Adventures in Crowdsourcing Webinar Series: Vendor Engagement

Thursday, December 5, 2019

FHWA EDC-5 Crowdsourcing for Operations Program

Source: Pixabay, Unsplash and FHWA
Today’s Agenda

Welcome and Introductions
Greg Jones, FHWA

Acquiring Probe Vehicle Crowdsourced Data
Colby Fortier-Brown, Maine DOT

Acquiring Crowdsourced and Traditional Data
Chris Lambert, Kentucky Transportation Cabinet

Vendor Engagement for Big/ Crowdsourced Data
Michael Pack, CATT Lab, University of Maryland
FHWA EDC-5 Crowdsourcing Innovation Goal

To increase the number of agencies that use crowdsourcing to better operate the transportation system through new, cost-effective, and proactive operational strategies and applications.

Crowdsourcing for Operations
Crowdsourced data allows agencies to more effectively operate the transportation system.

30+ States are pursuing Crowdsourcing for Operations.

Source: FHWA
Operations Application Area Highlights

- Most common applications of Crowdsourced data are for traveler information and incident management.
- Agencies are now expanding its use across application areas such as maintenance, road weather and work zone management.

Source: Unsplash, FHWA, and Pixabay
Initiative Resources

- **Webinars**
  - Adventures in Crowdsourcing: Data Management & Use
    January 28, 2020 1:30 - 3:00 PM

- **Workshops (see downloadable workshop flier)**

- **Peer exchanges**
  - Validation & Filtering of “Free” Crowdsourced Data
    Dec 10-11, Nashville, Tennessee
  - Making the Most of Real-Time Vehicle Probe Data
    February/March 2020 (Date/Location TBD)

- **On-site technical assistance**

- **Training materials and case studies**
EDC-5 Deployment Funding Opportunities & News

State Transportation Innovation Councils (STIC) Incentive Program
✓ Up to $100,000 per STIC per year to standardize an innovation
✓ [https://www.fhwa.dot.gov/innovation/stic/](https://www.fhwa.dot.gov/innovation/stic/)

Accelerated Innovation Deployment (AID) Demonstration Program
✓ Up to $1 million available per year to deploy an innovation not routinely used
✓ [https://www.fhwa.dot.gov/innovation/grants/](https://www.fhwa.dot.gov/innovation/grants/)

For more information on the EDC Program
Email: [https://www.fhwa.dot.gov/innovation/](https://www.fhwa.dot.gov/innovation/)

to subscribe to weekly and bimonthly newsletters
Polling Question
Contacts

James Colyar
James.Colyar@dot.gov
(360) 753-9408

Paul Jodoin
Paul.Jodoin@dot.gov
(202) 366-5465

For more information on the EDC Program

www.fhwa.dot.gov/innovation/everydaycounts/

Email: https://www.fhwa.dot.gov/innovation/

to subscribe to weekly and bimonthly newsletters
PRESENTATION OUTLINE

- MaineDOT Probe Data History and lead-in
- What we learned along the way
- Where we are now
- Takeaways
HOW IT STARTED

- Maine is relatively new to ITS
- We do not have an established network of Bluetooth devices, but we needed O/D and travel time data
- With the growth, we regularly access multiple crowdsourced datasets, mainly for planning and operations

- Travel Time Sign Project advertised in May, 2018
- Probe Data was purchased from TomTom as part of the contract and integrated into ATMS for operation of the signs

- Sole Source Justification completed with TomTom and negotiations started
- These same considerations can be used for RFP process
OPENING PERSPECTIVE

- Historically, agencies wanted data and collected it themselves. They could do what they want with it. “Work for Hire”

- The role of commercial data providers is to provide licensed, off the shelf data.

- From the beginning, we have to create those bridges between historical “Work for Hire” and commercially licensed data

- It is important to keep an open mind about what these vendors can provide beyond the standard product
WHY DO YOU WANT DATA?

- What will you use it for?
- What CAN you use it for?
  - O/D
  - Climate studies
  - User delay costs
  - 85% speed for horizontal curves
  - Etc.
WHAT TIME PERIOD DO YOU NEED?

- Live is necessary?
- Archived is okay?
- Somewhere in the middle with more processing?

- Answer needed NOW or answer can be calculated later?
  - If later, how far in the future?
  - Do you need to keep it beyond license?
CURRENT OR PERPETUAL LICENSE?

- Current Licensing is typically less expensive
- Current Licenses are temporary and renewable – so is the integrated licensed data from Current Licenses
- Data sharing constraints are the same with either type of licensed data
- Live data can typically be a great use of a Current license if you don’t need it later
- Certain parts of the record can be kept in perpetuity
- Licensing proprietary analytics platforms provide a lot of power and access to unique benefits
WHERE WILL YOU STORE THE DATA?

- Vendor system?
  - Move software will allow querying of TomTom archives
- Internal system?
  - Storage and standardization is expensive
- ATMS?
  - Must be defined in agreement, specifically
TRANSFORMATION OF LICENSED DATA

- Raw data or data product
- Display with no values (Green/Yellow/Red)
- Reports
- Derivative Products
  - Hard to qualify

* Make sure that everyone is on the same page about these options
WHO CAN SEE IT?

