National Data Warehouse: How the Netherlands Is Creating a Reliable, Widespread, Accessible Data Bank for Traffic Information, Monitoring, and Road Network Control

Article *in* Transportation Research Record Journal of the Transportation Research Board · December 2008

DOI: 10.3141/2049-21

5 authors, including:

Francesco Viti
University of Luxembourg

Serge Hoogendoorn
Delft University of Technology

Ben Immers
KU Leuven

Chris Tampère
KU Leuven

NATIONAL DATA WAREHOUSE: HOW THE NETHERLANDS IS CREATING A RELIABLE, WIDESPREAD AND ACCESSIBLE DATA BANK FOR TRAFFIC INFORMATION, MONITORING AND CONTROL OF ROAD NETWORKS

Francesco Viti

Delft University of Technology, Transportation and Planning Department, faculty of Civil Engineering and Geosciences, Stevinweg 1, P.O.Box 5048, 2600 GA Delft, The Netherlands Tel. +31 15 2786909

Katholieke Universiteit Leuven, Department of Civil Engineering, section Traffic and Infrastructure, Traffic and Infrastructure, Celestijnenlaan 300A - PO Box 2422, 3001 Heverlee, T +32 16 32 1673, F +32 16 32 2986

f.viti@tudelft.nl, francesco.viti@cib.kuleuven.be

Serge P. Hoogendoorn

Delft University of Technology, Transportation and Planning Department, faculty of Civil Engineering and Geosciences Stevinweg 1, P.O.Box 5048, 2600 GA Delft, The Netherlands, Tel. +31 15 2785475 s.p.hoogendoorn@tudelft.nl

Lambertus H. (Ben) Immers

Katholieke Universiteit Leuven , Department of Civil Engineering, section Traffic and Infrastructure, Traffic and Infrastructure, Celestijnenlaan 300A - PO Box 2422, 3001 Heverlee (Belgium). ben.immers@cib.kuleuven.be

Chris M.J. Tampère

Katholieke Universiteit Leuven, Department of Civil Engineering, section Traffic and Infrastructure, Traffic and Infrastructure, Celestijnenlaan 300A - PO Box 2422, 3001 Heverlee, T +32 16 32 1673, F +32 16 32 2986 chris.tampere@cib.kuleuven.be

Sascha Hoogendoorn Lanser

Project leader and ITS specialist, Rijkswaterstaat AVV - the Dutch Transport Research Centre, The Netherlands s.hoogendoorn@rws.nl

Word Count

No. of words:			5000
No. of tables:	2*250	=	500
No. of figures:	9*250	=	2250
Total:			7750

Submitted for presentation for the 87th meeting of the Transportation Research Board,, 13-17 January 2008, Washington D.C. and for publication in Transportation Research Record

ABSTRACT

Every day traffic managers as well as large part of the road users use available information regarding the current state of the road network to improve their decisions. However, the efficiency of these decisions strongly depends on how accurate, reliable and timely is the data they can gather. Moreover, information is typically scattered in space and time, large areas are usually unmonitored and the data collected is often at different quality levels depending on the (part of the) network they represent. Within this view, distribution of a unique dataset that contains sufficient levels of quality over the whole network may strongly improve the way information is provided to the user and management strategies can be more effective.

This paper investigates the requirements that such data bank should satisfy, namely the accuracy and reliability of information, which depend on the spatial location and aggregation time. The impact of these elements has been quantified through theoretical and numerical analysis, showing that both elements strongly affect good estimation and prediction of travel times and network states, especially under variable traffic conditions.

The need for a database that guarantees standard levels of quality of data provided to road authorities and service providers motivated in the Netherlands the establishment of the National Data Warehouse project. The National Data Warehouse has the goal of providing traffic information as well as information on the status of the road network system as a whole. This information is extended to a basic network level, which allows road authorities or service providers to combine this information with their own dataset and obtain a broader view to the problems that occur on the network they manage or monitor. In this paper we describe the current state of the project and we define the future steps that will be undertaken in the following years.

