MOBILITY IN THE AGE OF BIG DATA

Assoc. Prof. Ivana Semanjski

DATA ANALYTICS 2018, The Seventh International Conference on Data Analytics,
November 18 - 22, 2018 - Athens, Greece

INTRODUCTION

ADDING VALUE TO SOCIETY

Innovation Centre For Intelligent Information Processing

WHAT WE DO

UNIVERSITEIT GENT

HoGent

howest

BY ASSIGNING VALUE TO DATA

Innovation Centre For Intelligent Information Processing

WHAT WE DO

UNIVERSITEIT GENT

HoGent

howest

Spin off companies

Spin off companies

Industrial cooperation and partnerships

DATA COLLECTION

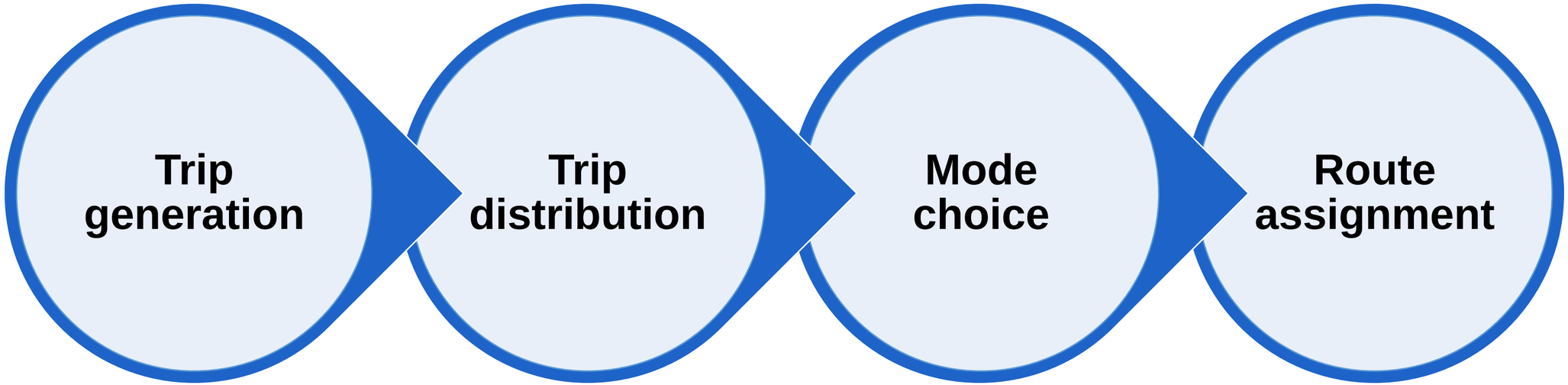
Transportation planning

prediction of usage demand in future travel

ensuring necessary facilities and services to cater to that demand

HOW DO WE DO IT?

