Cleansing GPS-data from person based travel surveys in urban environments

Peter Bro

Agenda

- The survey
- Challenges in GPS surveys
- Trip identification

THE SURVEY

Overall goal

- To evaluate GPS surveys as a mean of collecting travel information as a supplement to traditional trip diaries

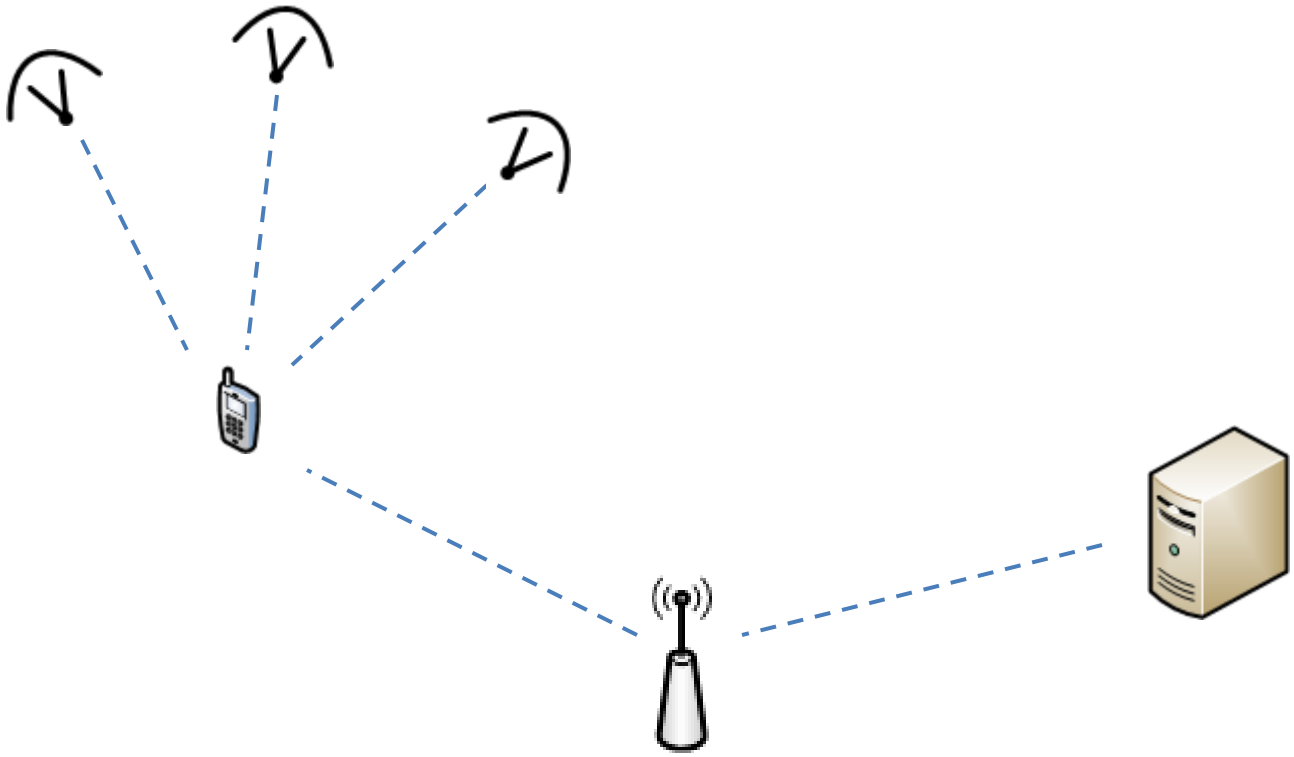
Hardware

På baggrund af flere tests besluttede vi at anvende Flextrack Lommy©, der har både GPS, GSM og GPRS enheder

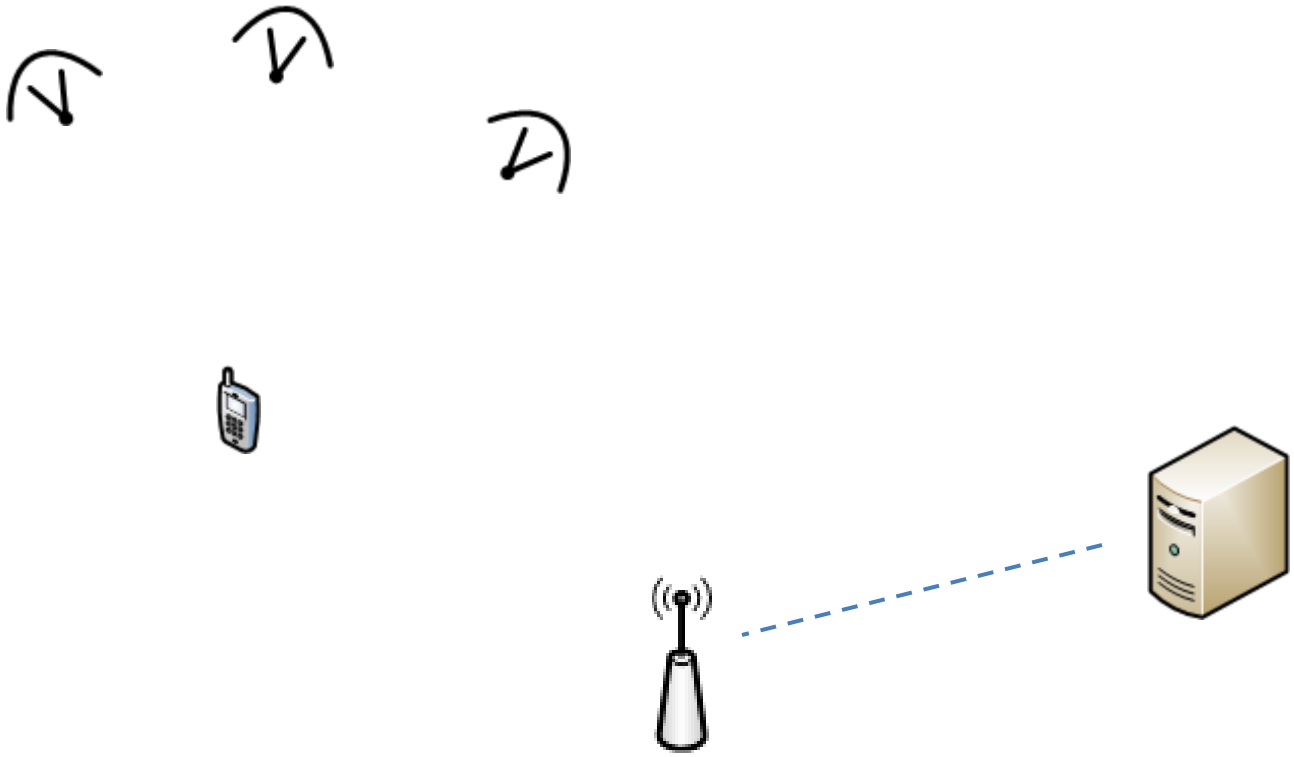
Designet af enheden er simpelt og den er ganske lille (74x61x23 mm og 99 gram) og den har kun én tænd/sluk knap