INTRODUCTION

Every day the traffic in the Netherlands produces serious delays to the drivers due to congestion, represented in an average working day by 200km of queues [1]. This problem is partly due to insufficient capacity at some parts of the road network, but also due to an inefficient use of the available road infrastructure. The installation of Intelligent Transportation Systems (ITS) measures and the adoption of Dynamic Traffic Management (DTM) strategies aim to achieve a better utilization of the system in the short term. Reduction of congestion is therefore achieved with these measures by means of a more efficient use of the existing physical capacity, i.e. by affecting the demand (e.g. through information, route guidance systems, pricing etc.) or that modify the capacity dynamically according to the actual need for road space (e.g. signal control, speed control, ramp meters, incident management etc.).

However, the achievement of this objective strongly relies on the knowledge of the current state of the network as much as the prediction power of the adopted traffic models. The quality level of the available data plays therefore an essential role for ITS and DTM applications. Minimum levels of information availability, accuracy and reliability depend on the network as well as on the adopted management strategies. Only with fast and efficient measures to redirect traffic or modify the available capacity it will be possible to prevent traffic from facing critical congestion phenomena such as blocked junctions, gridlock etc., which, once initiated, it is improbable to be manageable.

Moreover, the quality of information affects the way road users will respond to changes in the network level of service. If it is too low, then the drivers will rely more on their past experiences and the overall traffic conditions will not change significantly. On the other hand, travelers will comply more with information concerning route recommendations and predicted travel time information if this information is reliable. Stated preference surveys show in fact that users rate highly accurate (predicted) information positively, and highly inaccurate (random) information negatively, and that the perception of trip times for the travelers varies with the extent of information (complete/partial), the type (prescriptive/descriptive, realized/predicted) as much as their experience [2]. Since the aim is to provide the road user with up-to-date, reliable and accurate information, the requirements to the provided information must correspond to the requirements that they value most.

In this paper we describe the motivations for establishing a data bank, which collects, processes and distributes data with standard levels of quality and reliability, and we give an overview of the conceptual and technical structure of the data warehouse that is currently being developed in the Netherlands.

The Dutch National Data Warehouse (NDW) project has been created with the aim to collect accurate, detailed and reliable data from a large part of the national road network and to act as data bank for urban municipalities, information service providers, research teams and, in turn, to the road user. At the time of writing, the project is in the preparatory phase and it is expected to have the first prototype system running in the first half of 2008 and to grow progressively towards an extensive coverage of the road traffic system on national, provincial and local levels within three years.

The purposes of the NDW are fourfold:

1. To resolve the discrepancy between all sources of information that can be collected, i.e. different municipalities, different collection systems, different refreshing rates etc.
2. To fill the gaps in the existing traffic monitoring system in terms of spatial coverage and information accuracy,
3. To provide reliable and fast access to data concerning the current state of traffic for efficient traffic management and information systems and
4. To enable predicting actual traffic phenomena.

The paper is structured as follows. In the next section we list the general reasons for developing a data warehouse. The requirements for obtaining accurate, widespread and reliable data are described later, together with some theoretical aspects that have to be considered in this process. The structure and functioning of the Dutch National Data Warehouse is then described together with a description of the technical issues that this system needs to currently resolve. Finally, we give the main conclusions and describe the future steps that will be undertaken.

WHY ESTABLISHING A DATA WAREHOUSE

The information about the traffic on the road that can be obtained daily from various available sources (e.g. radio, TV, Internet, on-road information systems) is often incomplete, imperfect, inconsistent and untimely. This is due to the co-existence of different types of networks (at different levels) all having their own information systems, and which are not integrated. Furthermore, the service providers have often access to scattered and sometimes outdated information over the state of the monitored road system. In particular, provincial and urban roads are often not covered by any traffic monitoring system, and lack at this coverage level often reflects also to the predictability of the monitored part.

These are just a few motivations why data banks have been created all over the world since already many years. Extensive data warehouses that feed traffic information services and that are used for traffic analysis can be found already in, e.g., US, Japan and the Netherlands.
The PeMS data collection system web-based tool designed at UC Berkeley to host, process, retrieve, and analyze road traffic condition information for a large part of the highways in California. Since being launched over six years ago, the Freeway Performance Measurement System (PeMS) has evolved into a powerful freeway performance measurement tool for the California Department of Transportation (Caltrans), its partner agencies, and the research community. PeMS receives data from California freeway traffic detectors, as well as incident-related data from the California Highway Patrol (CHP) and Caltrans. PeMS extracts various performance measures, such as vehicle-miles-traveled or average-daily-traffic, from real-time and historic freeway detector data [3].
In Japan the system VICS enables one to receive real-time road traffic information about congestion and regulation at any time of the day. This information is edited and processed by the Vehicle Information and Communication System Center, and shown on the navigation screen of travelers by text or graphical form [4]. This represents the oldest of the data warehouse systems worldwide, as its establishment was done over 10 years ago.

