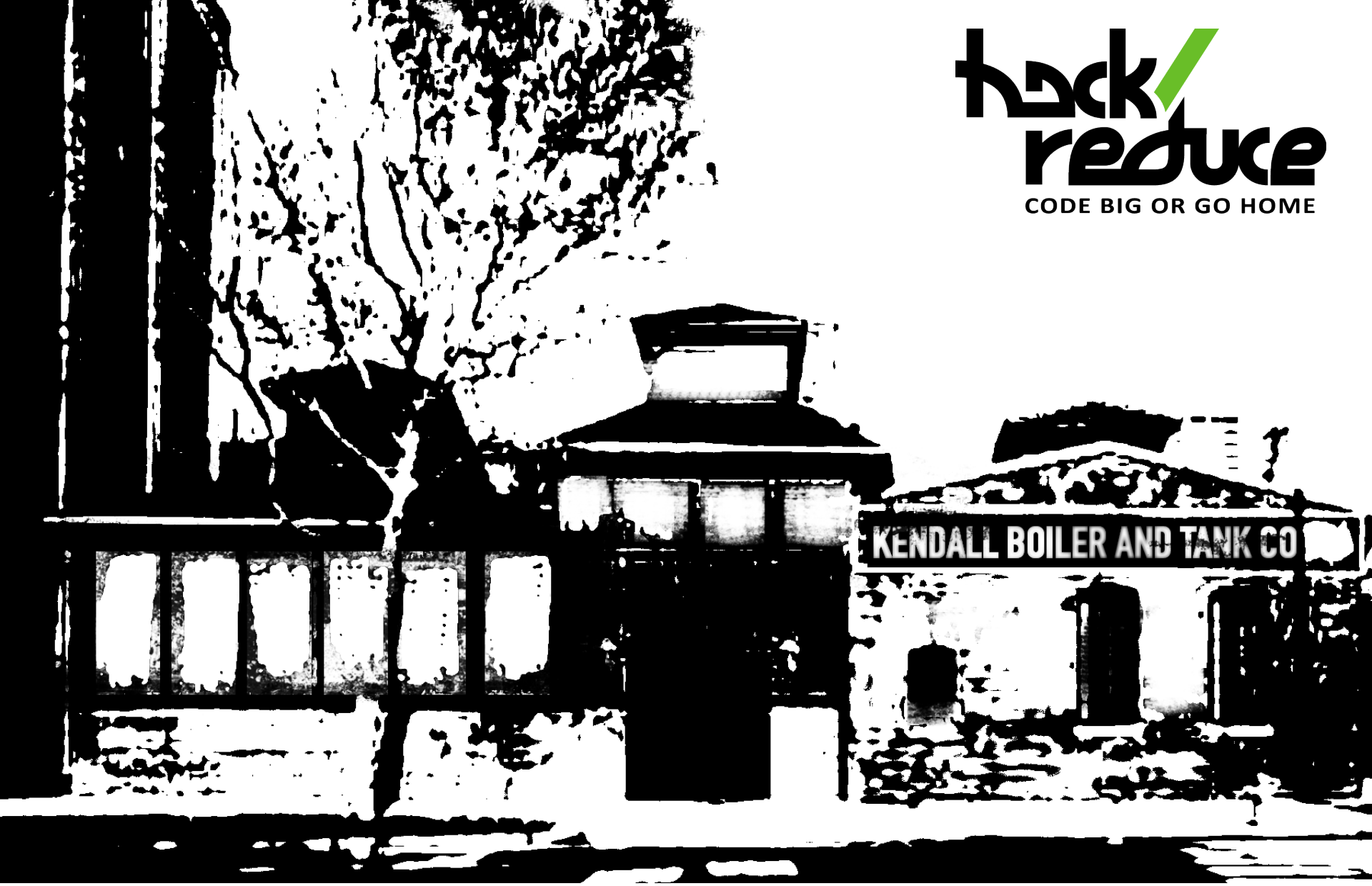


hack/reduce

CODE BIG OR GO HOME

Welcome!

WHY WE'RE HERE

Mission
To promote the growth and the development of the technologies and skills required to process, analyze, and apply data; and to help Boston create the talent and technologies that will shape our future in a big data-driven economy.