Lommyen giver desuden mulighed for at følge enheden online og i real-time, sådan undersøgelserne kan monitoreres løbende, og bortkomne enheder kan trackes og indhentes

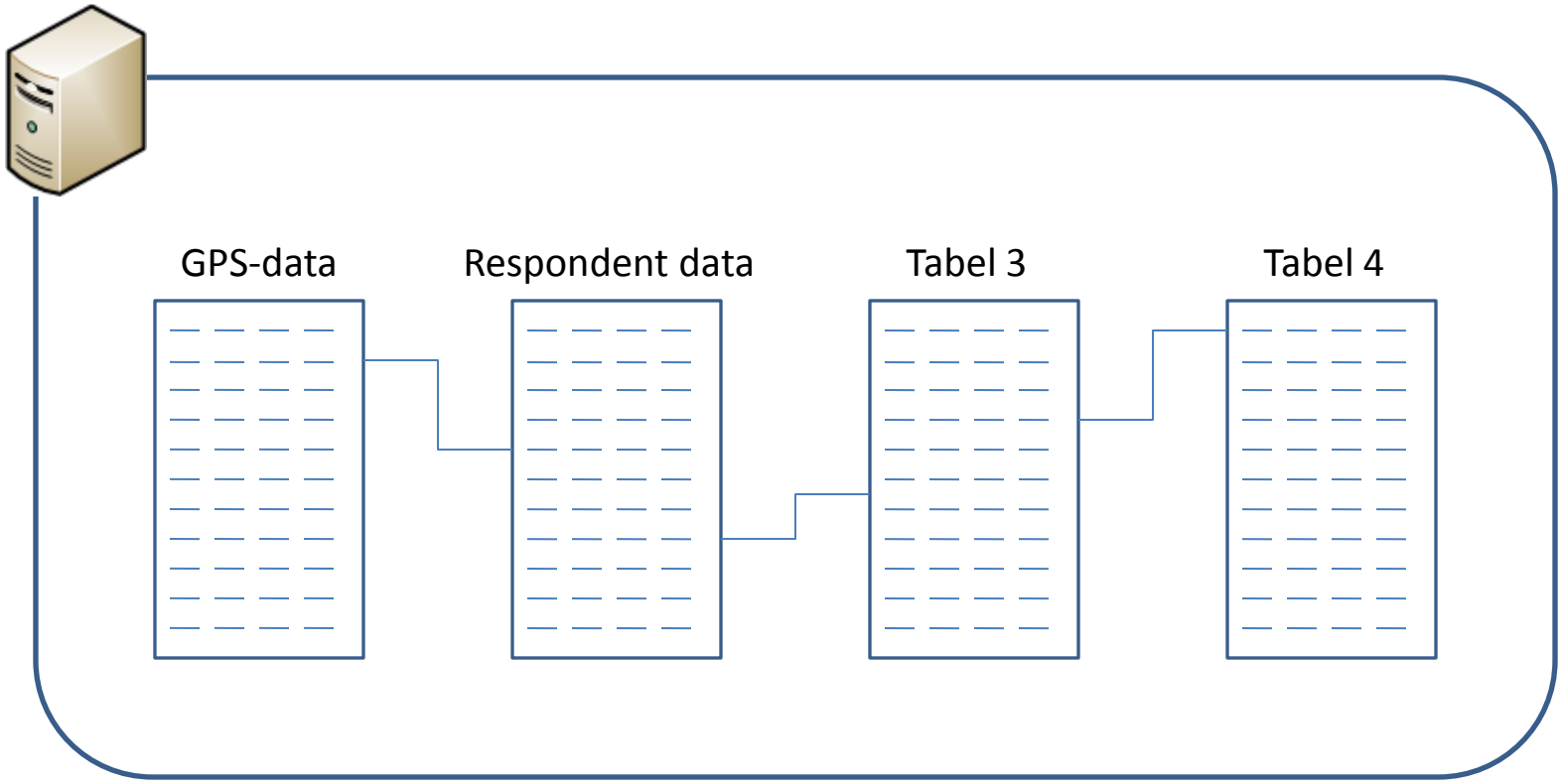
Data flow



Data flow



Data flow



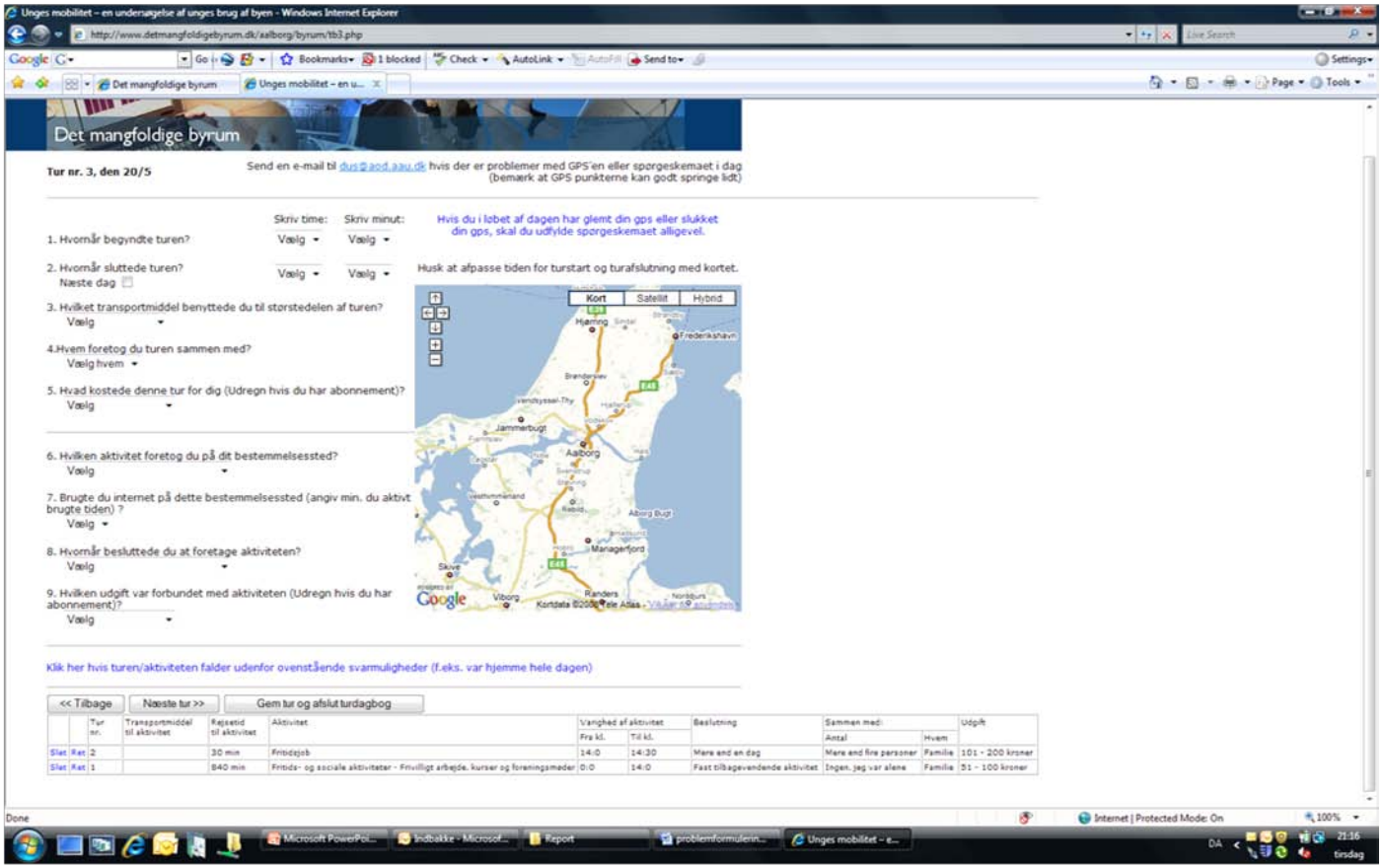
Methodological setup

- 250 young people were selected to participate and carry a GPS and answer a trip-diary each evening throughout a period of seven days
- All participants are students at high school level in Aalborg Municipality
  - 50 respondents from each school at the time
- Data collection were done outside the holidays
  - 4 surveys before the summer holiday
  - 4 surveys after the summer holiday

