastructure data such as the Highway Performance Monitoring System (https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm), or integrated data such as the Highway Safety Information System (https://www.hsisinfo.org). Many of the Department's key datasets are considered part of the National Spatial Data Infrastructure and are also geospatial in nature. Linear referencing systems are commonly used for roadway networks. The Department's core technology for geospatial data production and analysis is the Esri ArcGIS suite. Emerging tools in the geospatial data integration space include the 1Spatial 1Integrate product.

In addition to the USDOT's primary data sources (which can be perused at https://catalog.data.gov/dataset?organization_type=Federal+Government&organization=dot-gov), data from other Federal agencies are also relevant to safety data analysis. Representative datasets may include weather, demographic, economic indicators, employment, and more. Such datasets are available from other Federal agencies in a variety of formats and disseminated through a wide range of techniques. These include, but are not limited to: flat file downloads, application programming interfaces (APIs), and cloud-native offerings through public dataset programs (as is the case for weather data).

The Department has begun to collect and archive private sector data for a variety of purposes. For example, the Department maintains an archive of National data made available through the Waze Connected Citizens Program (https://www.waze.com/ccp/) and probe data used in support of calculating national highway performance measures (see tutorials at: https://npmrds.ritis.org/analytics/tutorials/). USDOT also envisions remotely accessing and

analyzing data that are curated by third parties, and encouraging voluntary exchange of safety-relevant data. For example, we are helping State DOTs make harmonized work zone data available in a federated model (see: https://www.transportation.gov/av/data).

The Waze CCP dataset is stored in the Amazon Web Services cloud in a proof-of-concept USDOT system called the Secure Data Commons (SDC). (see: https://www.its.dot.gov/presentations/2017/Waze_Partner_Summit20171017.pdf). The SDC enables collaborative but controlled integration and analysis of data at the moderate sensitivity level and is based on successful models in the health data space (see https://cbiit.cancer.gov/ncip/crdc-cloud-resources).

The Department is exploring additional solutions to cut across its data silos and to enable the use of modern data science tools in support of the Safety Data Initiative. These include the application of open source tools (e.g. R, Python) and notebook-oriented data analytics (e.g. Jupyter notebooks). Many of the Department's data sets are feature-rich but are not so big that they cannot be analyzed in memory. However, with increased use of real-time data and geographic information for safety data analysis, increased computing power and storage capacity are becoming critical needs. As such, the Department is moving ahead and establishing a "big data" analytics platform that will more broadly enable the type of analysis contemplated under the Safety Data Initiative. A platform may need to support a wide range of analytical applications, from the Department's own cybersecurity data, to the Department's management and accountability data, and data that will be part of the Safety Data Initiative.

Challenges with Safety Data and Analysis

USDOT identified three major challenges that impact our ability to use safety data to inform policy and decision making:

- Data integration is a time-consuming and difficult task, which reduces our ability to achieve quick results
    - Data linkage such as mapping and merging, as well as understanding the business processing from data source to data set, can be arduous
    - Data linkage requires clear documentation of how the derived integrated data were produced from the source data, which can be time consuming
- USDOT needs to significantly augment our capacity with regard to emerging data analytics, data integration, and data visualization to further produce predictive safety insights
- Acquiring new data sources and ingesting, processing, managing, and using new data – especially "big data" – is ad hoc and is dispersed throughout USDOT
    - New data sources require information on the source data, the context of their use, and an understanding of their limitations and biases
    - Management of personally identifiable information (PII), security sensitive, and business confidential data needs to have a clear and controlled distribution model, which may be difficult to automate or better conducted by a third party

Ideal End State:

USDOT has the systems, processes, data, and contract vehicles in place so USDOT can answer pressing policy-oriented research questions by quickly performing analysis through innovative techniques and entities. The data would be accessible, with previous work done to connect the data available for use, and high-quality vendors to perform the analysis would be easy to access.

Current Challenges	Potential Procurement Solutions
Data integration is a time-consuming and difficult task, which reduces our ability to achieve quick results	Data integration processes that permit the repeatable integration of data for analysis. The integrated data will be accessible and useful both internally and externally
USDOT has limited capacity with regard to statistical methods, data analysis, model development, and staff capacity for predictive safety insights	Access to vendors that can perform quick analyses using innovative techniques to answer research questions that broaden our understanding of transportation safety through predictive insights
Acquiring new data sources and ingesting, processing, managing, and using new data – especially "big data" – is ad hoc and is dispersed throughout USDOT	Creation of the necessary data systems and procedures to ingest, process, manage, and use new and more real-time data sources, especially big data

3. Information Request:

This RFI is intended to inform OST on the current status of industry sources, business practices, technical capacity, and operational capability related to data integration, data analysis, and data visualization that could be leveraged to further USDOT's capabilities to use safety-related data to better inform policy and decision making for multimodal surface transportation safety. Furthermore, this RFI is intended to better understand the potential roles and vendor opportunities available in a complex, Federated safety data ecosystem where data collection, management, integration, analysis, and visualization may come from a multitude of entities, including USDOT, other Federal government agencies, State, regional and local governments, researchers and universities, operators in the transportation system, transportation-focused organizations and associations, vendors involved in transportation data contracts and activities, and the public.