In the Netherlands data collection systems are currently done in a disaggregated way; many provincial authorities collect data for the main highways and for the most important urban arterial roads. One example is the Regiolab system, which collects and stores traffic data from a large part of the province Zuid-Holland, which contains the highways connecting three major cities: Rotterdam, Den Haag and Gouda [5].
The currently available data collection systems are however characterized by a few limitations. All systems at present collect data primarily from highways, whilst little information can be gathered from the urban roads. Within this aspect data banks are being made based on Floating Car Data, which has the advantage of covering a larger number of network links with respect to fixed sensors and with considerably less costs. Examples of these data banks can be found, e.g., in US by the Microsoft spin-off company INRIX [6] and in Belgium by the company Be-Mobile [7]. In the Netherlands a small part of the main urban arterials are covered by sensors but the data is currently available only for the provinces that are responsible for that part of the traffic network system.

We argue that improved utilization of the existing road infrastructure can be obtained only on the basis of a good understanding of the situation on the roads through an integrated, accurate and accessible data bank based on the following five cohesive measures:
1) Fill the blank areas, where traffic is currently unobserved or unpredicted;
2) Agree on the format and the quality of the data from all parties involved;
3) Organize the quality control of the structured information system;
4) Act as mediator between the many road authorities, institutions and information service providers by providing strict rules for collaboration and
5) Contribute to a clear division between traffic management and information.

The mission statement for successful creation and implementation of a national data warehouse is thus to "promote and establish the collaboration between municipalities, provinces, the government as well as private institutions and information service providers towards efficient and coordinated network traffic management and reliable traffic information" [8].
The establishment of an integrated database will allow all parties involved in the project to obtain a broader and consistent view on the traffic conditions on the networks they manage together with their surroundings. This agreement is expected to give many advantages, e.g.:
- Improve the traffic flow overall on the network, by restricting access to roads that operate close to their capacity and redirecting traffic to alternative routes;
- Facilitate the access to urban areas and city centers, by coordinating the demand with the available parking spaces and by an integrated monitoring of the incoming flow from the major accesses from the motorways;
- Allow more effective environmental management and control of the air quality, especially in bottleneck areas and "urban canyons";
- Contribute to improve traffic safety by monitoring all high-risk locations with high quality data levels.

Apart from short-term benefits, a data warehouse allows long-term data storage and retrieval. This adds further advantages to road authorities and institutions such as:
- Improve the scheduling of management and maintenance of the roads;
- Improve the integration and scheduling of public transports;

- Develop long-term traffic policies;
- Allow more effective planning and design of new management strategies, new infrastructures and improvement of the existing ones;
- Offer an invaluable dataset for assessment and research analyses.

By covering areas where currently there is no monitoring system, and by having a broader view on the current situation on the roads, including the ones external to the managed network, road authorities will be able to improve their mobility policies, to anticipate developments and make actions quickly.

Above all, the establishment of a national traffic data collection and monitoring system will, primarily, give benefits to the individual road user. Travel choices are in fact determined at present by incomplete information, which leads to uncertain predictions of the future traffic conditions that travelers will encounter when making their journey. Improving the level of detail and quality of this information and its timeliness will contribute to reduce the gap between actual utilization of the road system and optimal one. This effect will be achieved if travelers will be able to obtain reliable information before their departure as well as, once on the road, they will be updated quickly if road conditions change significantly, for example because of an accident.

REQUIREMENTS FOR ACCURATE AND RELIABLE INFORMATION GENERATION

The objective of a national data warehouse is to collect all available traffic data and to generate a database that contains all the relevant information content gathered from each fragment. These input elements are provided by road authorities and external providers both in terms of traffic conditions (travel times, flows, speeds, etc.) and of road conditions (lane closures, road works, status of DTM measures, etc.).