Methodological setup

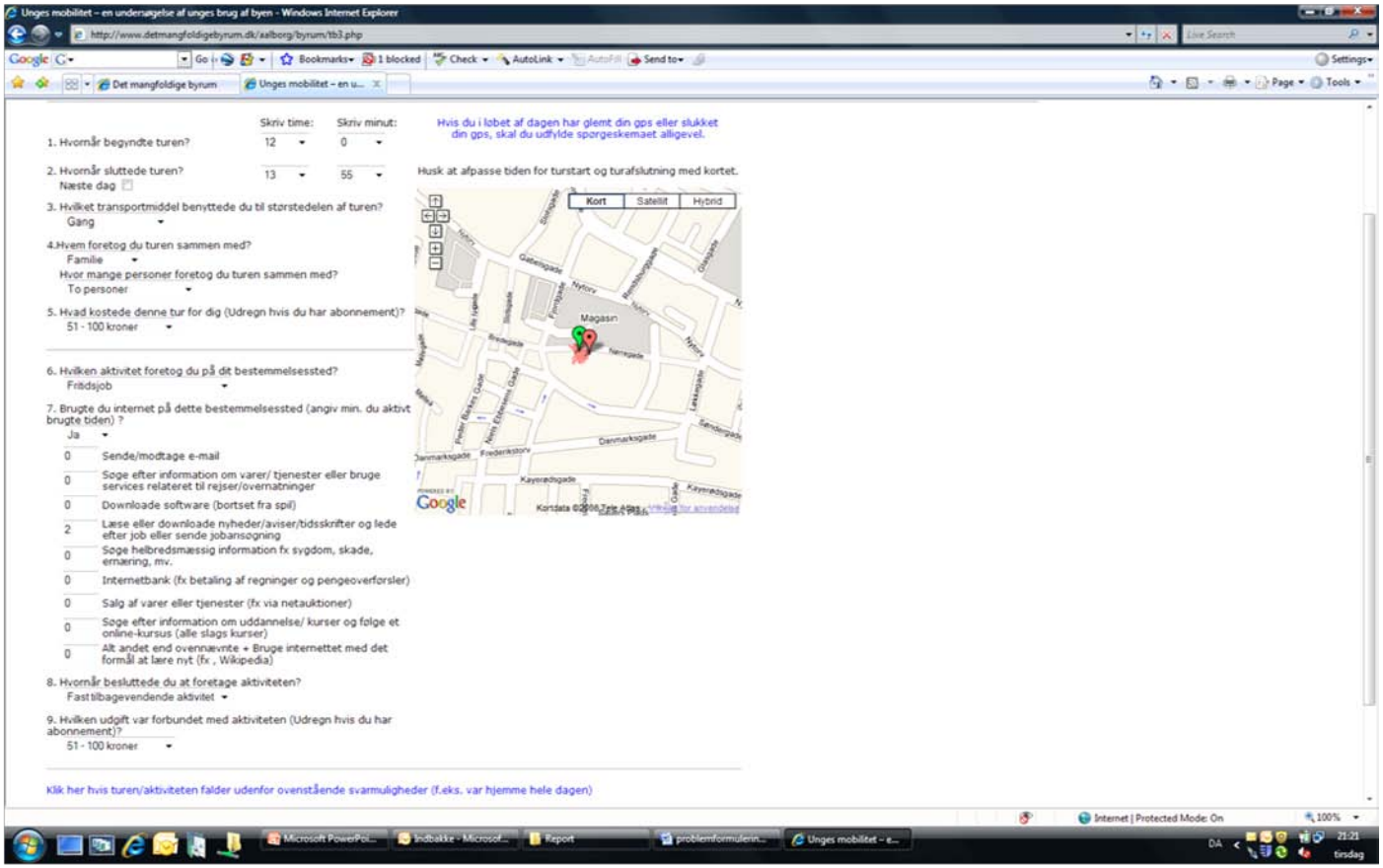
Sending e-mail to all schools in Aalborg → Sampling frame → Calling every potential respondent → Sample → Delivering GPS to each respondent → Data