It is important to note that this RFI covers a large range of the data lifecycle, and respondents should respond to RFI questions that are relevant to their expertise and capabilities, as appropriate. Please provide concrete examples and applications that detail, clarify, and/or visually display the responses to the questions for each section.

A. Assumptions, Objectives, and Anticipated Deliverables:

If USDOT determines that procurement activities are necessary, there are underlying assumptions and objectives that shape our path forward:

- Data and Tools
    - Enable a more effective, frequent, and rapid exchange of data and information between USDOT and the rest of the transportation community; data connections and linkages will be created and shared amongst the transportation community
    - USDOT will not be the sole source of data, nor the sole entity able to ingest, process, and manage data into usable information
    - Data ownership will vary depending on the data and their use
    - Data integration will build on previous research in order to reduce the time it takes to perform innovative analysis
    - Data governance, use, accessibility, and technology standards that are flexible and appropriate for data with a spectrum of restrictions
    - Balance appropriate data stewardship and access with encouraging open, available, and transparent data

- Task Order Management
    - Sufficient flexibility to enable a multitude of on-demand projects that span scope and purpose, safety activity and safety topic, participants, datasets, and analytical tools used
    - Flexibility for active participation, collaboration, and exchanges between USDOT and other members of the transportation community
    - Effectively manages and leverages complex relationships that may go beyond the traditional government to vendor relationship

- Outcomes
    - Build USDOT's capacity to convert data and research into information that proactively and effectively informs policy and decision making across surface transportation modes and for all users of the system
    - Expand USDOT's capacity to perform prospective analysis focused on risk and predictive insights through a mix of traditional and new, innovative analytical tools and statistical methods
    - Identify data quality and integration challenges, and initiate both top-down and bottom-up solutions to data use barriers
    - Develop and deploy shared data service models with adequate security and privacy protections to foster a robust exchange of information among transportation community members

Anticipated deliverables could include one or a mix of the following:

- Identifying, collecting, and analyzing new safety data, data sets, data systems, and tools
- Integrated, connected, and/or fused data sets from multiple sources that are ready to be used for analysis, which could include data quality control, data cleaning, data

standardization, and readily available and accessible data for individual projects or recurring safety data needs

- Facilitate the establishment of data management and sharing agreements amongst transportation community partners, both with and without USDOT as a direct partner, and vendors for safety data analysis
- Descriptive and statistical analysis using analytical tools and techniques for short and long term information needs
- Conduct exploratory safety analysis on available data sets using analytical tools and techniques
- Complex analysis using a variety of tools, including inferential statistics, modeling and simulations, artificial intelligence and machine learning, and non-statistical approaches
- Data visualizations that translate analysis and research results into useful information
- Data visualizations that permit the exploration of data by both the sponsoring entities and the wider transportation community unaffiliated with the project
- Continual shared services as the lynchpin of a data exchange network over a period of time

Questions related to Assumptions, Objectives, and Anticipated Deliverables

- What frameworks and strategies should USDOT consider based on our focus areas (integration, visualization, and insights), strategies, and ideal end state?
- Are there additional deliverables that USDOT should consider based on our focus areas, strategies, and ideal end state?

B. Data Integration
   - Describe and provide examples of your capabilities and past performance in integrating, connecting, and/or fusing multiple data sources. Describe the data integration methods used, including deterministic and probabilistic linkage techniques, the use of data attributes such as time and space, or other integration methods. If applicable, describe how you have made integrated data accessible and repeatable for others to use in future analyses and projects.
   - Describe your demonstrated approaches to providing services that employ data science platforms, cloud-native data analysis products and services, and ongoing governance/management of such platforms and the resulting analytics.

C. Data Exchange
   - Describe successful models that enable the exchange and use of information amongst a number of contributing parties in an open or controlled-access system, and the role(s) vendors could play in establishing and maintaining such as system.

- Describe and provide examples of your demonstrated approaches to establish and maintain a successful, effective data exchange system. Propose new, innovative approaches to establishing data systems and data exchanges that are flexible and can be adapted over time, if applicable.
- Describe successful voluntary arrangements that are facilitated by vendors and enable effective data integration and analysis between Federal, State, and/or local government entities. Highlight the participant benefits of using third-party vendors for intragovernmental information exchanges.