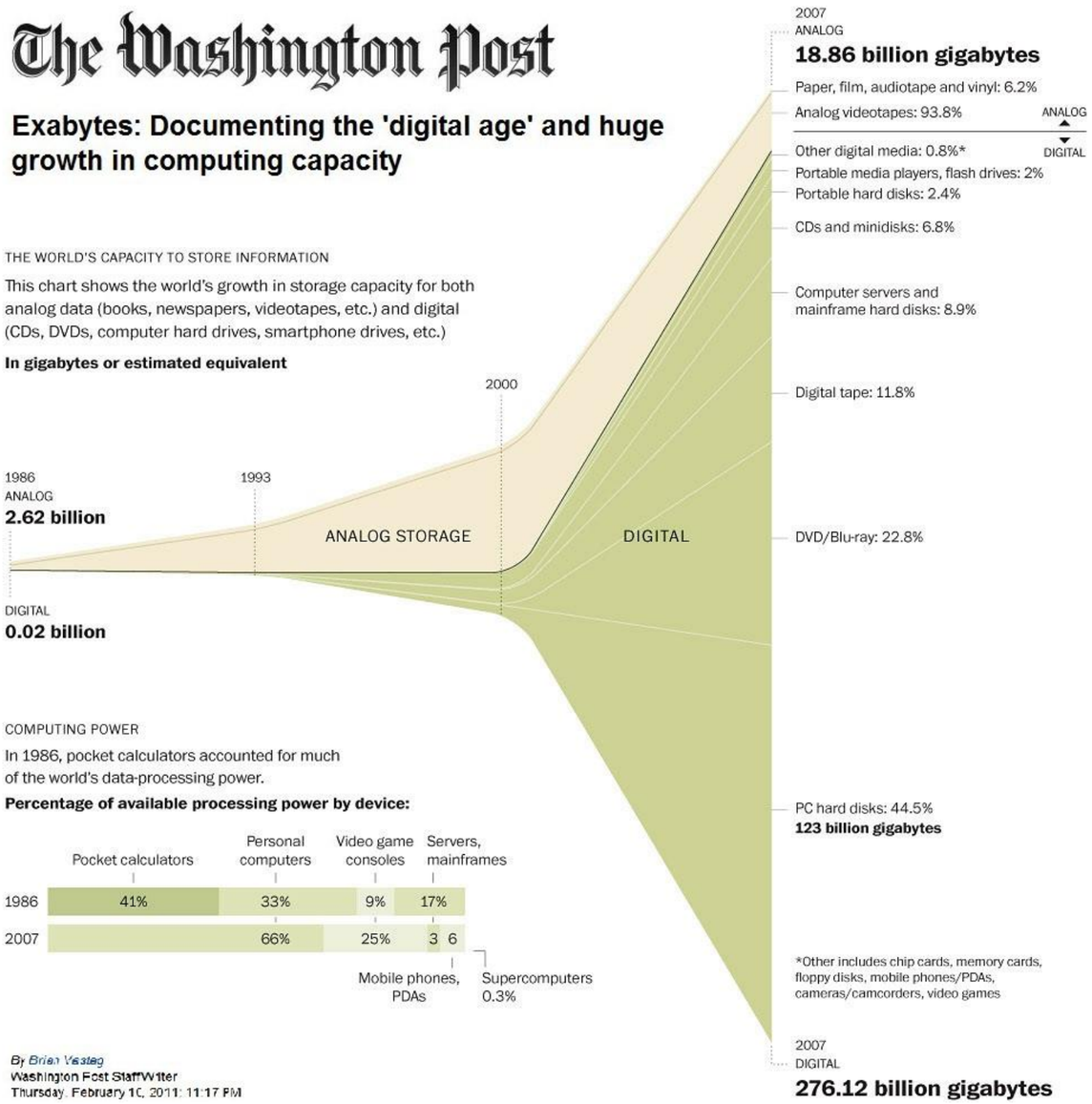
Four-step models

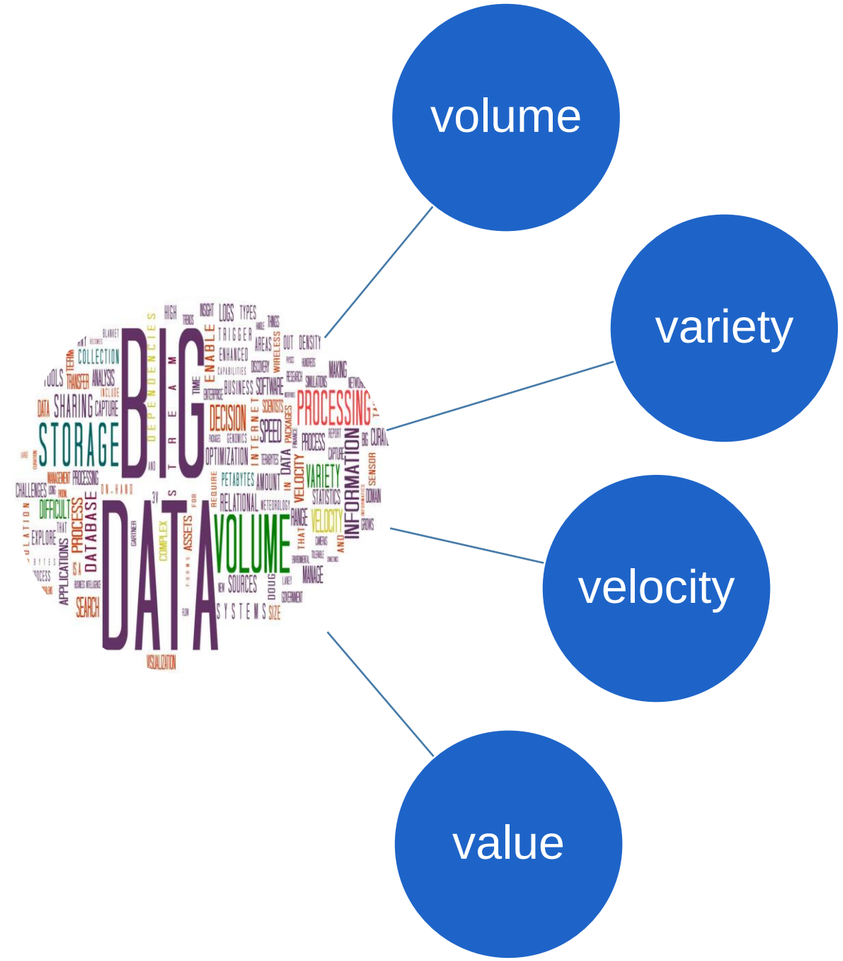


How do we collect data for transportation planning

- Travel diaries
- Interviews


78% of European citizens live in cities
2014 54% of the total global population lived in urban areas
average each person produces 1 GB of content daily
Pizza


The Washington Post
Exabytes: Documenting the 'digital age' and huge growth in computing capacity
THE WORLD'S CAPACITY TO STORE INFORMATION
This chart shows the world's growth in storage capacity for both analog data (books, newspapers, videotapes, etc.) and digital (CDs, DVDs, computer hard drives, smartphone drives, etc.)
In gigabytes or estimated equivalent
1986
ANALOG
2.62 billion
DIGITAL
0.02 billion
1993
2000
ANALOG STORAGE
DIGITAL
COMPUTING POWER
In 1986, pocket calculators accounted for much of the world's data-processing power.
Percentage of available processing power by device:
Pocket calculators
Personal computers
Video game consoles
Servers, mainframes
1986
41%
33%
9%
17%
2007
66%
25%
3
6
Mobile phones, PDAs
Supercomputers 0.3%
By Brian Vastag
Washington Post Staff Writer
Thursday, February 10, 2011; 11:17 PM
2007
ANALOG
18.86 billion gigabytes
Paper, film, audiotape and vinyl: 6.2%
Analog videotapes: 93.8%
ANALOG
DIGITAL
Other digital media: 0.8%*
Portable media players, flash drives: 2%
Portable hard disks: 2.4%
CDs and minidisks: 6.8%
Computer servers and mainframe hard disks: 8.9%
Digital tape: 11.8%
DVD/Blu-ray: 22.8%
PC hard disks: 44.5%
123 billion gigabytes
*Other includes chip cards, memory cards, floppy disks, mobile phones/PDAs, cameras/camcorders, video games
2007
DIGITAL
276.12 billion gigabytes


volume
variety
velocity
value
BIG
DATA
STORAGE
PROCESSING
VOLUME
INFORMATION
VARIETY
VELOCITY
DECISION
DATABASE