Supporting key areas:

- Big Data Technologies and Analytics Development
- Collaboration between Industry & Academia
- Workforce Development and Training

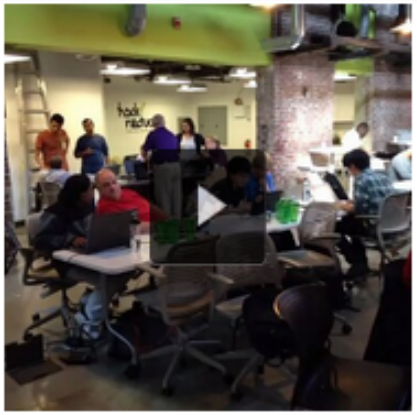
hack/reduce collaborative projects & workshops:

- Focus on addressing issues and solving problems through development of big data analytics
- Serve as a forum for engagement between industry and academia
- Provide programs/events that support training for both students and professionals in big data-related skills

THANKS TO OUR PARTNERS

hack/reduce

CODE BIG OR GO HOME



MassDOT Visualizing Transportation Hackathon

HOSTED BY:

massDOT
Massachusetts Department of Transportation

MASSACHUSETTS
TECHNOLOGY
COLLABORATIVE

hack/reduce

Tweet about it!
#MassBigData @MassDOT
@mass_tech @hackreduce

AGENDA - Friday

6:00pm -- Check-in and grab a slice of pizza.

6:45pm -- Welcome!

- Adrienne Cochrane, hack/reduce
- Christopher Scranton, MassTech Collaborative
- Rachel Bain, MassDOT
- Celia Blue, MassDOT
- Andrew Lamb, Nutonian

7:30pm -- Find a team and get started!

10:45pm -- Goodnight!

Tweet about it!
#MassBigData @MassDOT
@mass_tech @hackreduce

AGENDA - Saturday

9:00am -- Check-in, coffee and bagels.

9:45am -- Welcome back

- Adrienne Cochrane, hack/reduce
- Christopher Scranton, MassTech Collaborative
- Andrew Lamb, Nutonian

10:00am -- hacking!

12:30pm -- Lunch

1:00pm -- hacking!

6:00pm -- Dinner and Presentations begin

8:00pm -- Special Guest Speaker…

8:30pm -- Winners and Prizes!

9:00pm -- Goodnight!

Tweet about it!
#MassBigData @MassDOT
@mass_tech @hackreduce

Mentors

Andrew Lamb
Software Engineer, Nutonian

Russell Bond
Planner, MassDOT

David Barker
Manager of Operations Technology, MBTA

Judges

Clinton Bench
Deputy Executive for Operations, MassDOT

Jeff Mullen
Partner, Foley Hoag, LLP
Former Secretary of Transportation

Christopher Scranton
Senior Manager for Big Data & Tech Initiatives, Mass Tech Collaborative

Speakers

Deval Patrick
Governor, Commonwealth of Massachusetts

Richard Davey
Secretary, Transportation

David Mohler
Deputy Sec. for Policy, MassDOT

Speakers (con't)

Celia Blue
Assistant Secretary, Performance, MassDOT

Rachel Bain
Deputy Registrar for Operations,
MassDOT

Christopher Scranton
Sr. Mgr, Big Data & Technology Initiatives,
Mass Tech Collaborative

Prizes

Best Use of Data
$2,000

Most Visually Compelling
$2,000

Crowd Favorite
$2,000

Rules, Submissions, Voting, etc.

http://masstransporthack.challengepost.com/

Mass Big Data Initiative

Launched in 2012 by Governor Deval Patrick, driven by the Mass Tech Collaborative's Innovation Institute

massbigdata

Mass Big Data Initiative

MASSACHUSETTS TECHNOLOGY COLLABORATIVE

hack/reduce
CODE BIG OR GO HOME

- **Strengthening & Expanding Cluster by:**
- **1- Raising Awareness:**
  Web Portal – Dec Launch
  Major Industry Study – Jan release
- **2 - Workforce: Student Engagement/ Mass Intern Partnership**
- **3 - Accl. Regional Innovation:**
  Sponsorship & Launch of hack/reduce
  Massachusetts Green High Performance Computing Center
- **Enhancing Access to Public Data Sets**