D. Data Analysis
   - Describe your demonstrated approaches to performing the types of analysis below. If applicable, provide example(s) for each of the four types of analysis.
     1. Inferential statistics. Linear models; non-linear models; spatial regression; Bayesian statistics; econometrics; and survey sampling.
     2. Modeling, simulations, and forecasting. Commodity, freight flow, and market forecasting; probabilistic risk analysis and actuarial sciences; transportation-related modeling, forecasting, and simulations.
     3. Artificial intelligence and machine learning. Big and unstructured data analytics; neural networks and deep learning; qualitative content analysis and language processing; video, photo, and image analysis.
     4. Non-statistical approaches. Geospatial analysis, including analysis that involves linear referencing systems; root cause investigation; content and survey questionnaire design; audio and video digitization and analysis.
     5. Any other type of data analysis that was not mentioned prior that would be applicable to transportation safety analysis.
   - Describe and provide examples of your approach to exploratory analysis of data sets, especially big data sets, to determine their suitability for a particular analytical application or research question. Address your approach for discovering and correcting for bias and other factors that might impact a dataset's fitness for purpose.

E. Data Visualization
   - Describe your data visualization capabilities, including tools and techniques applied. If applicable, describe your experiences visualizing: multiple data sources into a cohesive narrative; very large data sets and "big data" (due to volume, variety, and/or velocity); and advanced analytical techniques described in part D. Provide examples of data visualization performed, if applicable.

F. Data Management
   - Describe and provide examples of your capabilities in ingesting, cleaning, storing, managing, and maintaining data. If applicable, describe how your data management led to a flexible, accessible data environment.

- Describe your demonstrated ability to work with "big data" (primarily high variety data. You may also discuss high-volume and/or high velocity data).
- Describe your experiences in ensuring privacy, business confidentiality, and how security sensitive information is managed in a data system environment used by multiple government, private, and non-profit entities. What systems and protections would need to be in place to build trust and encourage the exchange of information?

G. Partnerships
- Describe and provide examples of successful vendor consortiums for data integration, analysis, and data visualization that leverage the capabilities of multiple vendors to have a holistic set of competencies to perform a wide variety of work.
- Describe successful vendor consortiums that involve a mix of entities, which could include large for-profit companies, small and disadvantaged businesses, non-profit entities, universities, quasi-public entities, research consortiums, and other applicable vendor types.

H. Procurement Method
- Describe government and private sector arrangements used by other government entities that successfully facilitated data analysis, integration, and visualization objectives similar to the descriptions provided above. For example, Broad Purchase Agreements, Indefinite Delivery/Indefinite Quantity contracts, etc.
- Describe and provide examples of vendor experiences with successful business model approaches that would provide mutual value to the vendor and USDOT, and permit the level of on-demand, flexible services for data systems and services described in the RFI.

Instructions for Submissions:

Email Point of Contact: USDOTSafetyDataInitiative@dot.gov

Contracting Office Address:
Department of Transportation (DOT) / Office of the Secretary of Transportation (OST)
Office of the Senior Procurement Executive 1200 New Jersey Avenue, SE
Washington, DC 20590

Submit Responses Electronically: USDOTSafetyDataInitiative@dot.gov
Email Subject Line for Response Submissions: **"USDOT Safety Data RFI Submission"**
Submit Questions Electronically: USDOTSafetyDataInitiative@dot.gov

In addition to submitting questions via email, USDOT is hosting a webinar for potential respondents on **November 15, 2018 from 2pm to 4pm Eastern Time** to allow for questions and clarifications. In order to register to attend the webinar, further information will be provided at the Safety Data Initiative website: https://www.transportation.gov/policy/transportation-policy/safety/safetydatainitiative.

The due date for the submission of responses to the overall RFI are due by no later than December 7, 2018 at 2 pm Eastern Time.

In the email with the RFI submission, please note which section(s) of the Information Request the response is addressing. The Sections with questions are alphabetized and structured "A through H;" for example, "C. Data Exchange" and "G. Partnerships."

DISCLAIMER:
THIS SOURCES SOUGHT ARE FOR INFORMATIONAL PURPOSES ONLY. THIS IS NOT A REQUEST FOR PROPOSAL. IT DOES NOT CONSTITUTE A SOLICITATION AND SHALL NOT BE CONSTRUED AS A COMMITMENT BY THE GOVERNMENT. RESPONSES IN ANY FORM ARE NOT OFFERS AND THE GOVERNMENT IS UNDER NO OBLIGATION TO AWARD A CONTRACT AS A RESULT OF THIS ANNOUNCEMENT. NO FUNDS ARE AVAILABLE TO PAY FOR PREPARATION OF RESPONSES TO THIS ANNOUNCEMENT. ANY INFORMATION SUBMITTED BY RESPONDENTS TO THIS TECHNICAL DESCRIPTION IS STRICTLY VOLUNTARY.