Big data sources for mobility studiers

- CDR & signalisation
- GNSS
- Mobile phone

CDR & NETWORK SIGNALIZATION DATA

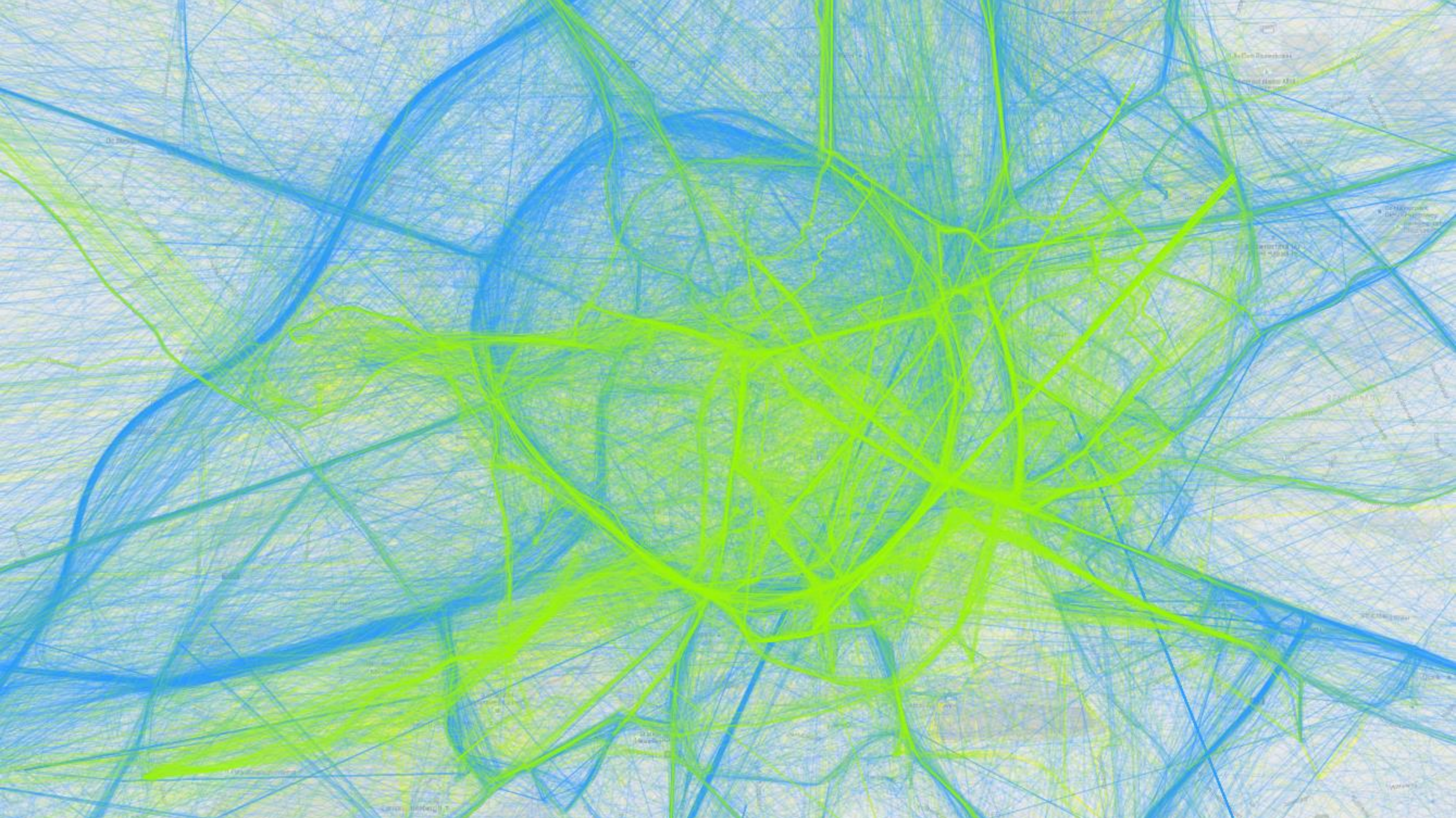
Global Navigation Satellite Systems

- GPS
- Galileo
- Glonass...

Mobile phone data

Passive

Active



TRIP GENERATION

Literature	OD estimation	OD Accuracy	Vehicle occupancy rates	Occupancy rate Accuracy	Duration of test data	Number of users	Dataset	User Validated
(Bohte and Kees, 2009)	No	-	No	-	1 week	1104	GNSS, GIS	Yes
(Munizaga and Palma, 2012)	Yes	**82 %**	No	-	2 weeks	N/A (74 million observations)	Smartcard, GNSS (PT only)	No
.(Shen and Stopher, 2013)	No	-	-	-	3 days	2059	GNSS	No
(Xiao et al., 2016)	No	-	-	-	1 week	321	Smartphone	Yes
(Ge and Fukuda, 2016)	Yes	77%	No	-	1 day	N/A (650,000 observations)	Smartphone	No
(Dong et al., 2015).	Yes	-	No	-	1 day	N/A	CDR	No
(Wolf et al., 2001)	Yes	37%	No	-	3 day	13	GNSS, GIS	Yes
(Lu et al., 2012)	No	-	No	-	**13 weeks**	N/A (3188 trips)	GNSS, GIS	Yes
(Feng and Timmermans, 2011)	No	-	No	-		329		