massDOT
Massachusetts Department of Transportation

Technical Details

1. Access to Datasets
2. Dataset Overview

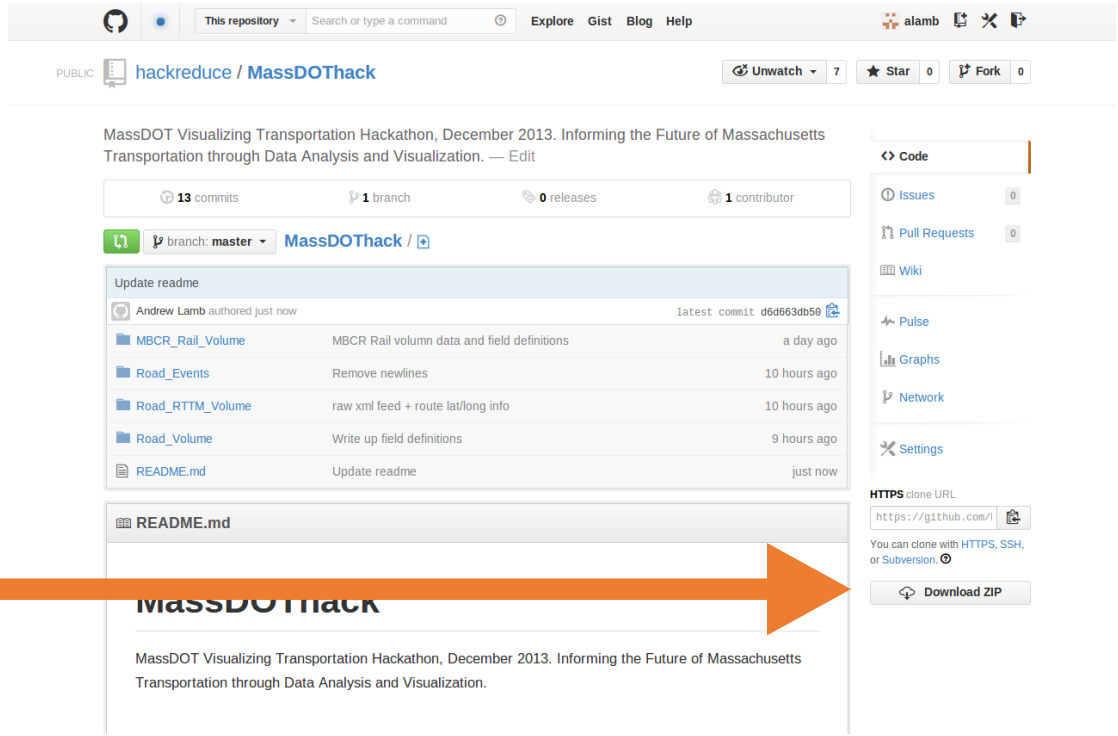
Andrew Lamb
Nutonian, Inc.
andrew@nutonian.com

Dataset Access -- on Github

https://github.com/hackreduce/MassDOThack

Access Options:

1. Download via browser
2. clone git repository




```
$ git clone https://github.com/hackreduce/MassDOThack.git
```

Dataset Format

- Data are in CSV (comma separated value) format
- Easy to use in
  - Excel, LibreOffice, etc
  - Most programming languages
- All datasets have a ..._fields.csv file which describes each column

* Please come and find me if you want/need help transforming the data into a different format

Datasets as SQL (New!)

https://github.com/hackreduce/MassDOThack_SQL

- Due to popular demand last night, a copy of the data is now available as a MySQL database dump -- a .SQL file you can load directly into MySQL
- Also contains the scripts I used to originally load the CSV data (Dump.sql)

Available Datasets

MBCR Rail Volume: MBTA Commuter Rail trip volume

MBCR Rail Locations: Position of each commuter rail train

Road Events: Planned (roadwork) and unplanned (accident) information for major roadways

Road Real Time Traffic Management (RTTM): Fine grained (5 minute) average speeds for major road segments

Road Volume: Average volume (cars/day) on various roadway segments

NOAA Weather: Weather information for stations in Eastern MA

MBCR Rail Volume

- 111 MB of data for 2011, 2012 and 2013
- Individual MBTA Commuter rail trips details
  - Date, route
  - Scheduled vs Actual departure and arrival times
  - Passenger, bike, crew counts