The functional requirements imposed on the gathered data depend on the information content that each fragment contains. Therefore, high quality is required for example to the data collected at the major arterials of a city, on the motorways, and at the main provincial roads. Quality levels are based on the likelihood of problems and their effects on the traffic in the network. Therefore, junctions and traffic bottlenecks are monitored at the highest level of accuracy, while roads characterized by generally low inflows and seldom changes in terms of traffic conditions require a lower level of detail.

The provided dataset should guarantee standard levels of quality, which can be quantified according to the following characteristics:

- It should have satisfactory levels of availability, accuracy, timeliness and reliability according to the application they are intended to be used for;
- It should quantify the quality level of the input data gathered from the various parties and bring them to the desired level;
- It should remove the inconsistencies in the datasets and in between datasets and provide unambiguous information;

- It should bring all parties to a standard level so that they can communicate using the same definitions, measures etc.

The trade-off between information quality and costs for its distribution suggest that standard pre-requisites are determined on the basis of the following considerations:
- Minimum quality levels are required for each specific management strategy applied and for the network context; the system should enable one to obtain only the relevant information needed without or with small redundancy;
- In case of application of network-wide traffic management they should add important information about management decisions such as when and where a certain strategy is adopted or how frequent the information provided is refreshed;

It is therefore necessary to have a sufficient knowledge of the specific quality requirements that each strategy needs for achieving effective management of a network both in terms of spatial coverage and of timeliness.

From application to minimal quality levels

It is important to clarify that the information in the data warehouse should not yet be valid and ready for traffic information. Road authorities and the service providers should have the task to collect the data and to derive estimates of the future traffic conditions from the processed data, to decide in which format the information is given to the road users and how and if to give advices to eventual alternatives of travel. In sum, the data warehouse provides data that *can* generate good traffic information data.
Once the data are collected, processed and connected one to another, the same authorities and external providers can use this consolidated dataset for two main purposes:
- traffic management
- traffic information

This implies that the information service provider may also ask for a specific level of quality for the dataset depending on the level of accuracy established for the information distributed to the road users. The data warehouse provides accurate and reliable *realized* travel times or *instantaneous* measures like local speeds and flows, while it is not intended to provide any estimate of the *actual* travel times, i.e. what the traveler has not yet experienced. It is task of the service providers to choose appropriate methods and models to obtain an estimate of such travel times.

It is therefore important, in the estimation of future traffic conditions, to have a sufficient knowledge of the current states. Within this respect, frequent updates and large numbers of data collection points certainly contribute positively. However, increasing the pace of data sending or installing data collection systems imply costs that are often bound to a limited budget.
Two factors typically influence the quality of data and accuracy of traffic flow estimation and prediction: 1) the updating time at which new information is sent to the information service provider and elaborated for distribution to the road users and 2) the spatial distance between data collection points. The impact of two elements is discussed with some examples in the following of this section.

Updating time

It is important to define the updating time needed from raw data to become information provided to the road user. This time can be subdivided into four elements:

- **Aggregation period**, i.e. the interval of time in which information data is aggregated and/or averaged;
- **Delivery time**, i.e. the time offset between the end of the measurement period and the delivery of the aggregated data to the authority or service provider;
- **Data processing time**, i.e. the time needed to transform the data provided by the NDW into useful information to the driver;
- **Operational time**, time needed to feed the information systems or the management systems with the new updates and the new management actions.

Figure 1 shows the relationship between raw data and processed information. Data is extracted from the traffic system from all available sources. Once collected for a limited period of time it is processed, i.e. it is cross-checked, cleaned of errors and outliers, and interpolated to the whole basic network. The processed dataset is finally loaded into the data warehouse and made available to road authorities and service providers, which use it for intervening on the infrastructure or elaborate it through prediction models and historic data. Only at this stage service providers can update the information given to the road users. This process is repeated continuously, since the drivers will in turn react to the new information by updating their travel choices (e.g. route, departure time, mode of transport). The time needed to refresh the information given to the road users is therefore a fundamental factor in the evolution of the traffic system.

