WHAT BIG DATA DO NOT TELL US: WHAT WE CAN LEARN FROM TRAVEL SURVEYS FOR BUS AND LIGHTRAIL IN THE NETHERLANDS

Wouter Kuhlman
TU Delft/Panteia BV
Jan Kiel
Panteia BV

1. INTRODUCTION

1.1 Background

With the introduction of the chipcardfor public transport (OV-chipkaart) as a digital ticketing system in the Dutch public transport (PT), a new data source for PT was created. The data registered by this smartcard contain every transaction (check-ins and check-outs) with according time and location. In 2013, the OV-chipkaart system operator Trans Link Systems registered 1.9 billion transactions, which were allstored in the central back office(Trans Link Systems, 2014). This Big Data[1]source contains interesting information for strategic transport models using a pivot point method(Pelletier, Trépanier, & Morency, 2011). This method uses origin-destination (OD) matrices to describe the current situation, on which growth factors are applied to determine forecast matrices.

The OV-chipkaart replaced the National Ticketing System (NVB),with the according WROOV studies. The NVB contained themain ticket types and allowed travellers to travel with all PT operators in the country. The WROOV studies consisted of surveys that were used to allocate the revenues of the NVB to the operators and governments.The WROOV surveys were the primary data source of travel behaviour regarding bus and light rail until 2009, the year the OV-chipkaart was introduced.The data availability on travel behaviour was one of the reasons for PT operators to introduce the OV-chipkaart(Bergmans, Bottenberg, & Hilferink, 2012).In contradiction to WROOV data, however, the OV-chipkaart data havehardlybeen used in strategic planning. This is caused byseveral issues: the availability of OV-chipkaart data is deficient and, moreover, the OV-chipkaart data do not contain all required information(Bagchi & White, 2005).

1.2 Goal

This paper describes the possibilities and limitations of the use of Big Data by means of a case study: the use of OV-chipkaart data in transport modelling. A methodology is presented that combines the strengthsof surveys and Big Data in order to apply the OV-chipkaart datato the construction of base matrices. Additionally, some initial results of data analysis of the WROOV surveys are presented. The paper culminates with a discussion on how to continue this method of data enrichment.

---

[1]There is no consensus on the exact definition of Big Data. Frequently mentioned characteristics are the three V's in the definition by Gartner: high volumes, high velocity and high variety in the data. These characteristics make Big Data hard to process and analyse. OV-chipkaart data comply with all these characteristics.

1.3 Structure of this paper

The paper is structured as follows. Chapter2providesthe opportunities and limitations of the use ofOV-chipkaart data, including the discrepancy between the available andthe required information. Chapter3 describes two alternative sources of information on travel behaviour: the survey data from WROOV and MON/OViN. Chapter4introducesa method of combining these sources to create base matrices suitable for transport models. Chapter5 presents the first results of this method, mainly from analysis of the WROOV data. Chapter6 contains the conclusions with respect to the use of Big Data and the method of data enrichment. Lastly, chapter7 provides a set-up fordiscussion.

2. OV-CHIPKAART DATA

2.1 Availableinformation

Without addressing the technical details of the OV-chipkaart data, the availableinformation can be assorted by directly recorded information and derivable information (see Figure 1).

Basically, all trips made withthe Dutch PT are directly recorded since most of the ticket types are converted to the OV-chipkaart. However, there are some exceptions: in some regions it is still possible to buy paper tickets at the bus driver and forgotten check-ins and check-outs result in incomplete trips. In the data,the amount of incomplete trips is less than 2% of the total trips in The Netherlandsmade withbus, tram and metro (BTM) (Schepers & Zwart, 2014). The OV-chipkaart data are a very rich and accurate information source because of the uniquely high coverage of Dutch PT.