MBCR Rail Locations

- 578 MB of data for November 2013
- Individual MBTA Commuter train locations
  - Timestamp
  - Scheduled route
  - Latitude and Longitude
  - Heading
  - Speed

Note: I also have raw data for anyone who wants it

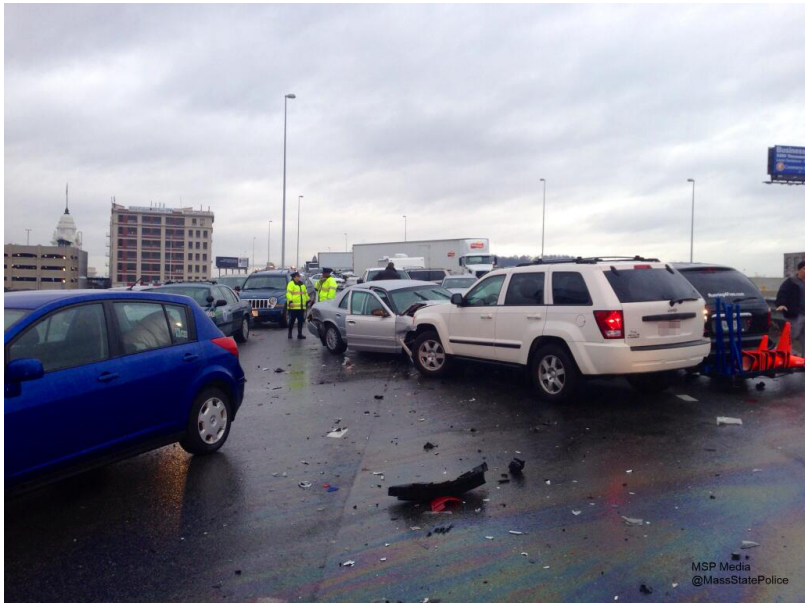
Road Events

- 18MB of data from 2012 to 2013
- Major planned and unplanned events
  - Start and end time
  - Road and weather conditions
  - Textual descriptions of what happened

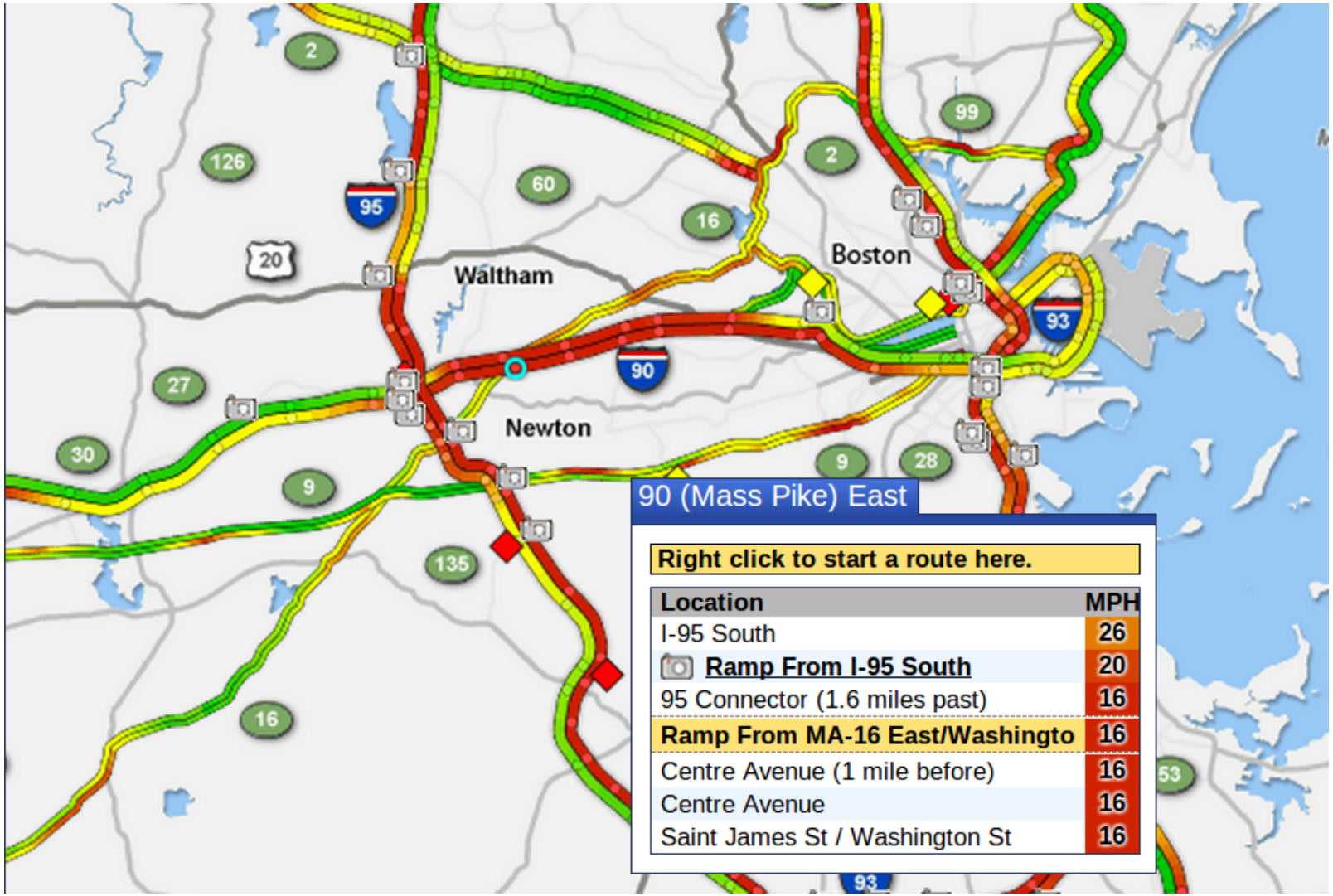
"MSP Charlton reporting Several Thousand Bee's have been abandoned at Rest Area 5E"

