From Data to Knowledge to Action: Enabling a Revolution in New Transportation

Sebastian Thrun — **Stanford University**
Chase Hensel — **Computing Research Assoc.**
Erwin P. Gianchandani — **Computing Research Assoc.**

Version 5: September 11, 2010[1,2]

The U.S. transportation system has provided considerable flexibility, speed, reliability, and efficiency over the past century. It has enabled us to live farther from where we work, to visit family and friends more frequently, and to ensure that essential goods such as food and water may be transported over long distances. However, the system is showing signs of aging, as we force upon it a growing population increasingly on the go. For example, consider the following challenges that have emerged in the past decade:

- **Energy and the environment:** Transportation now consumes more than 28 percent of all U.S. energy[3] and generates roughly the same proportion of CO_2 emissions. (By comparison, U.S. petroleum imports comprise just slightly more – 33 percent – of U.S. energy use.)

- **Efficiency and Productivity:** Congestion in the U.S. is responsible for 3.6 *billion* vehicle hours of delay annually, and today's average daily commute time exceeds one hour. Similarly, over 20 percent of commercial airplane flights since 2003 have been delayed or cancelled[4], and the percentages of delays and cancellations are increasing every year. The European Union, which faces similar problems, estimates that the cost of congestion is about 0.5 percent of its gross domestic product (GDP).[5]

- **Safety:** Over 41,000 people are killed[6] and 2.5 million others injured[7] in highway accidents each year in the U.S. alone. Most of these accidents are due to driver neglect, and are caused by vehicles either leaving the road or traveling unsafely through intersections. However, our highway infrastructure is also showing signs of overuse. For example, in 2007, an eight-lane, steel truss arch bridge that carried Interstate 35W across the Mississippi River in Minneapolis, MN, collapsed, killing 13 people and injuring another 145. The National Highway Transportation Safety Board (NTSB) concluded that a design flaw compounded by an increased load placed on the bridge over time was to blame for the accident[8]. Meanwhile, the Federal Aviation Administration (FAA) and NTSB are tracking a higher incidence of

---

1 Contact: Erwin Gianchandani, Director, Computing Community Consortium (202-266-2936; erwin@cra.org). For the most recent version of this essay, as well as related essays, visit http://www.cra.org/ccc/initiatives.

2 This white paper is adapted from a previous CCC document, "Surface Transportation 3.0" by Sebastian Thrun and Henry Kelly: http://www.cra.org/ccc/docs/init/Surface_Transportation_3.0.pdf.

3 http://en.wikipedia.org/wiki/Energy_in_the_United_States.

4 http://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp.

5 http://www.ertico.com/.

6 http://www-nrd.nhtsa.dot.gov/Pubs/811016.PDF.

7 http://www-nrd.nhtsa.dot.gov/Pubs/811017.PDF.

8 http://ntsb.gov/Recs/letters/2008/H08_1_Design_Adequacy_Report.pdf.

runway incursions (i.e., defined as incidents in which an aircraft inadvertently enters an "active" runway being used for takeoff or landing by another plane); failure to maintain minimum separation standards, particularly during takeoff and landing; and wrong runway usage[9]. For example, in 2006, Comair flight 191 crashed on take-off in Lexington, KY, because the runway the pilots used was much shorter than the one air traffic controllers had assigned the plane[10]. We have also witnessed more accidents in recent years than at any prior time in history on U.S. railroads. For example, the NSTB is currently investigating four accidents in just the past 12 months on the Metrorail subway system in Washington, DC, including a deadly rush-hour collision between two trains attributed to faulty track circuits[11].

- **Equity:** Urban development patterns have forced many low-income families to move to suburban or exurban neighborhoods to find affordable housing. This shift is necessitating long commute times, with few mass transit opportunities and rapidly rising costs. For instance, in the Washington, DC, area, fuel prices combined with hour-plus drive times due to near-gridlock conditions are prompting many suburban residents to consider alternative forms of transportation. But the region's primary transit system can cost as much as $14 each day roundtrip, including parking, between a suburban station and downtown offices. In addition, the number of people over the age of 65 will increase by 80 percent by 2025; more than half of the people in this age group stay at home on any given day because they lack transportation[12] – and driving is not an ideal solution, as people in their seventies have nearly four times the accident rates as those aged 25-65.

- **Homeland Security:** While the U.S. has been spared terrorist incidents on trains and roadways, the threat remains very real. New strategies – such as tracking freight through multi-modal journeys – must be implemented to detect danger and manage reaction to natural and man-made disasters. Likewise, despite the many advances in aviation security in the past decade, more work remains, as evidenced by the recent Christmas Day terror plot aboard a Northwest Airlines trans-Atlantic jetliner bound for Detroit, MI[13].