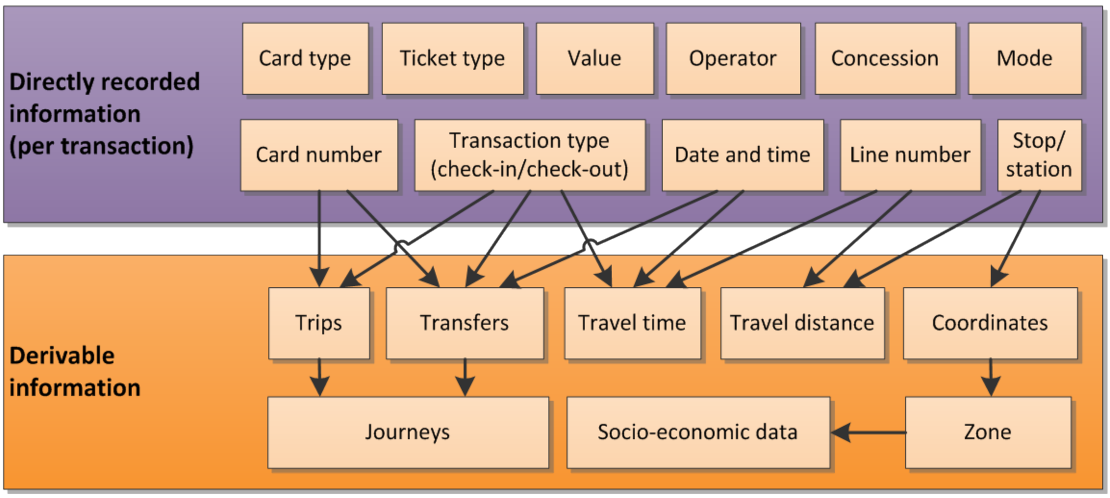


Figure 1: Available information from OV-chipkaart data

2.2 Public availability

Up until now, the availability of OV-chipkaart data for research and policy matters is deficient. This can be attributed to two issues: privacy matters that arise when analysing travel data and the structure of data management. Travel patterns of individuals can be deduced if the data are analysed over a longer period, triggering the sensitive subject of privacy. For this reason, it is prohibited to store the data longer that 18 months. The Dutch Board for

Protection of Personal information (CBP) monitors this act thoroughly, which causes data possessors to be very cautious with their data.

The data management can be partitioned into several levels. This comes down to the fact that operators only have data of their own concession. Authorities responsible for procurement only get strongly aggregated data from the operators, unless specifically stated otherwise. Because the aggregated data from different operators are hard to couple, authorities encounter problems forming a comprehensive overview of all PT in their area. Trans Link Systems the system operator, stores the data of all operators in a central back office. However,Trans Link Systems has been reluctant in sharing data for research purposes.

2.3 Required information in transportation modelling

The input of transportation models consists of network data, socio-economic data, base matrices and counts on cross-sections in the network. The base matrices consist of journeys per OD pair by modefor the base year (recent past). Socio-economic data contain information on the population, the economy and geographical information. Predicted changes in these data are one of the driving forces in the transport prognosis. To model the impact of these changes more accurately, the base matrices are broken down by time of day and by travel purpose (Rijkswaterstaat, 2012). The focus of this case study is on the formation of base matrices for BTM as main mode.

2.4 Discrepancy between available and required information

The OV-chipkaart data are a different type of source than thoseused for input in the current model set-up, like the surveys WROOV and OViN. The OV-chipkaart practically records all trips made in the Dutch PT.The data are unlike regular counts that can be used for calibration: they form a good description of the travel behaviour in the recent past. However, the implementation of OV-chipkaart data in the model set-up is complex. Although the data present an accurate description of travel behaviour, they are not equivalent to base matrices. For the use as base matrices, some key components are lacking.

As explained above, the travel purpose is an important component in establishing base matrices. This information is unknown forjourneys described by OV-chipkaart data. Furthermore, the access and egress trips are not obtainable from the data. These are required for the conversion of OV-chipkaart data into OD matrices, since OD matrices assume addresses of homes, offices or schools as start and end points of a journey (Bagchi & White, 2005). The WROOV and OViN studies are possible sources ofthis lacking information.Information from these surveys can be used for enrichment of the OV-chipkaart data. The enriched data can then contain all required information for the formation of base matrices.

The implementation of this Big Data source is not straight-forward en needs to be coordinated with the model structure. This raises the question what is the best approach for this coordination. Chapter 4 continues on this subject.

3. DUTCH TRAVEL SURVEYS

3.1 WROOV

The WROOV (Commission for Farebox Allocation in Dutch Public Transport) studies started in 1984 to allocate the revenues of the Dutch PT ticketing system to the PT operators, together with the implementation of the NVB. The NVB consisted of several tickets types that were valid with all PT operators in the Netherlands, except most tracks of the National Railways (NS). The allocation of ticket revenues was based on this survey.

In the period 2003-2009 the WROOV studies consisted of a yearly survey. This resulted in a number of annual trips and tours made with the NVB based on 100.000 to 150.000 completed survey forms per year. In 2010 the studies were discontinued because of the replacement of the NVB by the OV-chipkaart system.

The data generated by the WROOV studies contains information of used PT stops and lines, departure times and travel purpose. The survey also included questions on the origins and destinations of travellers, as well as background information like age andgender. The WROOV studies resulted in a large dataset that contains the required information for the construction of base matrices, which is lacking in OV-chipkaart data. Even though the NVB covers the largest part of all trips made with BTM, some other tickets were available to the Dutch traveller: students had special tickets and in some parts of the country regional tickets were available. The coverage of journeys with NVB tickets therefore differs per region. The kilometres travelled can give an indication for the coverage. For the NVB, this lies just above 70% of all kilometres travelled on a national scale (Ballhaus, 2012). Hence, the WROOV data is a valuable and accurate source of information for BTM.