- **Internal**
  - Data
  - Displays
  - Reports
  - Derivative Products

- **Public**
  - Data
  - Displays
  - Reports
  - Derivative Products

- **Other Agencies**
  - Data
  - Displays
  - Reports
  - Derivative Products
  - Executing your Agency’s role in cooperative work

* Those you share with may be potential customers
DATA PROTECTION

- Data is valuable

- Protect your own data
  - Unless there is a business case for it, do not allow ownership of your collected probe data
  - Standard licenses are not necessarily for Government groups
  - Regardless of Vendor

- Protect their data
  - Do not make proprietary data openly available
  - Be careful combining data sources in the same database
  - Know your other licensing agreements (ATMS or other defined software partners who will use this crowdsourced data)
SO WHAT HAPPENED?

- Negotiations complete…Rejected by AG’s office
  - We spent the entire negotiation process working on solely the business case in an established vendor contract
  - Re-writing the contract with the state template

- Proprietary Data Products are powerful, but are not as common and require unique considerations. There is no real history for this type of product. These agreements may make your legal staff uncomfortable.
SOME THINGS TO REMEMBER

- Know what you want; consider other potential uses
- Don’t be afraid to ask the provider “what else?”
- Stay problem-focused instead of product-focused
- Create a positive working relationship with providers that you are working with
- Carefully negotiate, but understand that you may not get a Perfect contract
- Stay engaged- it is easy to lose interest when waiting
- Have everyone that you need at the table
  - Technical staff
  - IT
  - Legal
  - Contract specialists
Vendor Engagement

Forming Successful Partnerships
Lessons Learned

• Understand the Data
• Promote Data Standards
• Prepare to Change
Core Partners Being Utilized

• HERE Traffic
• GeoTab
• Waze
• iCones
• Northrop Grumman
  • TRIMARC
  • Truck Parking
  • RWIS
• KYMesonet
• CoCoRahs
• National Weather Service
• Government Entities
  • Louisville
  • Lexington
  • KY Emergency Management
## KYTC “Big Data”

<table>
<thead>
<tr>
<th>Source</th>
<th>Records per Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>HERE</td>
<td>2.5 billion</td>
</tr>
<tr>
<td>Waze Jams</td>
<td>128 million</td>
</tr>
<tr>
<td>Waze Incidents</td>
<td>180 million</td>
</tr>
<tr>
<td>AVL (Snow Plows)</td>
<td>175 million</td>
</tr>
<tr>
<td>All Others</td>
<td>123 million</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3.1 billion unique records</strong></td>
</tr>
</tbody>
</table>
HERE Traffic

• Licensed
• Competitive Costs
• Allows for Sub-Licensing
• Allows Republishing Derivatives
• Previews of Upcoming Products
GeoTab

• Contract
• Data Included at No Additional Cost
• Automatic Vehicle Location
• Licensed Through Ky Fleet
Waze

• Data Exchange Agreement
• Free Real-Time Data
• Protected from Open Records
• Terms of Service:
  • Attribution
  • Road Closures
iCones

- Contract
- Provides Smart Work Zone Sensors
- Real-Time Data Feed
Northrop Grumman

• Statewide ITS Contract
• Traffic Management Center
• Hardware Services
Incident Detection (2019)

District: 6 - Kenton – I-75 – Northbound – Milepoint Range: 188.0 to 190.5
A crash has been detected at milepoint 190.3 with an average speed of 28.0 mph and a historic free flow of 54.9 mph.
Maps: Google | TOC Incident Manager | Waze | Google | Here Traffic
Dashboards: Real Time View | After Action View
Weather: National Weather Service | National Weather Service Details | Windy

Graphs and charts showing data analysis and incident reports.
Incident Detection / After-Action

HERE Speed data being graphed, Cardinal and Non-Cardinal directions.

Dots are where Waze users are reporting issues.

The point in time where our TMC received notice about the event.

Detection script detects and reports an incident.
CoCoRahs
Open and Proprietary Products

• Back-End
  • Apache Kafka
  • Apache Spark
  • Apache Hadoop
  • Elasticsearch

• User Experience
  • Kibana
  • ArcGIS Online

• Languages
  • Python
  • Scala
  • Java
QUESTIONS?

Chris Lambert
Intelligent Transportation Systems
(Retired from KYTC, Dec 1, 2019)

New Contact Information:
chris.lambert@aemcorp.com
859-619-7757
Challenges & Opportunities for 3rd party data partnerships.

Michael Pack, CATT Laboratory

Pack’s Pointers on how agencies can better leverage 3rd party data and private sector relationships.
Data alone isn’t the answer.

- Agencies need:
  - Policy guidance,
  - Tools & technologies,
  - Research & development, and
  - Thought leadership that helps reduce anxiety and increase big-data capabilities

To prevent this scenario:

Data Providers

You (and your poor staff)

Data from Everywhere
For Example: Waze data can be a firehose!

