



To data warehouse or not to data warehouse...

Marlies Van der Wee Samaneh Bahreini

Mobility Data Forum, May 2016



Rise in mobility data, from different sources, setting different requirements...



Main research questions

- How can the large amount of mobility data be handled and structured?
- What is the best political, economic, social and technical strategy for Flemish mobility data?
- What is the value of this data for businesses, governments and research institutions?



How to handle mobility data?



Categorization of mobility data

- Levels: primary, secondary and tertiary
- Types: structured, unstructured and semi-structured
- Sources: internal, external
- Born digital or born analogue
- Features of the data source
 - Frequency of collection
 - Spatial coverage
 - Equipment requirements
- Usage goals



Categorization of mobility data example

Mobility data		Features of data sources			Usage goal	Туре	Source	Level	Parameters
		Typical collection	Spatial	Temporal					
		Equipment	coverage	coverage					
Traffic	In-road	Inductive loops	Usually	Report at	Incident detection	Structured	Internal	Primary	Speed
flow	sensors	Magnetic induction	spaced at	20-to 60-	Congestion or	Semi-			Travel time
data		coils	<= 1.6 km;	second	queue	structured			Volume
		Magnetometers	by lane	intervals	identification	Analogue			Vehicle
		Pneumatic tubes			Vehicle detection	Digital			classification
		Piezoelectrics							Occupancy
		Weigh-in-motion							
		equipment							
		Truck weighing							
		technologies							



Different requirements for data



- Raw or pre-processed?
- Individual or aggregated?
- Real-time or historical?
- Continuous or sample interval?



Single point of contact

One responsible within the company that manages the available data



Integration from multiple sources



Historical data





X No automated analysis and reporting (hence slow)

X No real-time data





Integrated, centralized website with search engine





Searchable



Both real-time and historical data





X No automated analysis and reporting







Data lake

Large storage repository, no requirements with regards to type or standard "Give me whatever you have"



Massive data sets, multiple sources

🗹 Reporting

Both real-time and historical data











Organized, structured collection of standardized data





Searchable

```
Both real-time and historical data
```













Intelligent database



Searchable, reporting, analysis

Data processing, cleaning, checking

Both real-time and historical data





Different requirements for data need different handling systems



One responsible within the company



Centralized online overview of available data



Quasi unlimited storage, "give me what you have"



Storage and reporting on standardized data



Storage, reporting and analysis on standardized data

Cost estimation of the data handling systems

- Based on
 - Literature review
 - Experts interview
 - BEMES model as a cost calculation tool for technical equipment



Cost estimation of the data handling systems

Comparing the total cost per TB of data per year





PEST for mobility data in Flanders



PEST analysis





PEST: mobility data in general





PEST: data handling systems Cost-benefit trade-Organizational off depending on Economic Political politics characteristics and Responsibilities needs Infrastructure and Socio-Energy footprint of • **Technical** components data centers cultural Amount of data



Value of mobility data









Data requirements

- Standardized
- Searchable
 - Categorized
 - Structured
- Minimal delay (latency)
- Pre-processed
 - Cleaned
 - Missing data points filled
- Processed
 - Analysed







From handling systems to requirements





Quantitative analysis



Find the nearest and cheapest place to park

Outcomes	Types of benefits
Economic	 Travel time saving
competitiveness	 Vehicle operation saving
	 Improve local and regional transit
	 Increased city parking occupancy
Quality of life	 Reduction in the commute time, thus more satisfaction
	 Decrease stress level through providing more convenient
	way of finding parking in the city
Environmental	 Reduction in travel time and thus reduction of emissions
impact	
Safety	 Reduction of vehicles looking for a parking spot within the
	network as result of improved guidance on parking
	availability in the area







Comparison of costs and benefits for Ghent





Quantitative analysis: Proximus analytics



Improved market insights

Get to know your customer better through Proximus analytics



Quantitative analysis: Proximus analytics

- Benefits of data analytics
 - Direct dialogue with your customers
 - Re-develop your products
 - Perform risk analysis
 - Smarter marketing



To data warehouse or not to data warehouse?

depends strongly on the specific case or application...



gen'





Thank you for your attention!

Any questions?

saminbahrini@intec.ugent.be marlies.vanderwee@intec.ugent.be