3.2 MON/OViN

The MON/OViN studies consist of a yearly survey to gain insight in the Dutch mobility. This study has been performed since 1978 under several names (OVG: 1978-2003, MON: 2004-2009 and OViN: 2010- present) (Rijkswaterstaat, 2010)(CBS, 2013), and forms one of the longest running mobility study in the world.For this case study, the data from 2003 untilthe present are of interest, implicating both MON and OViN data.

The MON/OViN studies cover all journeys made in the Netherlands, except those made by foreigners. The studies cover all modes. Therefore, the mode BTM is not as prominently present in the data compared to the WROOV studies. The surveys also include questions on origins and destinations of travellers, departure and arrival times, mode and travel purpose, as well as background information of the traveller like age, gender and income. Conversely, the used PT stops and lines are not included in the data. Furthermore, the survey set-up changed with the transition to OViN, causing inconsistency in the data(Wouters & Brakel, 2010).

The MON/OViN data contain the information that is required for the formation of base matrices. The yearly number of observations for BTM, however, is a factor 50 lower than in the WROOV data.This causes problems regarding the reliability when analysing the data on a small scale. On a larger scale, the data give a comparable representation of travel behaviour with BTM.

3.3 Usability of these surveys

Both studies contain valuable information regarding travellers (age, gender) and journeys (purpose, access and egress) that is not availableinOV-chipkaart data. Both sources have pros and cons regarding the usability for this case study. The WROOV data contain information on used stops and lines and are based on a large number of completed surveys. However,the study isdiscontinued, which means that the information is aging. On the other hand, the MON/OViN data consist of a continuous series, which allows for a trend analysis up to present-day. This data does not contain information on stops and lines and has less observations for BTM. This complicates the direct coupling of information onto OV-chipkaart data. For WROOV data, this seems more suitable, since the survey is concentrated on PT. The combination of these surveys holds efficacy for the enrichment of OV-chipkaart data.

4. METHOD FOR IMPLEMENTATION OF OV-CHIPKAART DATA IN TRANSPORT MODELS

The question is: how can these sources be coupled in order to enrich OV-chipkaart data? This chapter describes the proceedings to combine the aforementioned sources, in order to apply the strengths of both OV-chipkaart data and surveys in transport modelling. Because the elaboration of some of these proceedings is not definite yet,possible options are mentioned as input for discussion.

4.1 Which information from which source?

The goal of this case study is to use OV-chipkaart data in the formation of base matrices for the mode BTM. As described in chapter 2, the OV-chipkaart data give an accurate representation of the use of BTM, but some essential information is lacking. The OV-chipkaart can be enriched with information from travel surveys to construct base matrices (see Figure 2). The WROOV studies provide the most fit data for this purpose, while OViN can be used for continuous evaluation of the extracted information.

The OV-chipkaart data provide the number of journeys with BTM between stops by time of day: the boarding matrices. The journeys are then allocated to the existing travel purposes by means of an estimation model. This model is derived from analysis of the WROOV data. Next, the boarding matrices are converted into OD matrices through addition of the access and egress trips. The length of these trips is also estimated based on WROOV data.

