RFP Title: Coordinated Intelligent Transportation Systems Deployment in New York City
RFP Number: CIDNY-2013

Responses to Questions from Faculty:

TASK ASSIGNMENT # 8: Develop Data Storage and Access Platform for MTA BusTime Data

Question: Please provide additional detail regarding the task on MTA BusTime Data

Response:
Below is a brief background summary of NYCDOT's work with MTA BusTime data.

· MTA uploads the previous day's data daily (at 2 a.m.) to the Amazon cloud, as a single zipped .csv file – currently approximately 2 million records per day, 200 MB .csv file.

· Weekly, NYCDOT staff manually downloads the data to a shared server at NYCDOT.

· To date, NYCDOT has downloaded nearly 300 days of data files, which is approximately 7 GB archived and 50 GB expanded.

· The data files from MTA include bus, route, direction, phase and NMEA Recommended Minimum Communication (RMC) sentence. Note that the NMEA time field is GMT/UTC and has HHMMSS format; date has DDMMYY format.

· NYCDOT is currently working on an Structured Query Language (SQL) server to import, process, manipulate and store the data as monthly files for querying and analysis purposes. Typical queries for travel time analyses are by route, phase, date and time. These data are used to evaluate travel times, service and routing for projects (pre-/post-) and trends analyses.

· Under this Task, we are looking for the "selected" team to streamline the data access, storage and analysis processes. The team should create an interface that will assist the analysis and reporting of the data:

(1) automate Amazon cloud download and expansion,
(2) automate the bulk import and data manipulation processes,
(3) generate tables of monthly data in the SQL database, and
(4) develop an interface to query the SQL database by route, phase, date and time.

NYCDOT staff familiar with the data can provide guidance in the future on desired data formats to the selected team.