"HOC observed MVA FRNB at Mass Ave Connector. Small flames beneath engine."

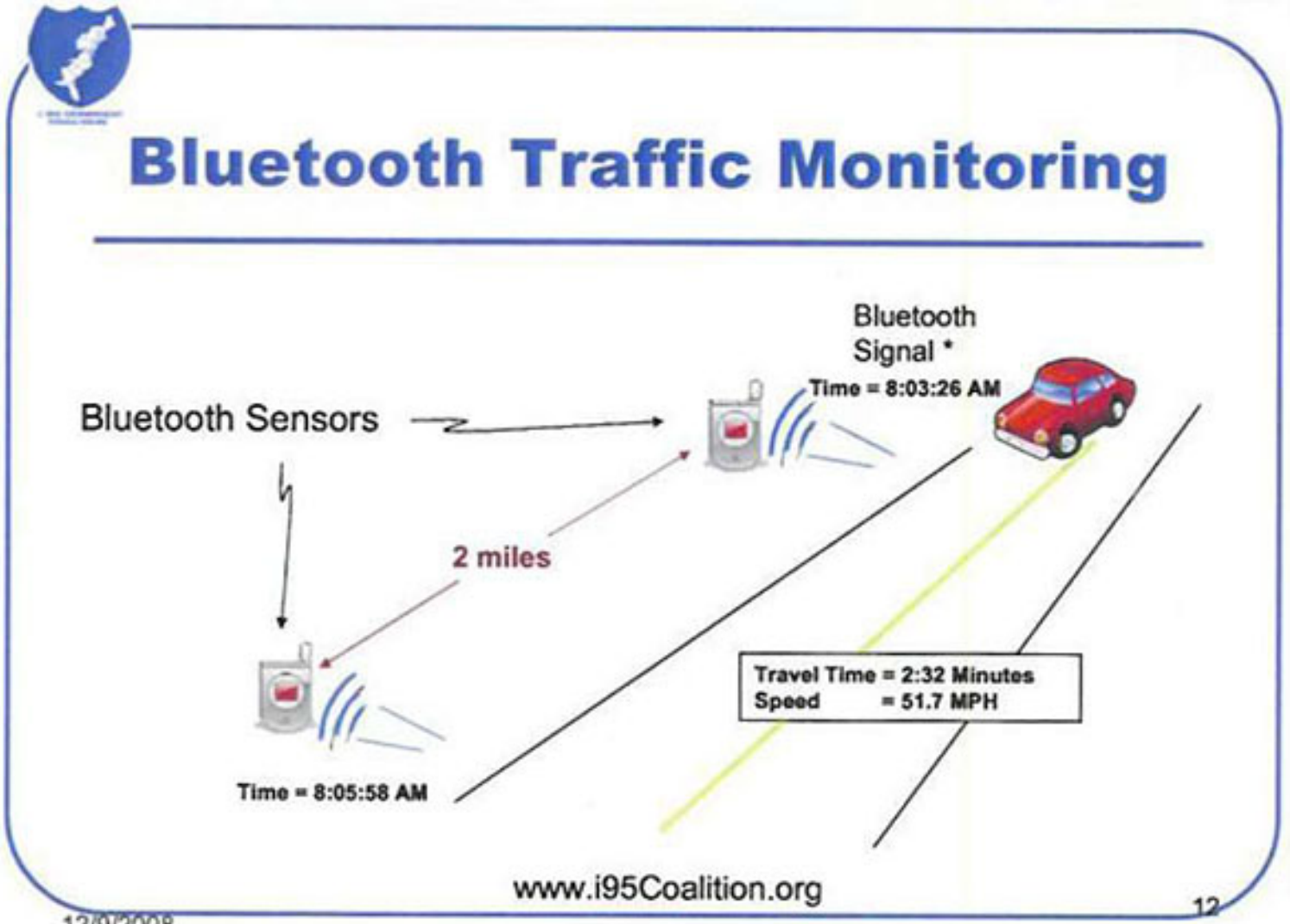
"Sumner Tunnel - bicyclist heading to Boston in the right lane per Boston Police."