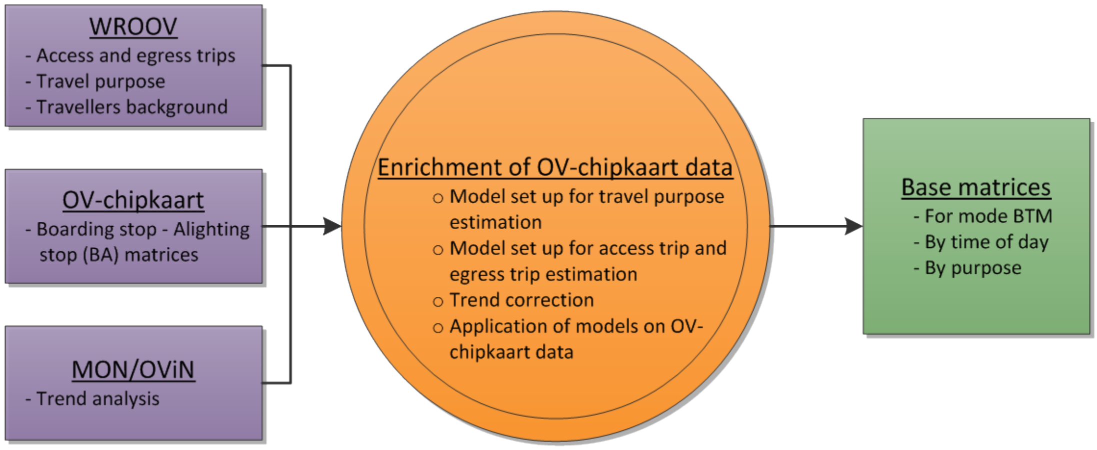


Figure 2: Research concept

OV-chipkaart data can be aggregated over a year to produce stop-stop matrices by time of day for the average working day. In this aggregated form, persons can no longer be identified. So where the original data needs to be deleted after 18 months in order to comply with privacy regulations, the aggregated matrices per year can be stored. The information from the WROOV studies is aging, therefore the validity of this information requires scrutiny. This can be done by means of a comparison with the MON/OViN data series. A yearly actualization can be realized with correction factors based on the OViN data.

4.2 Key variables

In order to combine the information from WROOV with the OV-chipkaart data, an estimation model has to be formulated. This model is based on explanatory variables that are existent in both data sources: the keyvariables. The following keyvariables are available:

- Stops/stations (with derived geographical information)
- Travel distance
- Departure time
- Ticket type

A permutation test can be applied to see if the datasets differ significantly regarding these key-variables. The travel behaviour could have changed over time between the current OV-chipkaart data and the last WROOV study. Depending on the test results, a model can be applied directly onto OV-chipkaart data or correction factors are to be applied.

4.3 Standardization of data

In order to combine different sources, the data have to be synchronized. A point of attention is the analogy of definitions. This concerns definitions such as trips, tours and journeys, but also definitions of explanatory variables like time of day periods.

The data analysis will be executed on journey level, where a journey is defined as a sequence of trips with BTM, without an interruption by a train trip or an activity. The interruption by a train trip is of significance because of the differentiation by main mode. Journeys with BTM that serve as access and egress for the train are excluded from the analysis, since the focus of this case study is on BTM as main mode. Transport models often consist of a separate module for train, including the access and egress for this mode.

4.4 Estimation of relations

Distribution of travel purposes

For the model estimation for the travel purpose, the influence of the key-variables on the travel purpose is determined. A number of different methods of regression can be employed, a frequently used onebeing the *multinomial logistic regression.*This method identifies significant explanatory variables and determines their influence on the chance that a journey was made with a certain purpose. Travel purposes are assigned to journeys according to the probabilities of the different purposes for that journey.

If the explanatory variables have a high correlation, the quality of the regression model can be affected. A large correlation can result in fewer significant variables. In that case, an alternative option is to perform a *principal component analysis* to cluster variables and determine the influence of the newly formed variables with a *generalized linear model.* Both methods result in a model that can be applied to OV-chipkaart data to allocate the journeys to travel purposes.

Access and Egress trips

The estimation of access and egress trips for BTM journeys is done by means of an estimation of the catchment area per stop. Subsequently, journeys can be allocated to the zones overlapping the catchment area of the used stops. These zones are then identified as the origins and destinations of the journeys. The boarding matrices are hereby converted to OD matrices. For the model estimations of the catchment area and the travel purpose, a comparable method can be used. However, a distinction is that the access and egress trip lengths are continuous variables,where the travel purpose is a categorical variable. So instead of a logistic regression, a linear regression can be applied. The influence of the key variables on the size of the catchment area can be determined. A differentiation can be made between access and egress, since these trips are not necessarily equal. For example, a traveller can have a bike available for the access trip, but not for the egress trip.

After determining the catchment areas of stops, these can be displayed onto a zoning map in a Geographic Information System (GIS). Journeys can be allocated to zones, for example postal areas, based on the shares of overlapping area (see Figure 3).