TRIP DISTRIBUTION

Literature	Duration of test data	Number of users	Dataset	Accuracy
(Fanhas and Saptawati, 2017)	2 months	N/A (16,337 records)	GNSS, taximeter	N/A
(Ge and Fukuda, 2016)	1 day	N/A	GNSS	**88%**
(Moreira-Matias et al., 2016)	**9 months**	441	GNSS, taximeter	79%
(Li et al., 2017)	1 month	12,000	GNSS	80%
(Ma et al., 2013)	1 month	128000	CDR and network signalization data	78%
(Bahoken and Raimond, 2013)	6 weeks	**10 millions**	CDR	45%
(Larijani et al., 2015)	1 day	1.4 millions	CDR and network signalization data	N/A
(Bonnel et al., 2015)	10 days	4,1 millions	network signalization data	82%
(Alexander et al., 2015)	2 months	2 million	CDR and network signalization data	65%
(Gundlegård et al., 2016)	2 weeks	300000	CDR	N/A

MODE CHOICE

Literature	Number of modes	Data	Duration of test data	Number of users	Accuracy	User validated
(Reddy et al., 2008)	3	Mobile sensed GNSS, accelerometer	240 min	6	90	No
(Bohte and Kees, 2009)	4	GNSS, GIS	1 week	1104	70	Yes
(Wang et al., 2010)	5	CDR	12 hours	**56,715**	70	No
(Reddy et al., 2010)	3	Mobile sensed GNSS, accelerometer	24 hours	16	93	No
(Manzoni et al., 2010)	**7**	Mobile sensed GNSS, accelerometer	1 day	4	82	Yes
(Hemminki et al., 2013)	4	Mobile sensed accelerometer	150 hours	16	60-85	Yes
(Biljecki et al., 2013)	7	GNSS, GIS	1 week	1104	92	No
(Xiao et al., 2015)	5	GNSS,	5 days	202	86	Yes
(Zhou et al., 2016)	3	Mobile sensed GNSS, accelerometer	6 days	12	**94**	Yes
(Semanjski et al., 2017)		Mobile sensed GNSS, GIS	**4 months**	8000	**94**	Yes

ROUTE ASSIGNMENT

Literature	Data type	Data Size	Time span	Transport mode	Success rate
(Marchal et al. 2005)	GNSS, low and high-resolution road data	3 cars (2.5 million points)	3 weeks	Car	Up to 10 m/pt
(Yang et al. 2005)	GNSS, road network	1 car (9500 points)	3 h	Car	**100%**
(Krumm et al. 2007)	GNSS, road network (NAVTEQ)	187 cars (1,351,669 points)	2 weeks	Car	N/A
(Chen et al. 2014)	GNSS, road network	**10,245 taxis (172,154 points)**	15 minutes	Car	90-97 %
(Lou et al. 2009)	synthetic GNSS, GeoLife trajectories, road network	unknown for synthetic data, 28 trajectories for real data	N/A	N/A	Real data: up to 85%, Synthetic data: > 95%
(Yuan et al. 2010)	GeoLife trajectories, road network	26 trajectories	30 hours	N/A	66- 84%
(Li et al. 2013)	GNSS, OpenStreetMap road network data	121737 trajectories	**1 month**	Car	85%

WHERE BIG DATA CAN ASSIST TRANSPORT PLANNING

- Data collection

- Complementing traditional models by:

  - Identifying TAZs properties as population, trip generation/attraction
  - Detecting trip purpose
  - Matching OD pairs
  - Detecting transport mode
  - Reconstructing trip paths….