Road RTTM


Waltham
Boston
Newton
90 (Mass Pike) East
Right click to start a route here.
Location
MPH
I-95 South
26
Ramp From I-95 South
20
95 Connector (1.6 miles past)
16
Ramp From MA-16 East/Washingto
16
Centre Avenue (1 mile before)
16
Centre Avenue
16
Saint James St / Washington St
16

Road RTTM


Bluetooth Traffic Monitoring
Bluetooth Signal *
Time = 8:03:26 AM
Bluetooth Sensors
2 miles
Travel Time = 2:32 Minutes
Speed = 51.7 MPH
Time = 8:05:58 AM
www.i95Coalition.org

Road RTTM

- 325 MB of average speed estimates at 5 minute increments for 167 road segments between June 2012 to October 2013
- Detectors pick up discoverable Bluetooth MAC addresses* from passing vehicles
- An algorithm processes raw data and produces an estimated travel time on road segments.

**RTTM collected MAC address data that is stored by MassDOT after calculating the travel time will be automatically translated to a non-repeating series of random numbers in order to eliminate any connection between the MAC address and the owner of the device. The MAC address information is not stored. This will enable MassDOT to continuously look for ways to improve the algorithm using historical (paired data) between Bluetooth readers and at the same time ensure privacy standards.*

Road RTTM (cont)

- pair_id is a specific pair of detectors in the field and represents a specific roadway segment.
- Three files:
  - massdot_bluetoad_data.zip: raw data
  - pair_definitions.csv: descriptions of road segments
  - pair_routes.xml (snapshot of http://www.massdot.state.ma.us/feeds/traveltimes/RTTM_feed.aspx)
- Contains geographic points which comprise each roadway segment

Road Volume

- 1.8 MB of data from 1988 to 2013
- Number of cars / day for certain road segments
  - Roadway name
  - Geographic location

NOAA Weather

- 72 MB of data from 1800s to 2013
- Weather observations for Eastern Massachusetts
  - Precipitation
  - Snowfall, Snow Accumulation
  - Min and Max Daily Temperatures

More details...

MBCR Rail Train Location (Raw)

We also have a 117MB (zipped!) containing one month of position data for all commuter rail trains. Basically it's ten million lines like:

11-08-2013 9:54:19 AM GPRS Comm Interface - Send Vehicle - Vehicle ID:1647 - Location[Operator:910, Workpiece: 212, Pattern:311, GPS:>RPV53656+4263222-07131242007329l2<]

Here is how to interpret the data:

11-08-2013 9:54:19 AM Date/timestamp of receipt of message.