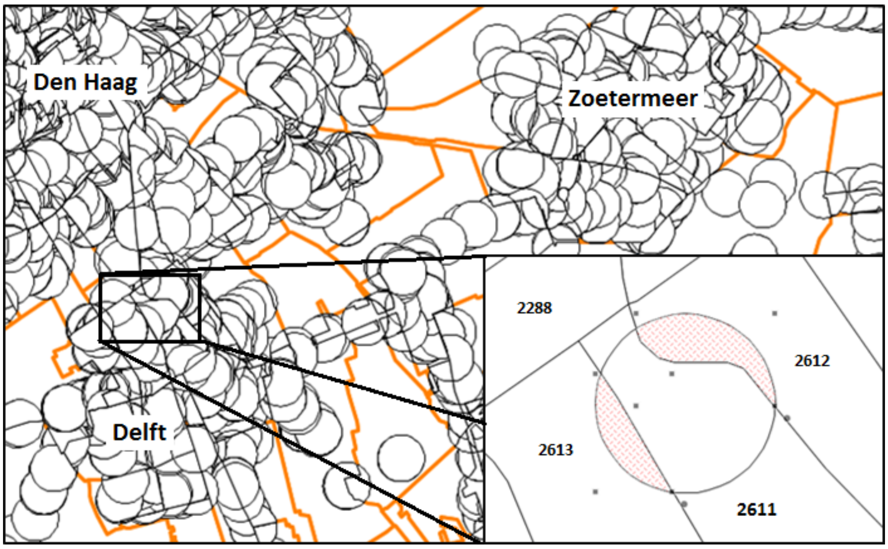


Figure 3: Partition of catchment area of a PT stop in Delft into three zones

A more refined option is to couple socio-economic data to the zoning map and allocate journeys based on this data, such as inhabitants or offices, instead of area. Extended research has been performed on the influence of socio-economic data on travel behaviour. Based on existing literature an allocation model can be developed to convert the stop-stop matrices to OD matrices.

4.5 Implementation of base year matrices in transport models

The OV-chipkaart data supply boarding matrices by time of day. After enrichment with travel purpose, access and egress, these matrices form OD matrices for BTM by time of day and by purpose: the base matrices. The set of base matrices function as input for a transport model.

The newly formed base matrices can be evaluated by means of a comparison with the results from existing methods. Different methods are currently in use in several transport models. Some models do not have base matrices for the mode BTM as input, and consequently neither as output. In this case, the implementation of base matrices is complex. Consequently, the quality of the output (the forecast matrices for BTM) is difficult to evaluate. An indication of improvement can be the reduction of applied corrections in the method.

4.6 Continuity

The value of the estimated relations of the added information possibly decreases due to the termination of the WROOV studies. To ensure continuity of this method, a trend analysis on the MON/OViN data can be performed. This data are broadly comparable with the WROOV data for the period 2004-2009. In case a trend is observed for one of the key-variables, the relation can be updated by means of a correction factor. It is possible to do so annually, based on the OViN results. In that case, both the boarding matrices from the OV-chipkaart data, as well as the enrichment models are updated annually.

5. FIRST RESULTS OF WROOV DATA ANALYSIS

The first stages of this study have been executed. This has resulted in several analyses of the WROOV data from the period 2003-2009. The influence of several key-variables on the travel purpose distribution was examined. The results indicate a clear distinction between purposes through key-variables. Furthermore, a method for allocation of journeys to zones based on catchment areas of stops has been tested in a GIS and a trend-analysis has been performed on the WROOV data for the period 2003-2009.

5.1 Travel purpose

Travel distance

The influence of travel distance on the distribution of travel purposes has been examined, with the underlying assumption that travellers are prepared to make longer journeys for certain purposes. In order to investigate this hypothesis, the number of journeys per travel distance is estimated with a top-lognormal distribution. This method is also applied in mode-choice models, where the assumption is that some modes are preferred more frequently for certain travel distances. For instance, the bicycle is popular for short distances and the train for long distances. Consequently, the chances of a traveller choosing a certain mode are determined for all modes. Travellers are then assigned to modes according to the chance distribution.

The estimation of the top-lognormal distribution was performed in the non-linear regression module in the SPSS software environment. Three parameters are estimated, all with a distinct influence on the distribution shape. The first parameter (α), determines the height of the peak, the second parameter (β) determines the location of the peak on the x-axis and the third parameter (γ) determines the slope of the peak (see Figure 4, left).

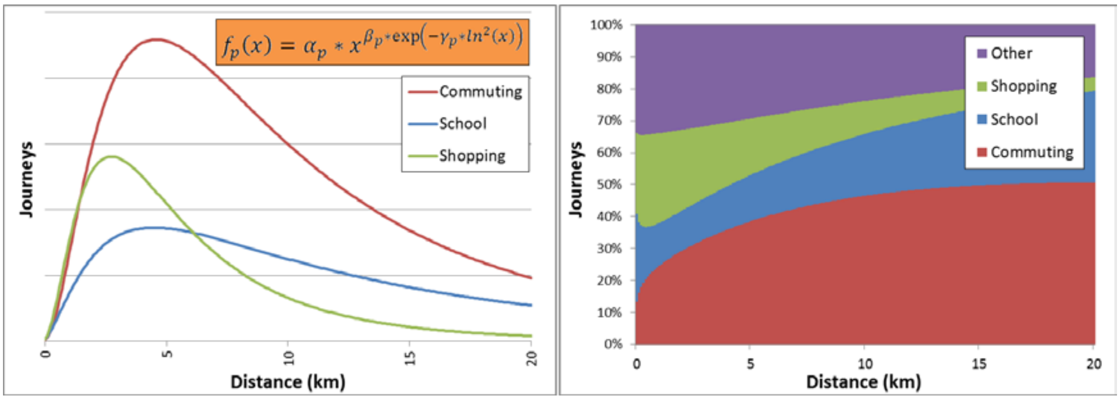


Figure 4: Estimation of the number of journeys (left) and the shares of purposes (right) over the distance