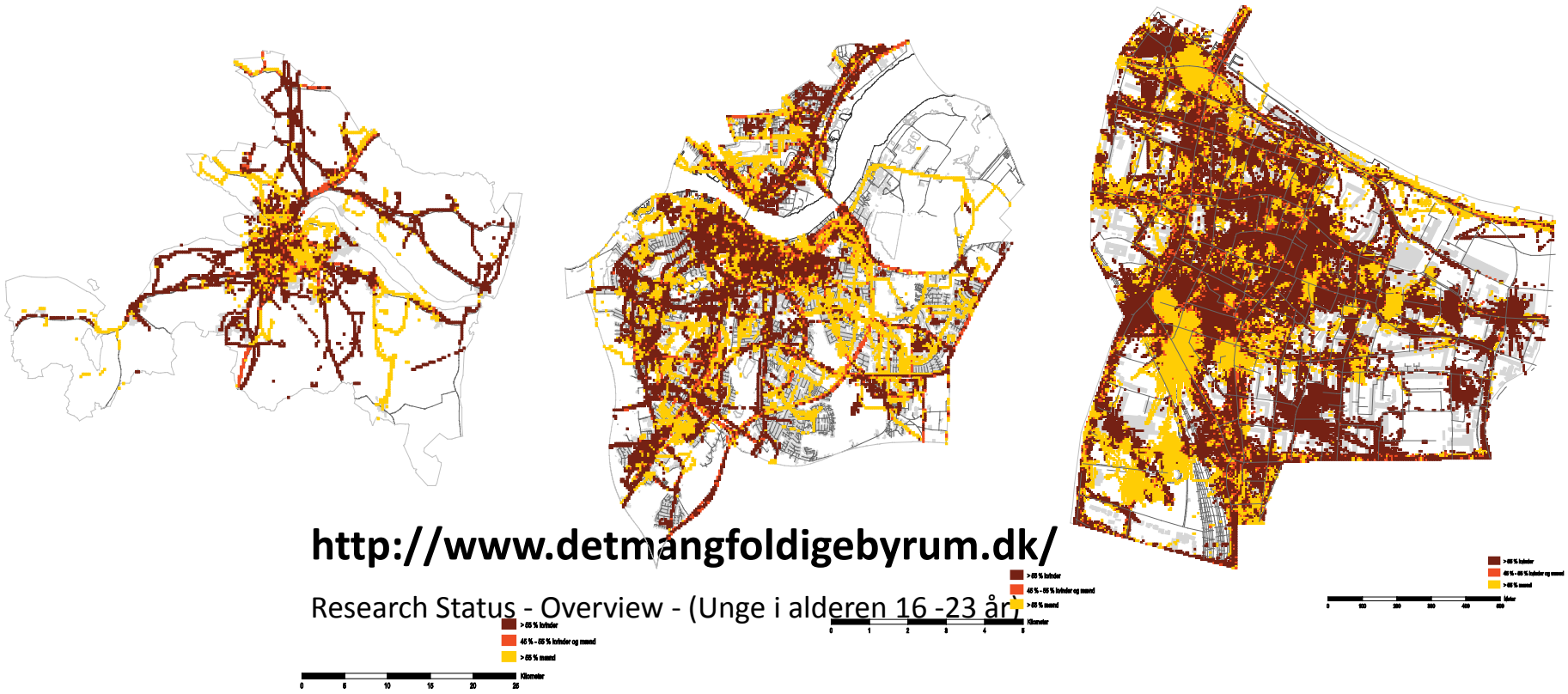
Trip diary


Unges mobilitet – en undersøgelse af unges brug af byen - Windows Internet Explorer
Det mangfoldige byrum
Tur nr. 3, den 20/5
Send en e-mail til dus@aod.aau.dk hvis der er problemer med GPS'en eller spørgeskemaet i dag
(bemærk at GPS punkterne kan godt springe lidt)
Skriv time:
Skriv minut:
Hvis du i løbet af dagen har glemt din gps eller slukket din gps, skal du udfylde spørgeskemaet alligevel.
1. Hvornår begyndte turen?
2. Hvornår sluttede turen?
Næste dag
Husk at afpasse tiden for turstart og turafslutning med kortet.
3. Hvilket transportmiddel benyttede du til størstedelen af turen?
4.Hvem foretog du turen sammen med?
5. Hvad kostede denne tur for dig (Udregn hvis du har abonnement)?
6. Hvilken aktivitet foretog du på dit bestemmelsessted?