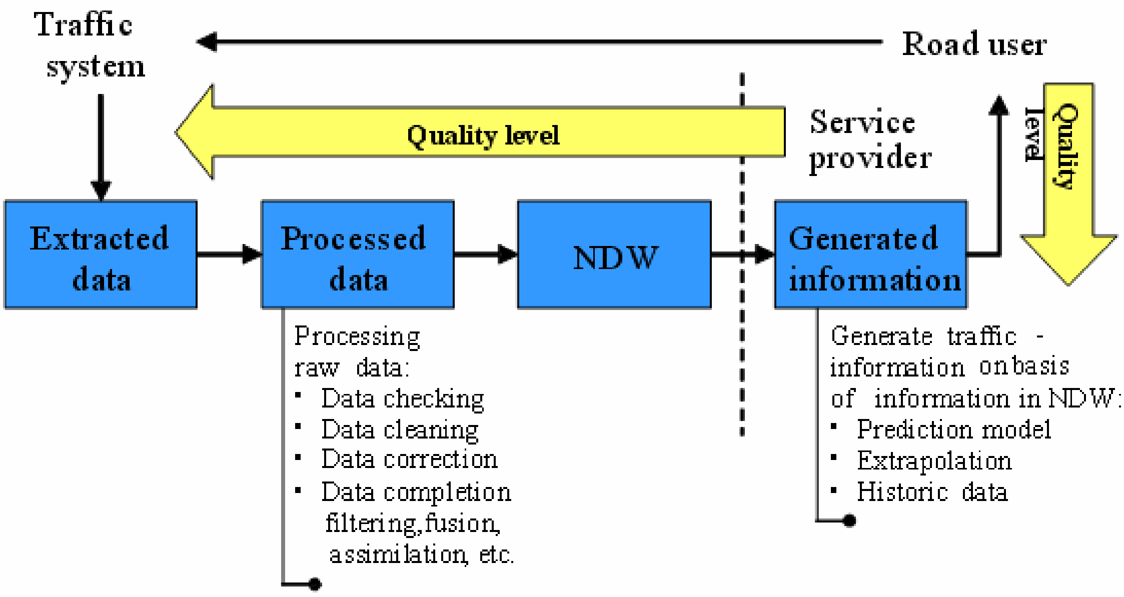

Figure 1: from data collection to information

Fast and reliable data is especially expected to give benefits in cases of non-recurring congestion due to e.g. special events, incidents, road work areas. In these situations the available management strategies can be effective only if there is enough information on the actual conditions of the roads. This information depends on how frequently updates are given of the current traffic conditions and how distant each data collection point is located one to another.