Note:
• Waze data excludes jams event type
• 3 Month Period of 3/17 – 5/17 displayed

<table>
<thead>
<tr>
<th>State</th>
<th>Avg DOT Events Per Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>3,184</td>
</tr>
<tr>
<td>DC</td>
<td>16</td>
</tr>
<tr>
<td>FL</td>
<td>1,895</td>
</tr>
<tr>
<td>IA</td>
<td>114</td>
</tr>
<tr>
<td>MA</td>
<td>14</td>
</tr>
<tr>
<td>PA</td>
<td>70</td>
</tr>
<tr>
<td>VA</td>
<td>681</td>
</tr>
</tbody>
</table>
## Waze Data Background

<table>
<thead>
<tr>
<th>Type of Event</th>
<th>Average Time that a Waze Event was Reported Before a DOT Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freeways/Ramps Crashes</td>
<td>3 Minutes</td>
</tr>
<tr>
<td>Primary/Secondary Crashes</td>
<td>3 Minutes</td>
</tr>
<tr>
<td>Freeways/Ramps Disabled Vehicles</td>
<td>14 Minutes</td>
</tr>
<tr>
<td>Primary/Secondary Disabled Vehicles</td>
<td>16 Minutes</td>
</tr>
</tbody>
</table>
Working with Waze

Working with the Data

- Redundancy
- Feedback loops
- Size
- Credibility and filtering
- Increased Coverage
- Faster Response
- The ability to truly influence route-choice

Working with the Company

- Legal
- Negotiations
- Nothing is really free
Common 3rd Party Data Providers and Services

- Speeds and travel times
- Data Feeds & APIs
- Map and data tiles
- O/Ds
- Trajectory/Trips
- Location-Based Services (LBS)
- Mapping
- Some are working on volumes and turning movements
- Much much more coming soon!!
- Not all provide the same type of data, the same format, etc. even for similar data types
3rd Party Data can be AWESOME!!!

• But...

YOU the purchaser can ruin it!!!
I mean, really really ruin it.
Procurements can go wrong.
And you can also get played.
Don’t make these mistakes

• DUAs – You have the power!
  • Fight for Great Acceptable use
  • Fight for (and think about) Sharing with partners
  • Don’t just do what your neighbor did (but ask them)
  • Look for model DUAs (I-95 CC for probe data)

• Sharing back with the provider the way YOU want to share it
  • (don’t permanently dumb down your data)

• Treat your provider as part of your team, NOT a whipping boy

• Be open to communication and vendor discussions

• Don’t blend “all” of the requirements

• Payment terms based on quality and uptime (where applicable)

• Stop focusing on how to pay less. Instead, work to try to get more!
Invest in Tools to Make Fused Data Easy to work with

- Data is only useful when it is
  - easily accessible,
  - usable, and
  - understandable

To managers, planners, operations, and ITS applications…
To be effective, you need the following:

- Data
- Tools
- Domain Expertise

Fusion, Statistics, & Integration
Analysis & Visualization

= Insights
Technical Capacity Needs to Increase (and diversify)
• Don’t just train Transportation Engineers to do this stuff.
  • Hire other skill-sets and teach them about transportation
    • Data Journalists / Analysts / Data Scientists
  • Consultants can do this, too, but….  
    • Think long-term (don’t hire then fire)
    • Train staff and transfer knowledge
• Partner with Universities (or other similar institutions)
• Invest in Research

Invest in your technical capacity
Beware of Distractions and Hype
Buzzwords, Shiny Objects, and Peer Pressure

- Blockchain
- Machine Learning
- Artificial Intelligence (AI)
- Business Intelligence
- The Cloud
- Agile
- Etc.

Know what they mean. Don’t confuse them. Understand their relevance. Don’t think they’ll solve all your problems.
Big Data: Savior or Big Fat Tease

Time

Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity

Expectations
The Cloud (hype, sales, or savior?)

- The cloud is EXTREMELY cost effective when you do things the way they want you to do them!
- Don’t assume the cloud will save you money or improve capabilities
- You don’t have to be in the cloud to be effective and innovative
- The cloud should not be used for everything
- The cloud is not “all or nothing”
- Not all clouds are created equally
- **Virtualization** is not the same thing as **cloud computing**
Know your terms...
• Well-intentioned people confuse open source and open data.

• Making institutional investments based on a misunderstanding of terms can have drastic impacts!

• Open Source typically applies to software and applications

• Open Data applies to DATA
Pack’s Predictions for the Future...

- Data isn’t going to get any smaller.
  - Deploying data collection infrastructure will become increasingly less necessary—even at signals!
  - Get your (Current) house in order
    - Or else the latest and greatest thing won’t matter.
    - You won’t be ready.
  - Tools (and newer staff should) make some of this easier:
    - Think of Tableau as the new Excel.
  - But that means that expectations are going to go up, too!
  - We need to invest together and pool our resources for data management and analytics.
Thank you!

Michael L. Pack
Director, CATT Laboratory
PackML@umd.edu
240.676.4060