The projected approximations show a clear distinction between the three most observed travel purposes: commuting, school and shopping. The total number of journeys consists of 40% commuting journeys, 20% with the purpose school and 15% shopping journeys. The other five distinguished purposes in the survey account for the remaining 25% of the journeys. These purposes are combined in the purpose ‘other’. The approximation formulas show that the number of journeys over smaller distances (<10 km) are more frequent than journeys over longer distances.

The peak for journeys with shopping purpose, which includes the daily groceries,is observed at lower distance compared to the purposes commuting and school. Additionally, the peak is sharper. Especially for school journeys, the peak is soft. The purpose commuting has the highest peak, as this purpose is the most observed. Theseapproximation formulas result in high shares for journeys with purposes commuting and school for longer distances and high shares for shopping journeys for shorter distances (see Figure 4, right).

Departure time

The number of journeys per hour (see Figure 5, left) clearly depict the peak hours. The peaks are caused by journeys with commuting and school purposes. The morning peak (MP) and evening peak (EP) are fairly symmetrical for commuting journeys. The school purpose shows a more asymmetrical day, which can be explained by the different school rosters. The return journey of students is more spread over the day and in general earlier than return journeys with commuting purpose. The purposes school and other are more frequent in the off-peak hours.

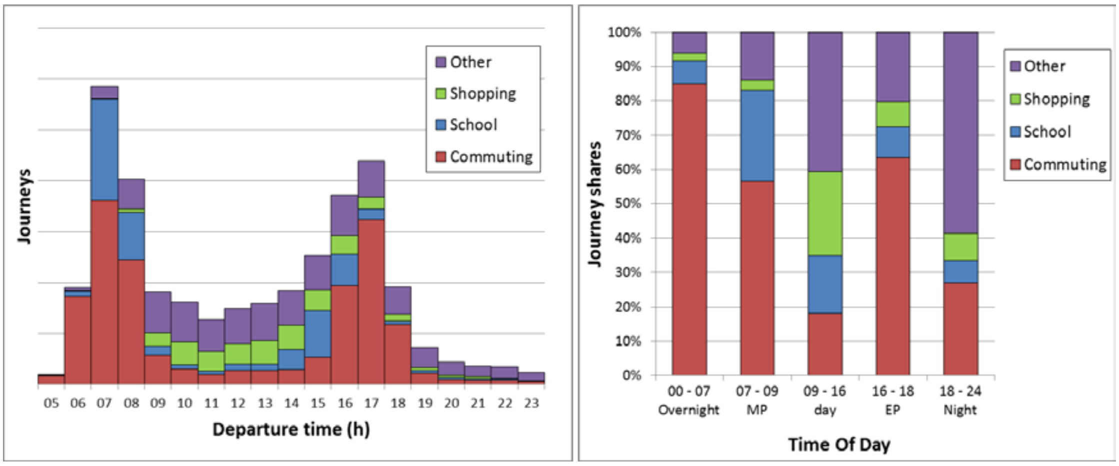


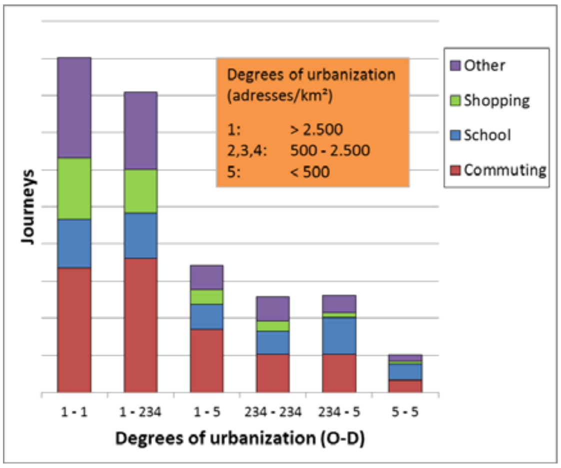
Figure 5: Number of journeys per hour (left) and shares of travel purposes by time of day (right)

Many transport models differentiate between three times of day: the morning peak, the evening peak and off-peak. Within the off-peak period, differences between shares of purposes are noticeable between the day, the night and overnight (see Figure 5, right). The number of observations overnight is low, so it might be a better option to combine this time of day with the night. The peak periods mainly contain journeys with commuting purpose. During the day period and the night period, the purpose ‘other’ contains a large share of the journeys. The aggregation of the remaining five purposes into one category causes loss of information here.