7. Brugte du internet på dette bestemmelsessted (angiv min. du aktivt brugte tiden) ?
8. Hvornår besluttede du at foretage aktiviteten?
9. Hvilken udgift var forbundet med aktiviteten (Udregn hvis du har abonnement)?
Vælg
Kort
Satellit
Hybrid
Aalborg
Klik her hvis turen/aktiviteten falder udenfor ovenstående svarmuligheder (f.eks. var hjemme hele dagen)
<< Tilbage
Næste tur >>
Gem tur og afslut turdagbog

Trip diary

Young people's urban mobility





http://www.detmangfoldigebyrum.dk/

Research Status - Overview - (Unge i alderen 16 -23 år)

Henrik Harder, Architecture & Design, Aalborg University,
E-mail: hhar@aod.aau.dk

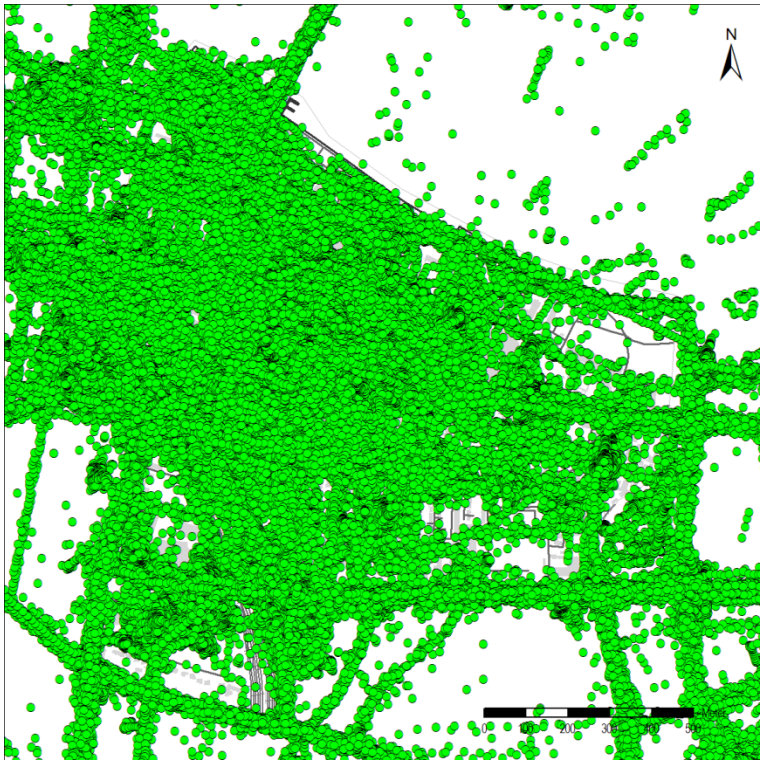
CHALLENGES IN GPS SURVEYS

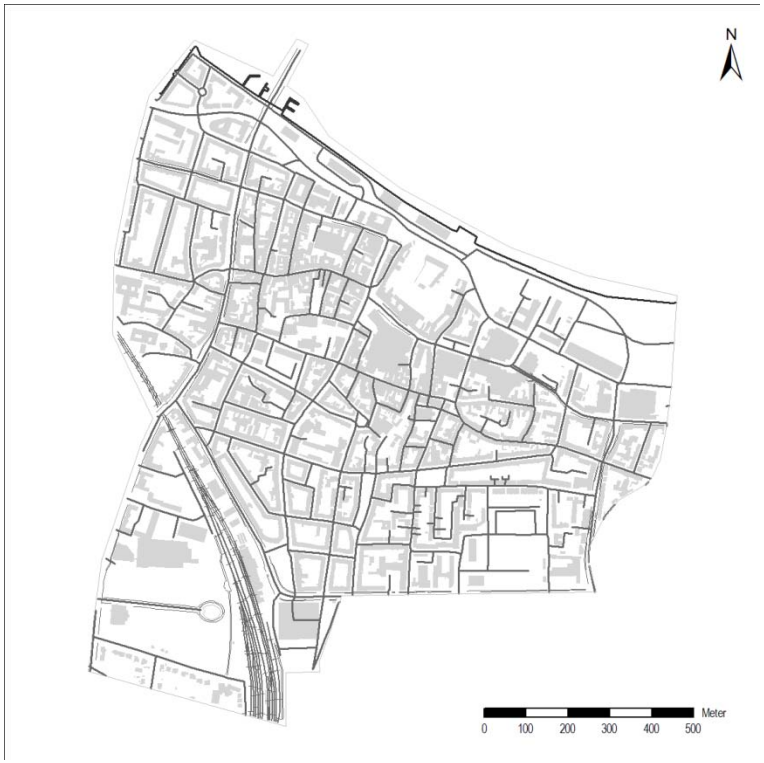
The pros and cons of GPS surveys

- Easy to collect a lot of data
- A possibility to eliminate errors due to limited memory