Because so many of our daily functions are dependent upon transportation, it is critical that we develop new approaches to addressing these problems. New technologies – from "black boxes engineered into vehicles to capture the last few seconds of information prior to an accident to roadside sensors that measure traffic speeds in real-time – are being developed and deployed. At the same time, data mining and machine learning techniques are advancing, allowing us to analyze these data to not only inform large-scale decisions for reengineering the transportation system for the twenty-first century but also much more fine-grained issues such as which routes a given individual might wish to take to avoid traffic jams. As we describe below, Federal support of the *data* → *knowledge* → *action* paradigm is critical for improving our transportation system and, in turn, ensuring our economic productivity well into the future.

---

[9] http://www.asias.faa.gov/portal/page/portal/ASIAS_PAGES/ASIAS_STUDIES/PDFS/ASIASWRONGRUNWAYREPORT.pdf.
[10] http://en.wikipedia.org/wiki/Comair_Flight_191.
[11] http://www.ntsb.gov/Dockets/RailRoad/DCA09MR007/default.htm.
[12] http://www.apta.com/research/info/online/documents/aging_stranded.pdf.
[13] http://en.wikipedia.org/wiki/Northwest_Airlines_Flight_253.

Moving toward "New Transportation"

The costs of sensors, communication tools, and data analytics approaches have dramatically decreased in recent years, and consequently these now have the potential to completely change the transportation landscape in the coming years. However, they have not yet been effectively explored or exploited; the presence of many stakeholders, coupled with a variety of complex incentives, has hindered innovation in transportation. **Federal leadership is therefore essential** to overcome this logjam, to encourage new approaches, and to yield new programs that incentivize change. The following basic research elements are vital to any Federal investment in "New Transportation":

- **Improved urban design**: A number of urban and suburban regions have developed creative plans for converting traffic-clogged sprawl into areas that mix residential and commercial development. Most trips can be made by walking, biking, or in short-range electric vehicles for people with limited mobility. High-density housing and commercial development is encouraged around transit hubs. **Shifting to these more efficient designs requires collecting and analyzing data on local traffic flows and population densities, including modeling future urban plans and assessing their effectiveness *in silico* prior to adopting and deploying them.** For example, based on computational models, New York City has increased the number of pedestrian/bike lanes, which in turn are yielding – as predicted – increased business for retailers and decreased numbers of accidents. In addition, the city is following the lead of other urban centers and introducing bus and traffic light control systems that can anticipate a bus's approach to a traffic light and adjust the signal timing to ensure the bus's efficient passage through the intersection. In the near future, data analytics will enable signal timings to be altered on the fly on the basis of real-time traffic flow data to lessen overall congestion. Ultimately, the shift to more efficient urban designs may take many years but can only occur if communities develop clear goals and utilize these goals to guide decisions about new construction and infrastructure investments. This approach has worked particularly well in housing developments built around new transit projects.

- **Personalized, real-time information for choosing travel options:** New information tools should make it possible for individuals to use hand-held GPS navigation and other devices to identify a variety of options for travel – and make the best decision in terms of time and cost.[14] By entering a destination, a traveler could be given a price and estimated time of arrival for options including walking (including directions), mass transit (where to go, what bus/train to catch, next available arrival, etc.), and jitney, taxi, and "zip car"/bike locations. Selecting a jitney or taxi would instantly send an order and update routing. Some bus companies already let people find the next bus at a stop using conventional cell phone text messages, or form "just in time" car pools using services from companies like Ride Now[15]. The Irish firm Avego[16] is experimenting with methods that use the iPhone to let people offer rides to others headed in the same direction and receive appropriate payments as compensation for their services. "Zip cars" are a particularly attractive option in urban areas;

---

[14] http://www.google.com/intl/en/landing/transit/#mdy.
[15] http://www.ridenow.org/.
[16] http://www.avego.com/ui/index.action.

Americans appreciate the convenience of personal vehicles, but the average personal vehicle is utilized less than five percent of the time over its lifetime, meaning that the economic and environmental costs of manufacturing it are not well-amortized (and also that it is occupying an expensive parking space more than 95 percent of the time). **These approaches will require feeding data from roadside (as well as on-board car, bus, and train) sensors into computational models that predict traffic patterns and travel times at different times of the day – taking into account other factors, such as the likelihood of a crash on a particular roadway at a given time of day, weather conditions, the anticipated fuel consumption and cost, and CO_2 emission rates, etc. These models will enable predictions of train and bus arrival times, compare these mass transit approaches with different routes one could take by car, and make recommendations of how to most efficiently travel from point "A" to point "B." Similarly, zip car systems require capturing and analyzing usage patterns over time in order to ensure that each zip-car location has cars at all times.**

