

From the masses

Crowdsourcing has become a popular buzzword with the recent trends in data-based transportation-system management.

The term *crowdsourcing* is relatively new, but the basic idea is not. Crowdsourcing involves leveraging the combined intelligence, knowledge or experience of a group of people to answer a question, solve a problem or manage a process. It is a natural solution to problems inherent in the management of complex systems. Transportation research is increasingly investigating digital crowdsourcing platforms as sources of valuable data.

Establishing a working definition of crowdsourcing is complicated because the word is applied in common usage to a wide variety of activities and processes. Yet many traditional activities that would seem to qualify as crowdsourcing are often not thought of as such (e.g., public meetings). In an attempt to lend some structure to discussions about crowdsourcing transportation data, we have developed a taxonomy that defines four specific types of crowdsourcing that may be used by transportation agencies. These are:

1. Third-party aggregated data;
2. Social media engagement;
3. Internet-as-a-sensor; and
4. Dedicated platforms.

Third-party aggregated data

Many transportation agencies already obtain crowdsourced data through contracted commercial providers. Commercial providers offer clearly defined products and services, as well

as customer support and professional expertise. Such arrangements allow agencies access to proprietary data that would be otherwise difficult to obtain, and allow them to avoid tedious data cleaning and management tasks. Agencies can also avoid dealing with sensitive personally identifiable information by receiving only aggregated data from providers.

Transportation agencies most typically contract third-party vendors for historical and live traffic-condition data. This often includes average speed, vehicle counts and (possibly) incident detection. An emerging trend is the use of aggregated location data for assessment of broad mobility patterns across a network. Such data can be used for travel-demand modelling, long-term planning and project prioritization. Traditionally, this information is collected through public surveys

and travel journals; these low-tech crowdsourcing methods are generally expensive, have a limited sample size and are prone to errors. Fortunately, the broad adoption of cell phones and connected devices provides an opportunity for accurate and objective measurements of peoples' movements and use of the transportation system. Data vendors and transportation agencies are increasingly leveraging such capabilities to improve

Four ways crowdsourced information can work for transportation agencies

understanding of people's movements across a transportation network.

Social media for public engagement

Americans spend more time using social networks, on average, than any other online activity. Social media provides an opportunity to streamline and improve public relations and communications. Many agencies have already deployed social media strategies with positive results. Social media can be extremely useful for communicating system status during special events and disruptions. Agencies can simultaneously disseminate information, gather additional information from system users and take advantage of instantaneous, unmediated information-sharing on the platform. Additionally, some planners have successfully experimented with

Many traditional activities that would seem to qualify as crowdsourcing are often not thought of as such.

engaging constituencies via these platforms to supplement and support traditional public participation in the planning process.

The Internet as a sensor

Broadly defined, a sensor is any device that takes a measurement and converts it into readable data. Recent research in transportation-demand modeling has investigated the concept of using the Internet itself as a "virtual sensor"

Table 1. Summary of Crowdsourced Transportation Data Categories and Use-case Examples

Crowdsourced Data Category	Use-case Examples
Third-party Aggregated Crowdsourced Data	<ul style="list-style-type: none"> • Historical and real-time traffic data • Cellular geo-positioning data
Social Media for Public Engagement	<ul style="list-style-type: none"> • Collection and dissemination of system status information • Planning and project prioritization
Internet-as-a-Sensor	<ul style="list-style-type: none"> • Mining social media data to supplement public participation • Using available traffic data • Predicting traffic spikes
Dedicated Platforms	<ul style="list-style-type: none"> • Automated vehicle location for public transit • Pavement condition data collection • Bicycle travel and infrastructure data • Origin-destination studies • Planning and project prioritization

capable of providing valuable, actionable data. New paradigms of open data and big data analysis are making it possible to scan a wide range of Internet sites in search of posts that relate to a particular topic or agency. Social-networking sites and other Internet sources may be mined for relevant data using Internet-as-a-sensor approaches. Such methods are generally in early stages of development, and might require complex methods of data mining and sentiment analysis. As these methods continue to develop, transportation planners should consider these techniques to supplement traditional

methods of public participation.

Internet-as-a-sensor methods may also be used by agencies to supplement traditional traffic-condition monitoring and forecasting. Today, real-time comprehensive performance data is available through proprietary probe-vehicle networks, many of which make limited, live-traffic data available for free over the Internet. One study by New York University researchers found that travel time estimation provided by these Internet sources is within 95% of the measurements by traditional sensor networks (MVDS). The Internet data was also generally found to be less prone

to errors and gaps in data provision. The study concluded that this “virtual sensor methodology comes with almost no additional cost while the quality of the data obtained is quite satisfactory compared to physical sensors.”

Dedicated platforms for transportation data collection

The three previous categories of crowdsourced transportation data involve re-purposing existing platforms for transportation-related data. The fourth and final category includes dedicated crowdsourcing platforms—where the data generated by the platform is created specifically for transportation-system management. Such platforms can be developed in the public sector, private sector or (perhaps most effectively) as a public-private partnership. Dedicated platforms often take advantage of smartphone ubiquity and capability, utilizing the phone’s built-in GPS, accelerometers, camera and other sensors to collect a wide variety of information. Potential applications for dedicated platforms include:

- Automated vehicle location for public transit;
- Pavement condition data collection;
- Bicycle travel and infrastructure data;
- Origin-destination studies; and
- Planning and project prioritization.

TM&E

Dennis is a Transportation Systems Analyst with the Center for Automotive Research.