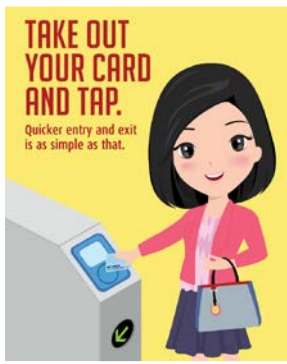# Smart card data use in transport modelling research

Use of AFC data in transportation modelling

Automatic Fare Collection data

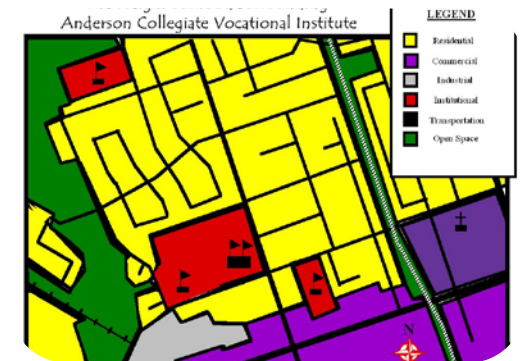Automatic Vehicle Location data

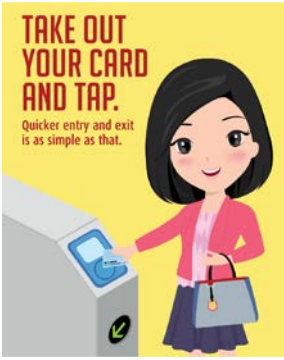General Transit Feed Specification data

OD matrix estimation

Trip chains detection

Vehicle load profiles
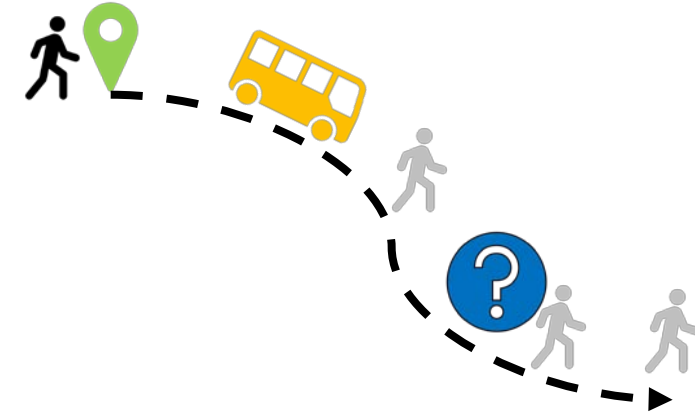
...

Land Use data

Data sources

Automatic Fare Collection data

Series of records, containing information such as:
- User id
- Id of tapping station (vehicle terminal, system ingress or egress point…)
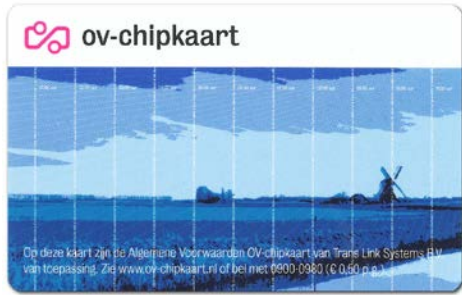- Timestamp
- Type of event
- Card type

Tap-in, tap-out

Tap-in only

Fare calculation

Nationwide

Regional

By operator

Aggregation degree

Data sources

**Automatic Vehicle Location data**

Series of records, containing information such as:
- Vehicle id
- Line id
- Service id
- Timestamp
- Type of event, such as: start of a trip, doors opening, re-assigning…
- Trip id
- Deviation from scheduled time

General Transit Feed Specification data

Standardized format to exchange transit schedules and their associated geographic information
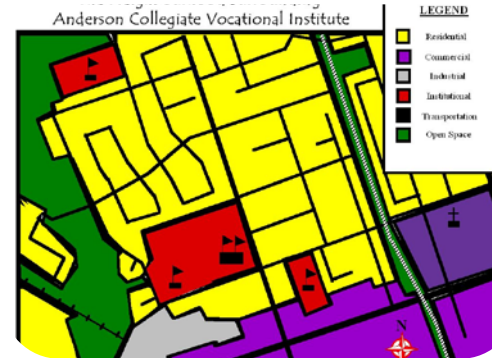
Land Use data

Geo-positioned socio-economic data

Some common issues

Ambiguous ids for
some elements

Missing timestamps
for some events

Multiple timestamps
for the same event

Trip ID
inconsistency

Uncertainty in trip
cancellation/execution

Lack of information
to match rides to
recorded vehicle trips

Lack of information
to match rides to
scheduled trips

Indirect availability of line
information about stops

Erroneous
check-in/out
records

Goals

Develop a series of tools that can be useful to other researchers while working with AVL data. These would be provided as a Python library, defining a standardized way to enter the information, and a series of helper functions to make dealing with the most common issues simpler.

We make heavy use of several Open Source libraries, such as:

Open Source toolkit

matplotlib

pandas

numpy

OD matrix and passenger alighting estimation

Building upon the current State of the Art, try different approaches to better model public transportation use patterns.

We are currently considering:

- Improving trip chaining modelling and alighting estimation in tap-in only systems, combining already stablished methodologies with lesser-used data sources in this field, such as land-use data, historical mobility, mobile phones MACs, public transport video feeds...

- Trying different strategies to estimate public transport OD matrices.

- Enhancing trip chaining validation algorithms, to decrease the dependency on survey data.
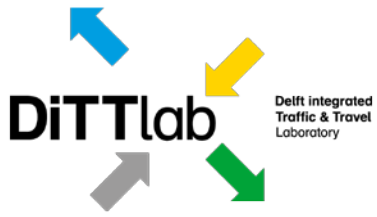
Thanks

juan.benavente@unican.es