- Too much data is collected
- Data is hard to interpret and process
- No guarantee that respondents carry the GPS all the time

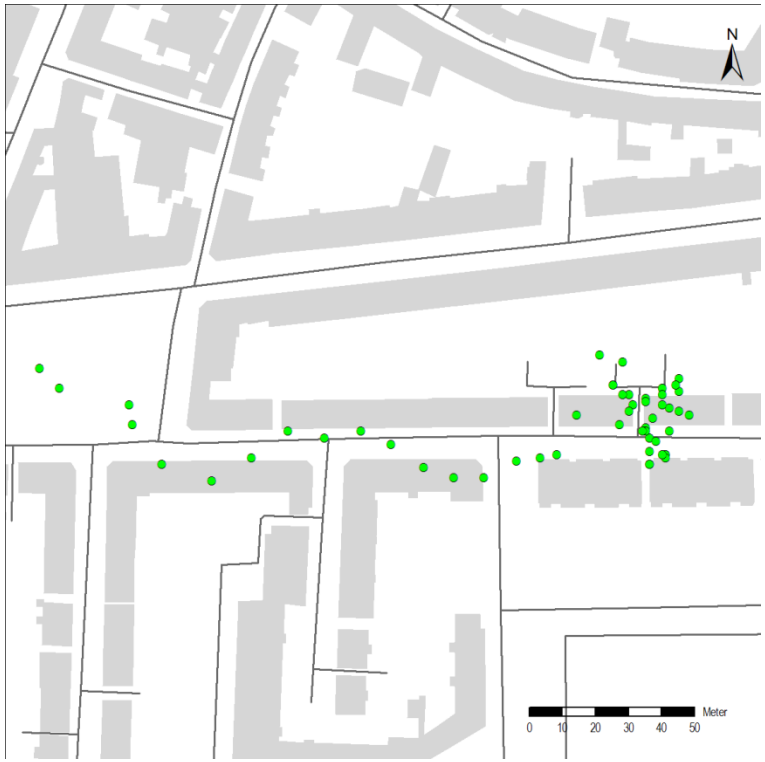
All the data is one big bunch





A closer look

- During trips the loggings are in a nice line
- During stays they scatter

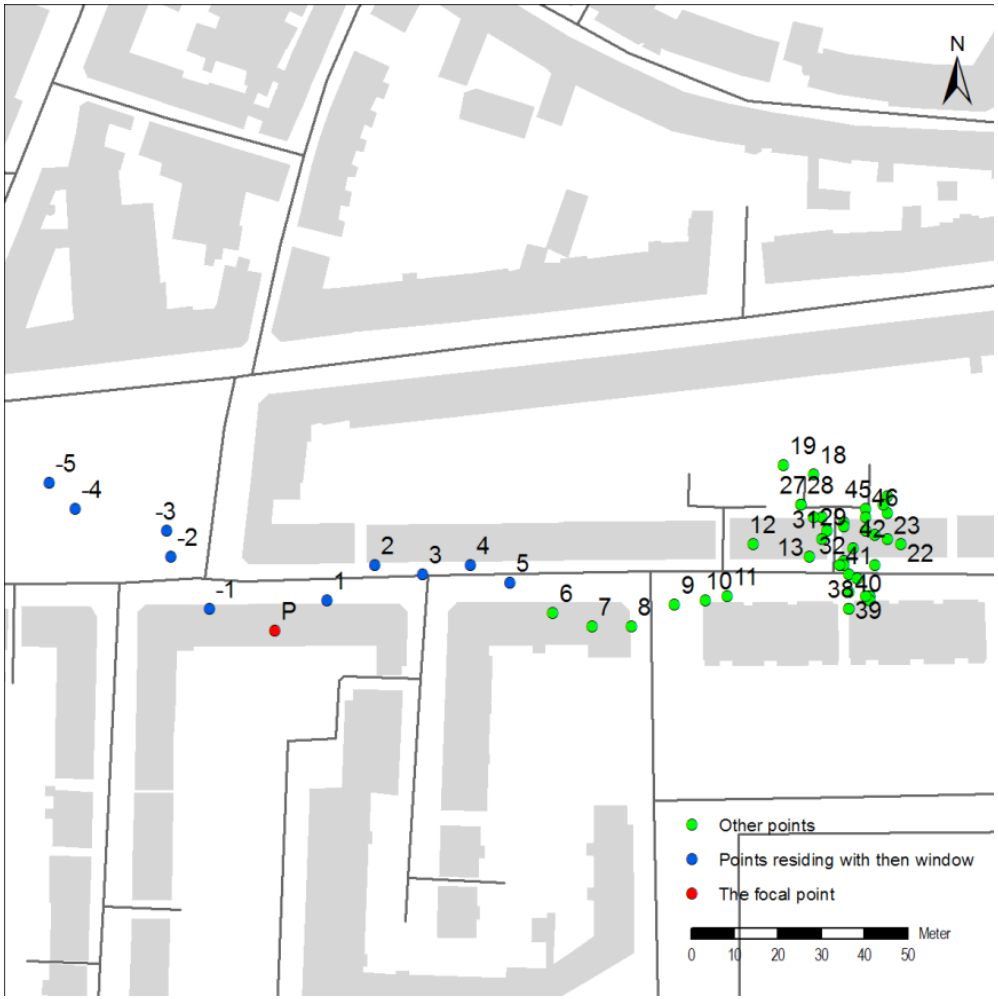


TRIP IDENTIFICATION

Data cleansing

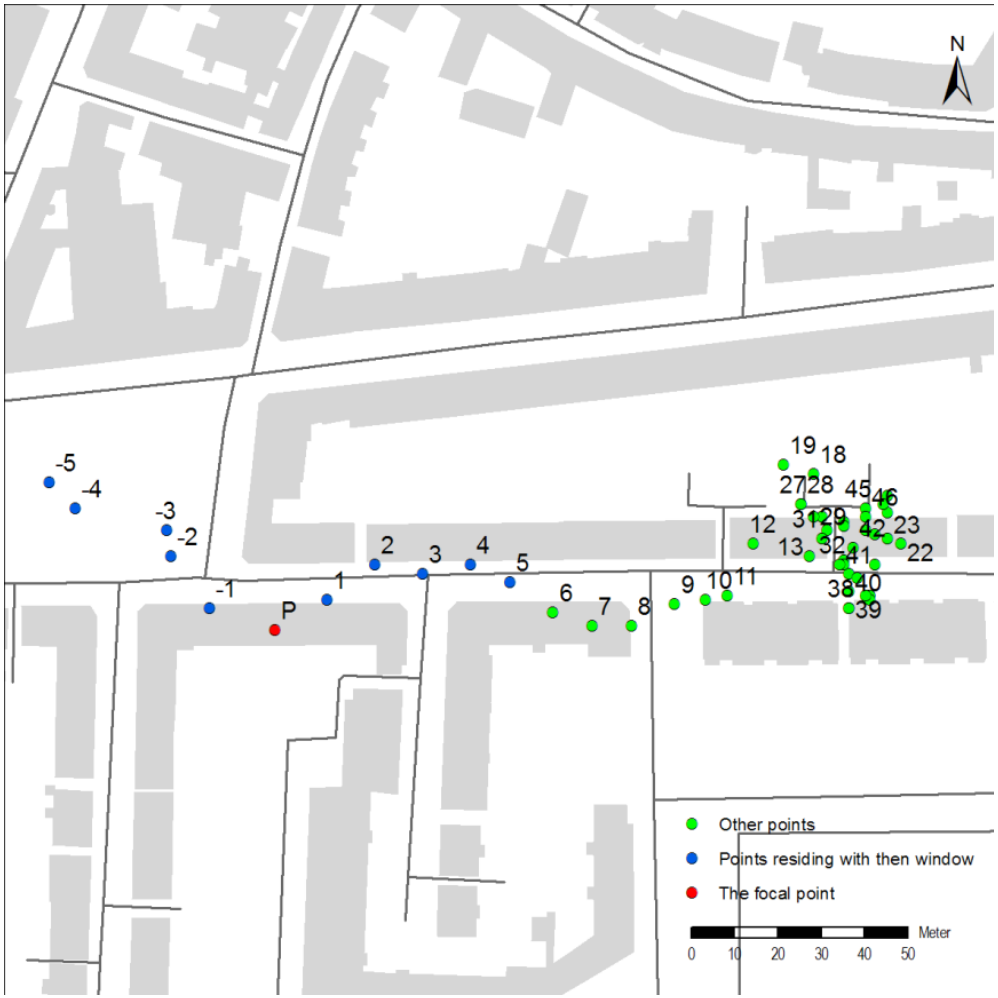
- A sample data set of roughly 120.640 loggings were manually sorted
  - 413 activities
  - 451 trips
  - 24 different respondents
  - 8 different days

The window approach



Point ID	Speed	Direction	Focal point relation
7	4	139	-8
8	3	144	-7
9	4	103	-6
10	7	142	-5
11	3	90	-4
12	0	177	-3
13	0	157	-2
14	5	140	-1
15	3	135	P
16	4	78	1
17	5	79	2
18	4	97	3
19	4	85	4
20	3	98	5