GHENT
UNIVERSITY

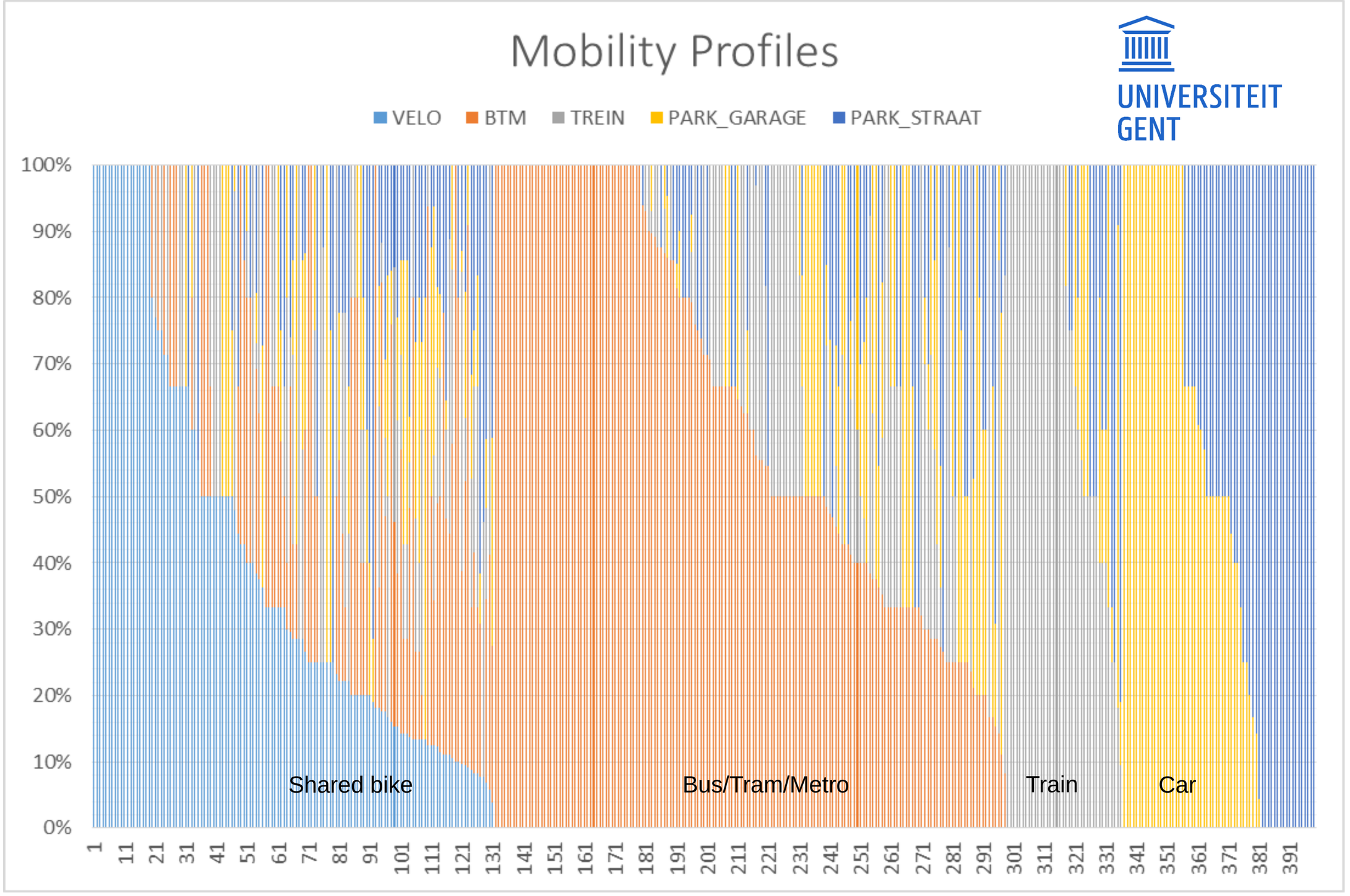

Information on popular pedestrian routes
and where people like to spend time

CONSUMER PROFILING

TARGET GROUP INSIGHT

- **Mobility behavior**
  mode, shift, travel time, ...
- **Customer metrics**
  recency, frequency, duration, ...
- **Spatial insight**
  O/D, visitor flow, POI, ...

GHENT UNIVERSITY


Mobility Profiles
VELO
BTM
TREIN
PARK_GARAGE
PARK_STRAAT
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
11
21
31
41
51
61
71
81
91
101
111
121
131
141
151
161
171
181
191
201
211
221
231
241
251
261
271
281
291
301
311
321
331
341
351
361
371
381
391
Shared bike
Bus/Tram/Metro
Train
Car

Thank you for your attention!!!