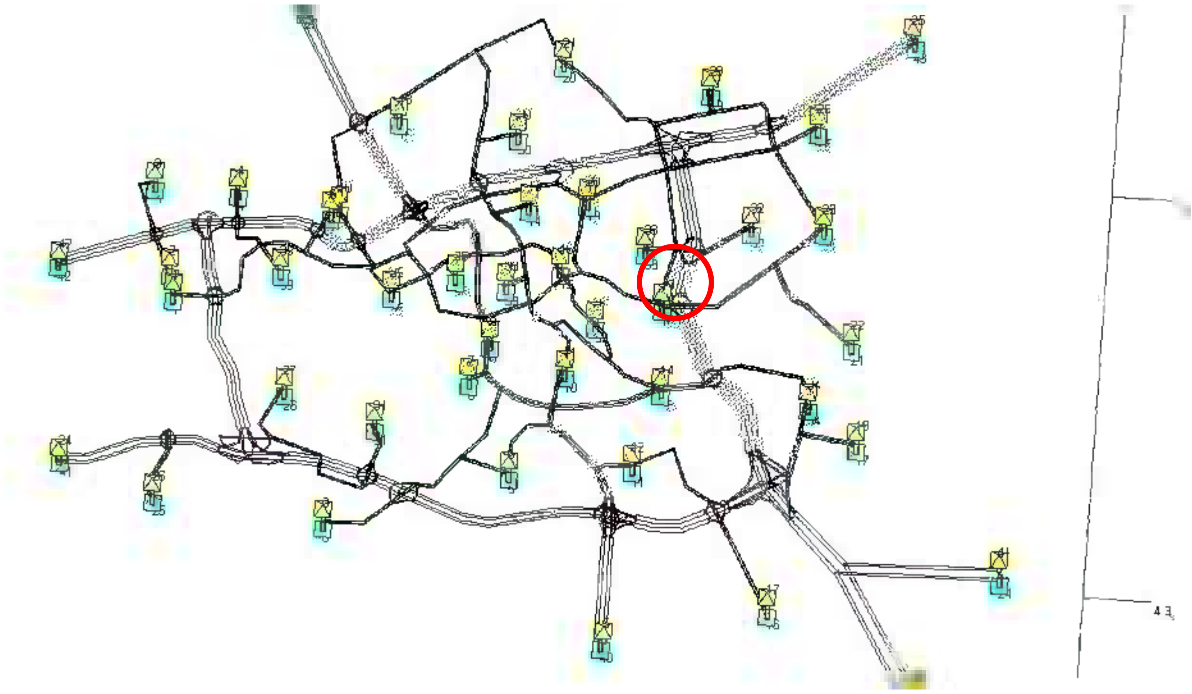

Figure 2: Representation of part of the road network of Rotterdam

To give an idea of how delayed information on the actual traffic conditions can influence the effectiveness of a management strategy we can consider the following example, where an incident has been simulated on the ring-road of Rotterdam (figure 2). All the travelers in the network are assumed to obtain information concerning this accident and to react to the route advice which provides realized travel times, but the updates are delayed.
The network contains 468 links, 239 nodes, 44 destinations, and it has been subdivided into 3920 cells and modeled with a time step of 10 seconds. The accident has been simulated on the motorway, which resulted in a reduction of the capacity from 1400 to 1000 veh/h. The user choice has been modeled via a simple Logit model.

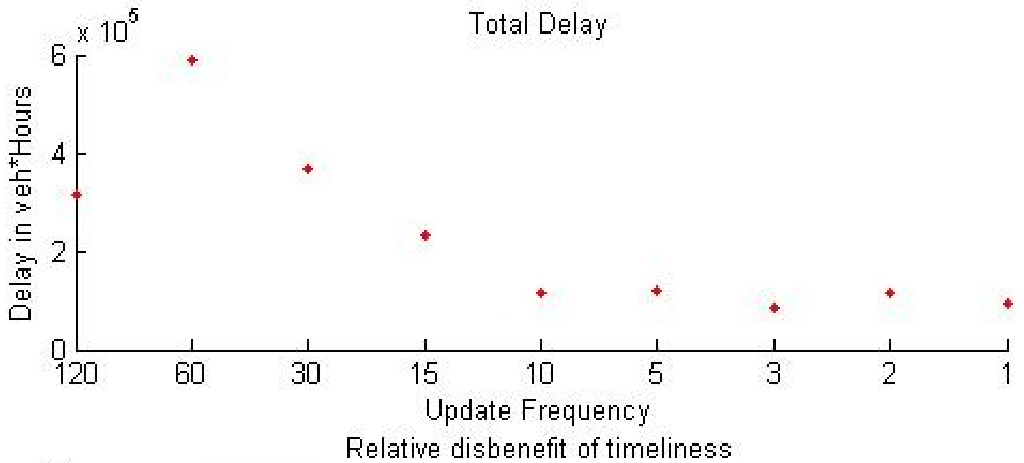

Figure 3: Relationship between effect of an incident and information delay

The network performance has been measured by means of the generalized network throughput, which is the difference between the cumulative summed departures and the cumulative summed arrivals at the end of the simulation. The smaller this value, the faster people have traveled in the network. Moreover, total delay and total travel time where computed. All values show similar trends. Figure 3 shows the performance in terms of total delay. Up to 10 minutes delay, the effects of delayed information (on this specific case) remain relatively small.

The effects of an incident vary depending on the functioning and topology of a network and 10 minutes delay in the information updates can be in some other cases a too long period. In cities for example this period can be sufficient to allow queues to grow and involve also other parts of the network. Phenomena like spillback and gridlock can occur, and if this happens then there will be not much left to manage.
To show these effects we simulated an accident at the urban level on a part of the network of Amsterdam (figure 4) using the simulation program Vissim [9]. The accident is expected to create a queue upstream which, if traffic is not re-routed efficiently, will lead to spillback effects on the roundabout located at the beginning of the link and, eventually, to full gridlock of the roundabouts and all links connected to it. The accident is assumed to block the road completely for 15 minutes. When the incident is detected, a part of the flow is re-routed by means of alternative routes according to the available information. The delay between information update and information on the road is in this example assumed 1 minute, 5 minutes and 10 minutes.