- **Improved highway vehicle management:** New technologies[17] also permit real-time, individualized information and advice for drivers and highway managers, including such services as: real-time reports on road conditions; incident detection and management; surveillance and detection of hazardous material; open road tolling; electronic border crossing and credentialing; electronic parking payments and guidance to free spaces; commercial vehicle inspection verification; variable message signs; on-ramp metering; improved incident management; and driving fees based on when and where a vehicle is driven (e.g., the fees charged for driving in downtown London during business hours). These and other steps can improve safety and reduce congestion using technology available today. **The key technologies are low-cost sensors embedded in highways, wireless communication systems (including analysis of cell phone signals), and low-cost sensors in vehicles (radar, GPS, and accelerometers).** Dedicated Short Range Communications devices (a variant of Radio-Frequency Identification) play a critical role since they allow vehicles to communicate with each other and with the highway. Additionally, sensors embedded in bridges and other parts of the highway system allow early detection of flaws. Improved highway vehicle management will be increasingly important in the coming years, as the American Recovery and Reinvestment Act (ARRA) of 2009 investment will accelerate the diverse mixture of hybrids, plug-in hybrids, and electric cars – thereby greatly increasing the number and type of vehicles on the roads.

- **Load-balanced transit: An even more advanced approach involves linking GPS navigation devices with cloud computing platforms. Each traveler could input his or her destination into his or her GPS navigation device, and then this device would transfer this information to the cloud computer. The cloud computer would run a dynamic routing algorithm to determine the collective fastest routes for all travelers in the system.** Furthermore, by synching this routing algorithm was real-time traffic data, vehicles could be re-routed in transit. This approach would cause roads to be used in a much more balanced manner, thereby minimizing congestion and decreasing overall travel time – all the while without necessitating expensive new transit networks. Indeed, imagine never

[17] http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPTS_te/14412.HTM.

pulling onto a gridlocked highway again. Drivers would be pre-empted to use less busy roads and to drive at different times. As the sophistication of the algorithm increases, required stops could be incorporated. For example, a user could tell the cloud that he or she wants to stop for coffee at some point during their trip. The system would then tell the user the ideal time for coffee (and the nearest shop), while minimizing system-wide transit time. Vehicles would also be re-routed in order to avoid emergency vehicles, parades, and funeral processions, etc. These load-balancing technologies can also be applied to both the trucking industry and freight trains, minimizing the resources needed to move essential goods cross-country.

- **Real-time driver assistance: Information available from the sensor network would also provide resources to help drivers navigate dangerous conditions through such things as adaptive cruise control and collision and rollover warning/avoidance.** These tools are particularly important for individuals with disabilities and for a population of increasingly older drivers. Over time – given proper advancement in machine learning and data mining – these "cruise control" technologies can evolve to take on an increasingly complex set of tasks and safety maneuvers. Given successful research outcomes, it may be possible to develop a new generation of "cruise control" that would make it possible to put more vehicles in the same highway space allowing an increase in highway capacity without decreasing safety. Research could also lead to an infrastructure for conveys of computer-controlled trucks traveling on dedicated guideways.

- **Improved aviation:** GPS technology will replace ground-based radar systems for tracking planes, allowing for more accurate positioning and allocation of aircraft in our skies. **GPS-based navigation will prevent planes from inadvertently flying too close to one another and provide a more accurate portrait of routes. Data mining approaches will allow us to simulate critical scenarios, and to optimize flight plans and better detect unsafe flying conditions automatically. Data on flights, flight plans, and delays will also be analyzed in order to optimize airport layout and design, routing, and responses to delay conditions such as weather or high volume.** For example, 70 percent of current aviation delays are attributed to weather; optimizing flight plans in real-time system-wide through GPS navigation coupled with model-based recommendations could drastically reduce these delays. **Moreover, the current jumbled set of aviation interfaces could be unified into a single system enabling all stakeholders to easily view the planes flying over the U.S.; this interface will mitigate data redundancy and facilitate information sharing.** Finally, there is increasing evidence that aircraft energy consumption can be cut by about five percent – or as much as $80 million in fuel savings for a large airline – using "continuous descent" to move aircraft continuously from cruise altitude to landing (and vice-versa) instead of the current star-step descents from one fixed altitude to another[18]. **Moving to this mode depends entirely on the development and deployment of new trustworthy algorithms.**

- **Automatic scheduling of mass transit systems:** Analysis of turnstile data, bus and train data, and sensor networks could allow for accurate counts of the number of people riding these systems and the number of people waiting at each stop. This information could, for

---

[18] http://www.sciencedaily.com/releases/2009/02/090203091356.htm.

example, be used to dynamically add and remove trains to and from a subway system based on demand. Dynamic train allocation would be enhanced – given appropriate advances in data analytics – if subway systems were fully automated because demand could be decoupled from staffing.