21	5	113	6
22	5	105	7
23	4	98	8
24	4	85	9
25	3	84	10
26	2	110	11

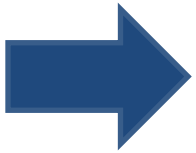
The window approach



Point ID	Speed	Direction	Focal point relation	Direction difference
7	4	139	-8	
8	3	144	-7	
9	4	103	-6	
10	7	142	-5	
11	3	90	-4	52
12	0	177	-3	87
13	0	157	-2	20
14	5	140	-1	17
15	3	135	P	62
16	4	78	1	1
17	5	79	2	18
18	4	97	3	12
19	4	85	4	13
20	3	98	5	
21	5	113	6	
22	5	105	7	
23	4	98	8	
24	4	85	9	
25	3	84	10	
26	2	110	11	

The window approach

Point ID	Speed	Direction	Focal point relation	Direction difference
7	4	139	-8	
8	3	144	-7	
9	4	103	-6	
10	7	142	-5	
11	3	90	-4	52
12	0	177	-3	87
13	0	157	-2	20
14	5	140	-1	17
15	3	135	P	62
16	4	78	1	1
17	5	79	2	18
18	4	97	3	12
19	4	85	4	13
20	3	98	5	
21	5	113	6	
22	5	105	7	
23	4	98	8	
24	4	85	9	
25	3	84	10	
26	2	110	11	