Urbanization

The degree of urbanization is a Dutch classification for the density of addresses. Five categories are distinguished, going from highly urban (1) to rural (5).Both the origin and destination of a journey can be categorized by this degree, resulting in 25 possible combinations. The classes 2, 3 and 4 are combined in order to reduce this number. Also, journeys in opposite directions (for example, from 1 to 5 and from 5 to 1) are combined. This results in six possible combinations for the degree of urbanization of a journey. In addition, the ratio between the most and the least observed category is reduced from 60 to 9.

The largest part of the country is classified as rural area. Most journeys, however, take place in highly urban areas(see Figure 6, left). The journey shares per degree of urbanization show a relation for journeys with origin or destination in highly urban areas. Assuming an origin located in ahighly urban area, the journey shares of commuting and school increase for the more rural destinations, at the expense of purposes shopping and 'other' (see Figure 6, right).



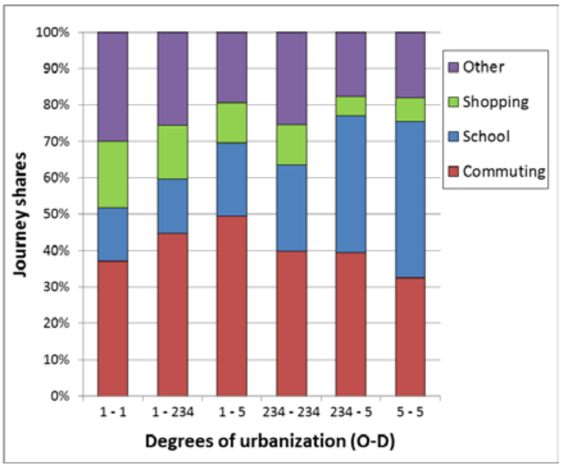


Figure 6: number of journeys (left) and shares of travel purposes (right) by degree of urbanization

Journeys with origin and destination in more rural areas show a different trend. As the degree of urbanization decreases, the share of commuting journeys decreases. On the other hand, the share of journeys with school purpose increases. This is likely to be correlated to the travel distance, which is generally higher in rural areas.

5.2 Access and egress trips

The average access trip length for a BTM journey is 400 metres, measured as the crow flies. For egress trips, the average distance is slightly lower: 377 metres. During 27% of the total access trips, the traveller crosses a postal code border. The origin then lies in another zone than the first boarding stop. Even though the average distance is shorter, the percentage of egress trips crossing a postal code border is slightly higher: 30%. Outbound journeys generally go to more urbanized areas, where zones are smaller than in rural areas.

5.3 Trendanalysis

The travel purpose distribution was also analysed based on WROOV data per year in order to discover possible trends. Overall, the journey shares per purpose are fairly consistent over time. The purpose commuting shows a deviation in the years 2003 and 2009, with 8% below average in 2003 and 5%

above average in 2009. The aberration in 2009 can be explained by the introduction of the OV-chipkaart in the same year. At that time, discount fares for PT subscribers were not implemented on the OV-chipkaart yet. Subscriptions were therefore still paper tickets and remained to be part of the WROOV study, while OV-chipkaart journeys were excluded. Subscriptions are the main ticket type for commuting journeys, while the OV-chipkaart was more frequently used for irregular travel.

In general, the WROOV data appear to be reasonably constant concerning the travel purposes. The introduction of the OV-chipkaart caused a shift in the last year of the WROOV study. A comparison with the MON/OViN data over the period 2003-2009 can improve the comprehension of trends in travel behaviour. The MON/OViN data also allow for the trend analysis to expand up to present day.

6. CONCLUSIONS

Limitations of Big Data
The decentralization of PT policy to local authorities resulted in a fragmentation of information. OV-chipkaart data are scattered across PT operators and accountable authorities. The availability of information suffers from this fragmentation since the sharing of information between these parties is limited. Furthermore, the data processing and storage techniques vary between operators and regions due to the open system structure. The resulting differences in the data complicate the interpretation. The analysis of Big Data takes a lot of time because of the high volumes and high complexity. The usability of the data can be improved by the removal of these preventable complications. Subsequent information processing needs to be taken into account during the formation of a coding template.

