# B2 - SYSTEM BASED PASSIVE DATA STREAMS SYSTEMS: SMART CARDS, PHONE DATA, GPS

ISCTSC 2014 Workshop B2

**Martin Trépanier & Toshiyuki Yamamoto**

**Workshop report**

# Main issues and research opportunities

- Privacy

- Data validation and other challenges

  - In general

  - By type: GPS data, phone, bluetooth data, smart card data

  - How to add socio-demographic information?

- Role of passive data streams in transport planning and modeling

# Privacy and data control

Passive
- Produced by other than our intention (secondary data)
- No user interaction during survey (primary data)

Privacy
- Primary data: Same as active survey
- Secondary data: More difficult problem
  - Legislation on reverse identification
  - Commercial = money (not really privacy)

Data control of secondary data
- No fully control by us. Data provider has
- Needs ethical review by IRB (institutional review board)
- Often no access to raw data and already agregated
- Cost to obtain the data

# General challenges for passive data

- Access to data: privatization or open source
- Processing time for big amount of data
- Imputation of stop, mode and purpose
- Validation especially when the data is controlled by data provider
- Representativeness of the sample
- Disaggregate vs aggregate analysis
  - Behavioral change
- Limited link with socio-demographic characteristics
- How to combine the data from different sources including passive and active survey

# Challenges related to phone data

- We do not know the sampling scheme
- Different spatial precision: Level of antenna, location area
- Translation of phone record to derive trip
- Unit of trip: phone $\neq$ person $\neq$ car, bus
  - Multiple devices to one person

# Challenges related to bluetooth data

- Multiple signals from one person, one car, one bus
- Determine minimum sample size
- Difference between travel speed and running speed
- Difficult to calculate the traffic volume

# Challenges related to smartcard data

- Data management: external information (e.g., bus route)
- Definition of trip: fare system definition is different from ours. Multiple cards for one person
- Fare evasion, non-smartcard users
- Distinguish type of unregistered users: visitors and tourist
- Exact origin and destination identification
  - System without tap-off has more problems
- Problem with detecting real changes (card replacement)
- Not obtained online -> Limited use for real-time operation

# How to add socio-demographic information?

- Induced residential location
- Observed travel pattern: trip frequency, time of day
- Registered card, fare type
- EMV payment by phone
- Combination with active survey (semi-active)

*Legal restriction might prevent*

# Role of passive data streams in transport planning and modeling

- Enhance the understanding of travel patterns (demand) and service level (supply)
- Less time and cost for analysis

Demand side

- Behavior dynamics, distribution, trend
- Social network analysis (e.g., Tweets, ...)

Supply side

- Real-time operation: Special event, incident, etc.

Added-value

- Giving marketing opportunity to data provider
- Detective use of data if court admits

# Thanks to authors & participants

- **Marcela Munizaga**
- **Ka Kee Alfred Chu**
- **Patrick Bonnel**
- **Yusuke Hara**
- **Tim Spurr**
- **Robert Chapleau**
- **Daniel Piché**
- **Salah-Eddine Bouterfif**

- **Martin Trépanier**
- **Catherine Morency**
- **Toshiyuki Yamamoto**
- **Robin Lovelace**
- **Qian Ge**
- **Daisuke Fukuda**

**And thanks to the 24 workshop participants!**