Point ID	Speed	Direction	Focal point relation	Direction change sum
7	4	139	-8	
8	3	144	-7	
9	4	103	-6	
10	7	142	-5	
11	3	90	-4	
12	0	177	-3	
13	0	157	-2	
14	5	140	-1	
15	3	135	P	282
16	4	78	1	
17	5	79	2	
18	4	97	3	
19	4	85	4	
20	3	98	5	
21	5	113	6	
22	5	105	7	
23	4	98	8	
24	4	85	9	
25	3	84	10	
26	2	110	11	

Data cleansing

- Different variables were developed in order to automate data cleansing
  - $\text{avg}(\text{DIRCHN})_{t1,t2}$
  - $\text{avg}(\text{SPEED})_{t1,t2}$
  - $\text{sum}(\text{DIST}_r)_{t1,t2}$
  - $\text{avg}(\text{HDOP})_{t1,t2}$

Data cleansing

$$P(X = trip) = \frac{1}{(1 + exp(1{,}688 + 0{,}278x))}$$

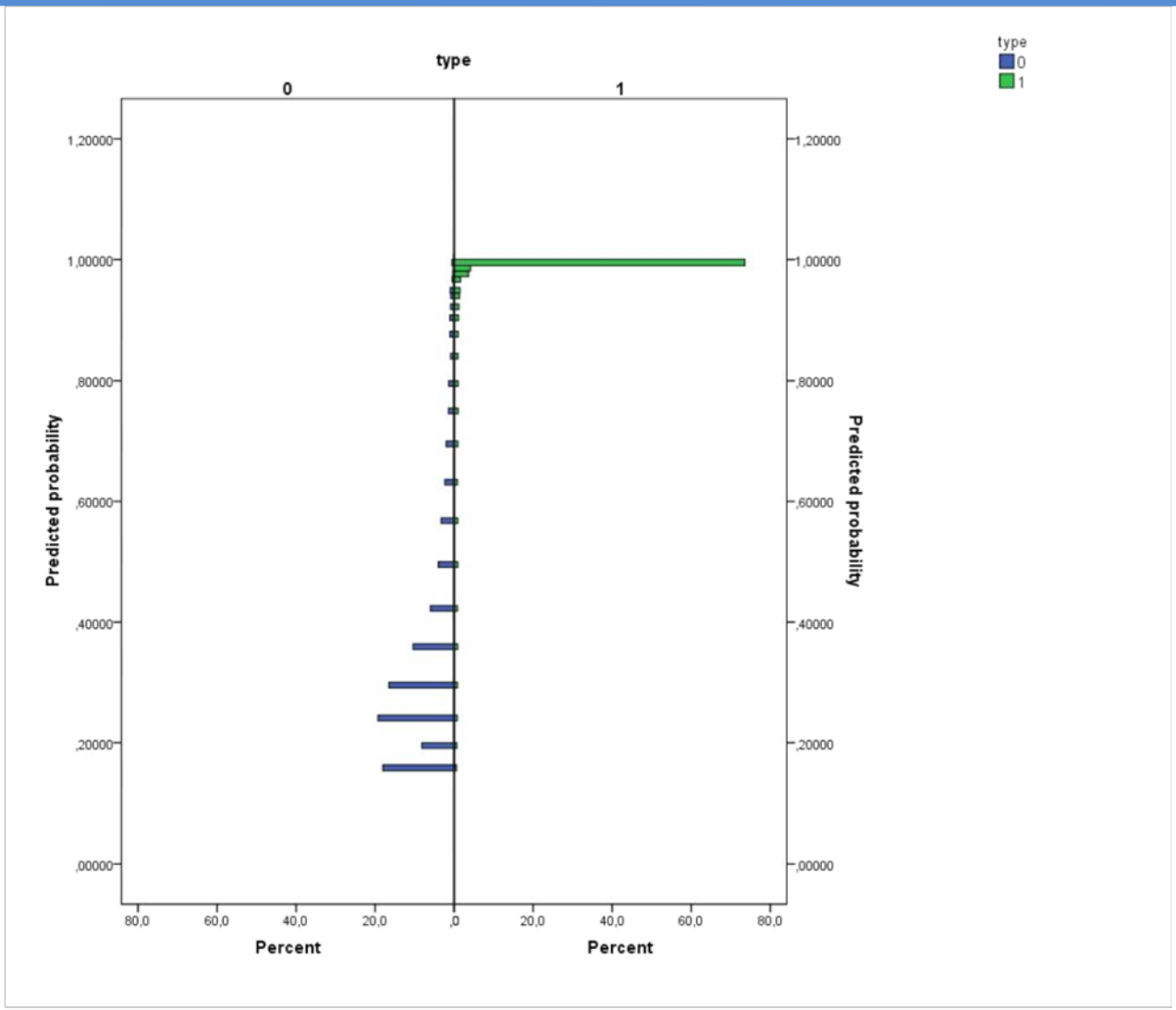
- Where
  - x = the number of loggings within a 20 meter radius from the focal point and registered within the time span of 120 seconds before and after the focal point

Data cleansing

- This classifies 92,8% of all loggings correct
  - 94,6% of the activity loggings are classified correctly
  - 82,3% of the trips loggings are classified correctly
  - Nagelkerke R^2: 0,695
  - Sig: 0,000

Cleansed data

Systematic errors?



Systematic errors?

CONCLUSIONS

Conclusions

- General
  - Collecting travel data with GPS is relatively easy
  - Data processing is time consuming and requires good computational power
  - It is possible to automatically cleanse the data based upon attributes in the loggings
- Specific
  - The developed algorithm classifies 92,8% of the loggings correct
  - It classifies 82,3% of the trip loggings correct
  - The algorithm tends to misclassify the first and the last loggings of trips