Possibilities of travel surveys
The WROOV studies have provided a large database containing the information that is required for the construction of base matrices. The potential of enriching OV-chipkaartdata with information based on this survey appears to be high after first analysis. However, the WROOV studies have been terminated and the data is aging. The validity of the information obtained from this data can be tested by means of a comparison with MON/OViN data. These studies contain less observations, but represent a continuous series up to present day. The combination of the surveys from WROOV and MON/OViN with OV-chipkaart data contains all the required information for the construction of base matrices in transport models.

Application in transport models
Base matrices constructed with OV-chipkaart data can improve the description of the current travel behaviour with BTM. The high number of observations result in more reliable base matrices compared to synthetic matrices that are calibrated with counts. At this moment it is still unclear to what extent this method can improve the modelling of BTM in transport models. The research is in progress. The next stage consists of the formulation of estimation models that enrich the OV-chipkaart data with

information from WROOV and MON/OViN. Subsequently, these models have to be applied on the OV-chipkaart data. Lastly, the resulting base matrices have to be implemented in a transport model to evaluate the quality of the method.

7. DISCUSSION

Public data availability

Currently, the availability for OV-chipkaart data for research purposes is insufficient for use in strategic planning. Operators only share strongly aggregated data with authorities responsible for procurement, according to a specified format (KpVV, 2008) (KpVV, 2011). Operators only have to deliver additional data if specifically stated in the concession terms, which is currently uncommon. These governments therefore do not have all information available for strategic planning, even though it can be argued that they are the owners of the data as authorizing body.

Standardization of data

The open structure of the OV-chipkaart system results in dissimilar data processing and data storage techniques. The resulting variety brings along complications with the interpretation of the data. One collective method of data coding and processing can improve the usability of OV-chipkaart data. In general, this issue plays a role in Big Data since the mining of information is often not the original or only goal.

Continuity of travel surveys

A successive survey has not been implemented after the discontinuation of the WROOV studies. Consequently, no similar source of PT information of the recent past is available. The termination of WROOV coincided with the transition from MON to OViN. Therefore only the OViN study is continuous at the moment, which has a potential inconsistency in the data due to differences in the methodology between MON and OViN. The OViN study can be expanded to gain more insight in travel behaviour in PT. Another option is the establishment of a new study to gather information for enrichment ofOV-chipkaart data. Travellers could, for example, enter their travel purpose, origin and destination in their onlineOV-chipkaart overview. However, this raises the sensitive issues of privacy and online data security.

Continuation of the research

The quality of the estimation models depends on the number of significant explanatory variables. The correlation between these variables could result in few significant variables, limiting the predictive value of the model. On the other hand, many explanatory variables could result in spurious precision due to *overfitting* of the model to the WROOV data. The balance between these extremes is however hard to evaluate, since there are no possibilities for a direct comparison of the results.

REFERENCES

Bagchi, M., & White, P. (2005) The potential of public transport smart card data,*Transport Policy, 12*(5), 464-474.

Ballhaus, F. (2012)*Ontwikkeling reizigerskilometers t/m 2010 K5.* Retrieved December 2013, from www.wroov.nl

Bergmans, j., Bottenberg, W., & Hilferink, P. (2012)*WROOV evaluatie,* Ministerie van Infratructuur en Milieu.

CBS (2013)*Onderzoek Verplaatsingen in Nederland 2012: Onderzoeksbeschrijving*,Rijkswaterstaat, Den Haag

KpVV(2008)*Model Informatieprofiel Openbaar Vervoer(MIPOV)*,Kennisplatform Verkeer en Vervoer, Rotterdam.

KpVV (2011) *Model Informatieprofiel Openbaar Vervoer (MIPOV): Technische uitwerking MIPOV 2008 m.b.t. OV-chipkaart*, Kennisplatform Verkeer en Vervoer, Rotterdam.

Pelletier, M., Trépanier, M., & Morency, C. (2011) Smart card data use in public transit: A literature review,*Transportation Research Part C, 19*(4), 557-568.

Rijkswaterstaat. (2010)*Mobiliteitsonderzoek Nederland 2009 Het onderzoek,* Ministerie van Verkeer en Waterstaat.

Rijkswaterstaat. (2012)*Documentatie Groeimodel 2011 Deel 1,* Rijkswaterstaat.

Schepers, B., & Zwart, G. (2014)*Onderzoek incomplete transacties,* Panteia, Zoetermeer.

Trans Link Systems (2014) *Feiten & cijfers - Trans Link Systems*. Retrieved August 2014, from www.translink.nl: https://www.translink.nl/nl-NL/Persinformatie/Feiten-cijfers

Wouters, B., & Brakel, J. v. (2010) Simulatiestudie naar Methodebreuken in het Onderzoek Verplaatsingen in Nederland,*Colloquium Vervoersplanologisch Speurwerk (CVS),* Roermond.