- **Enhanced safety, reliability, and redundancy:** Importantly, as highlighted by the June 2009 Metrorail crash in Washington, DC, briefly described above, semi- or fully-automated systems are not without their challenges. Extensive research into safety, reliability, and redundancy must be conducted prior to deploying such systems in order to ensure that automation does not jeopardize lives. Use of data analytics can also enable us to detect safety concerns as they arise in real time. For example, in June 2009, an Airbus A330-200 being operated as Air France flight 447 crashed into the Atlantic Ocean, killing all 228 people on board. No one knows why the plane fell out of the sky because its black boxes – those rugged, reinforced, waterproof cases housing the plane's flight data and cockpit voice recorders – have never been located on the sea floor. Lacking these data, investigators have no way of knowing the exact cause of this crash. Recently, researchers have proposed a "glass box," i.e., a system by which data from aircraft would flow to ground stations in real time using high-bandwidth radio or lower-bandwidth satellite links, to be analyzed on the spot or later on. New analytics are providing the capability to make sense of such data, which in turn will enable operators and government safety officials about problems as they arise in real time – as well as preventative measures that should be adopted in the future based on much more highly specified conditions and incidents. (For instance, currently, the data within black boxes are only analyzed when an incident occurs; "glass boxes" will enable researchers to evaluate data even from "normal" flights, leading to new insights that will enhance overall safety, reliability, and redundancy of airplane systems.)

The need for Federal investment in transportation

Improving the transportation system will involve deploying sensor networks and continuing to develop tools for analyzing the wealth of data that they are likely to generate. This work requires coordinated basic research funding from a variety of Federal agencies involved in overseeing various aspects of the transportation system. Specifically, we must:

- **Undertake a major upgrade of the Department of Transportation's (DoT) research program, making it responsible for managing an ambitious program of technical research as well as economic and policy analysis – possibly by greatly expanding the Research and Innovative Technology Administration in the Department of Transportation now funded at only $10 million/year.** DoT presently spends about $570 million on surface transportation in several different Administrations (Highway, Transit, Railroad, and Motor Carrier Safety). Close collaboration with the National Institute of Standards and Technology (NIST) and the Department of Energy (DoE) is essential. A fixed fraction of these funds should be dedicated to high-risk research on potentially disruptive technologies in the data analytics space.

- **Establish programs within the FAA to support analysis of the wealth of historical data about flights, flight plans, and delays, as part of the huge aviation modernization effort**

termed NextGen. The FAA received $56 million for R&D in 2010 for Air-Ground Integration, Self-Separation, Weather in the Cockpit, Environmental Research, and the Joint Planning and Development Office (JPDO). However, none of these efforts has directly enabled the basic computing research that is necessary to develop and optimize the aviation systems of the future.

- **Create a number surface transportation research centers at universities based on a competitive solicitation (each would be funded for at least five years), likely through the DoT.**

- **Request NIST to develop interoperability standards for intelligent transportation systems and safety (there is already incompatibility between U.S. and European implementation of Dedicated Short Range Communications devices).**

- **Create a competitive solicitation within the Department of Housing and Urban Development (HUD) for innovative intelligent transportation schemes for urban areas.**

- **Task the National Science and Technology Council (NSTC) with building a tightly integrated program involving DoT, DoE, FAA, NIST, and HUD to carry out these missions.**

The road ahead

In recent years, we have witnessed many signs of stress in our transportation system. Bridge collapses, train derailments, "near-misses" in our skies; higher fuel prices, increasing CO_2 emissions; and more frequent delays and cancellations – together, these problems illustrate how our current approach to transportation is simply not sustainable as more people and goods exhibit time-sensitive travel needs.

As we consider reengineering our transportation infrastructure for the twenty-first century, data analytics approaches offer tremendous promise. Massive, low-cost sensor networks are being deployed to collect a wealth of new data about transportation – from details about how frequently a car's accelerator and braking is used to highway traffic patterns. At the same time, data mining, machine learning, pattern recognition, computer modeling, security, and optimization techniques are enabling us to analyze these data to increase the safety and efficiency of transportation. For example, in some cases, researchers are already modeling traffic flows to improve urban designs. In other cases, continuously–updated analyses of traffic conditions will soon inform real-time decisions about which turns a driver should take in order to to go from point "A" to point "B."

Ultimately, a renewed Federal commitment to these technologies is essential in order to ensure that our transportation system is able to keep up with the demands of the future – and to help the U.S. continue its prominence at the forefront of the global economy, as so much of our economic viability is contingent on the safe, timely, low-cost, long-haul transport of people and goods.

For citation use: Thrun S., Hensel C., & Gianchandani E. P. (2010). *From Data to Knowledge to Action: Enabling a Revolution in New Transportation*: A white paper prepared for the Computing Community Consortium committee of the Computing Research Association. http://cra.org/ccc/resources/ccc-led-whitepapers/