Send Vehicle Ignore any lines in which this text is different.

1647 Unique identifier for train as in physical trainset, or more specifically its control cab.

Location Ignore any lines in which this text is different.

311 Unique identifier for train as in scheduled trip from point A to point B. Corresponds to the number that is publicly published on our schedules. But not exactly. See below.

>RPV53656+4263222-07131242007329l2< Position in Trimble ASCII Interface Protocol (TAIP) format. (Next Slide)

MBCR Rail Train Location (Raw)

Regarding the unique identifier for train-as-in-scheduled trip, there are exceptions. Note that our publicly scheduled train numbers are only unique systemwide if you account for which ones have leading zero's and which ones don't.

For the following train numbers in the data you must add 1-2 leading zero's to match with the published train number.

2,3,4,5,6,7,8,9,10,12,14,15,16,17,18,19,20,21,22,23,25,27,28,29,32,33,34,36,38,40,41,43,45,47,48,51,52,55,56,57,60,61,62,63,64,65,66,67,70,71,72,73,74,75,76,77,78,79,80,81,82,83,84,85,86,87,88,89,90,91,92,93

For the following train numbers in the data drop the last two zero's to match with the published train number.

6100,6200,6300,6400,6500,6600,6700,6800,6900,7200,9400,9500,9700,9800

For the following train numbers in the data add a prefix of the letter P to match with the published train number.

500,501,502,503,504,505,506,507,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,529,530,531,532,533,534,535,536,537,538,539,540,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,582,583

MBCR Rail Train Location (Raw)

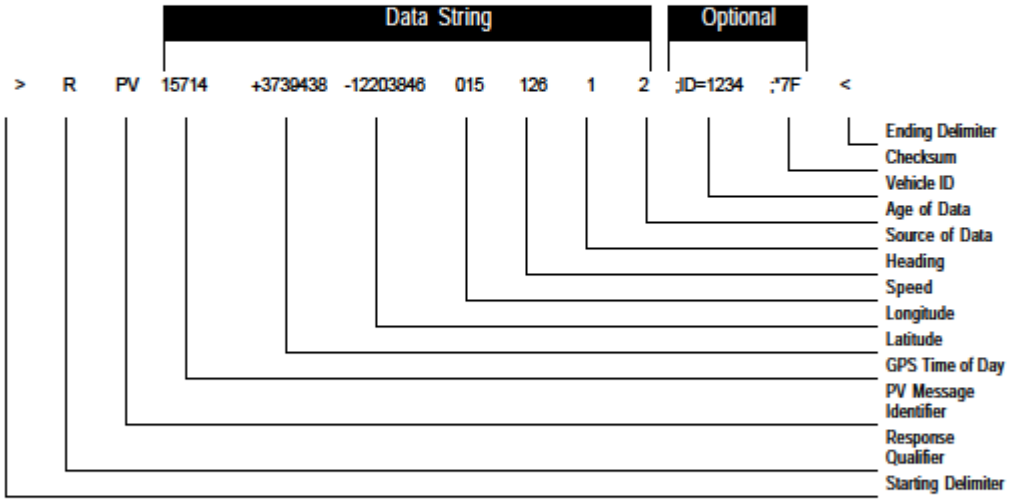
TAIP

Sample PV Message

The Position/Velocity Solution (PV) message is one of the more commonly used TAIP messages and most sensors using TAIP are set by default to output the PV message once every 5 seconds.

The following analysis of a typical PV message is provided to further explain the TAIP message protocol.

>RPV15714+3739438-1220384601512612;ID=1234;*7F<



Data String Information

GPS Time of Day :	15714 seconds, 04:21:54 GPS (time of last fix)
Latitude:	+37.39438 degrees
Longitude:	-122.03846 degrees
Speed:	15 MPH
Heading:	126 degrees
Source of Data:	3D GPS
Age of Data:	Fresh (<10 seconds)

◆ NOTE: *Refer to the discussion of the PV message data string for more detail on how this message is interpreted.*

Good luck!

Let me know how I can help.

I’ll be around all night and all tomorrow morning