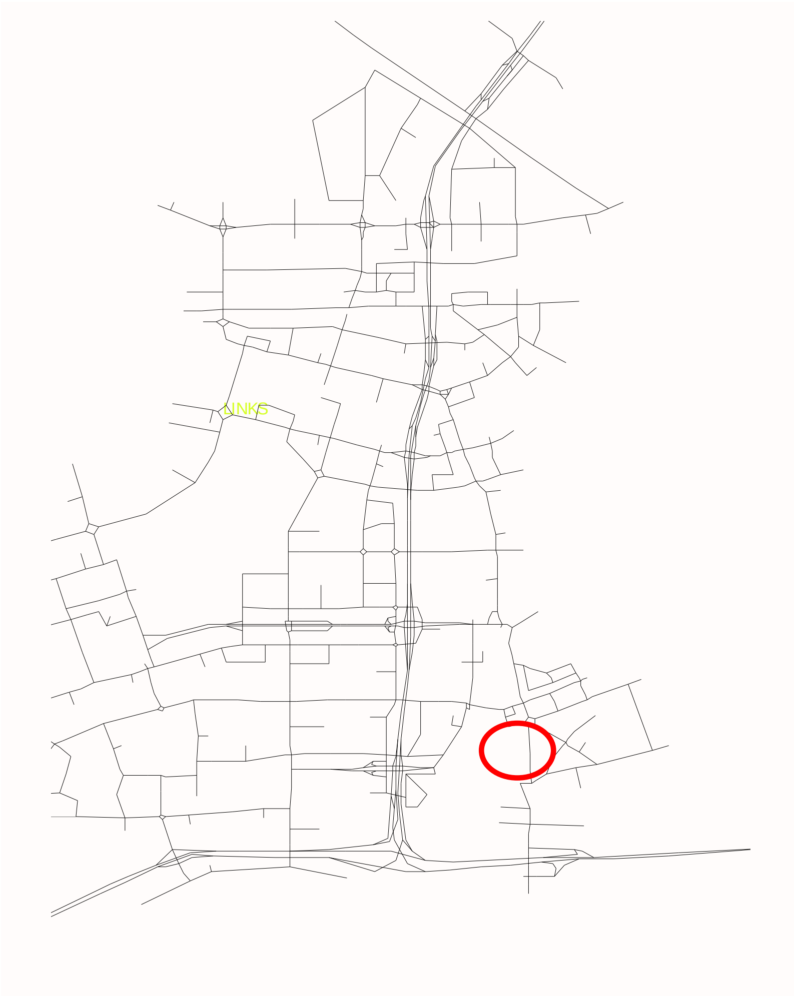

Figure 4: Representation of part of the road network of Amsterdam

Table 1: Performance indicators for the network of Amsterdam

	60 s	300 s	600 s
1. Number of vehicles in the network, All Vehicle Types	8034	8960	9291
2. Number of vehicles that have left the network, All Vehicle Types	15350	14295	13908
3. Total Path Distance [km], All Vehicle Types	127478.5	124818.7	123304.1
4. Total travel time [h], All Vehicle Types	3788.309	4235.224	4387.288
5. Average speed [km/h], All Vehicle Types	33.65	29.472	28.105
6. Total delay time [h], All Vehicle Types	2288.778	2707.031	2851.546
7. Average delay time per vehicle [s], All Vehicle Types	352.361	419.063	442.5
8. Total stopped delay [h], All Vehicle Types	533.866	795.888	878.151
9. Average stopped delay per vehicle [s], All Vehicle Types	82.19	123.208	136.271
10. Number of Stops, All Vehicle Types	576951	738235	798069
11. Average number of stops per vehicles, All Vehicle Types	24.673	31.745	34.401

Tab. 1 shows some indicators which describe the resulting traffic conditions in network. We see clearly that with an information update interval of 60 s, the network performance generally improves. It is very important to observe that the table concerns all vehicles in network, therefore not only those vehicles that pass where the incident takes place.

The total path distance is in scenario 1 larger than in the other scenarios. This is in line with the expectations, since more traffic is diverted. The total travel time decreases, showing that the information system led to a more efficient choice for the drivers. The same conclusion applies for all other performance measures.

Accuracy

Second important aspect for guaranteeing reliable information on the current traffic performances is represented by the accuracy of such information. Data gathered from the roads contains a certain percentage of errors due to e.g. missing data, interpolation and averaging etc.

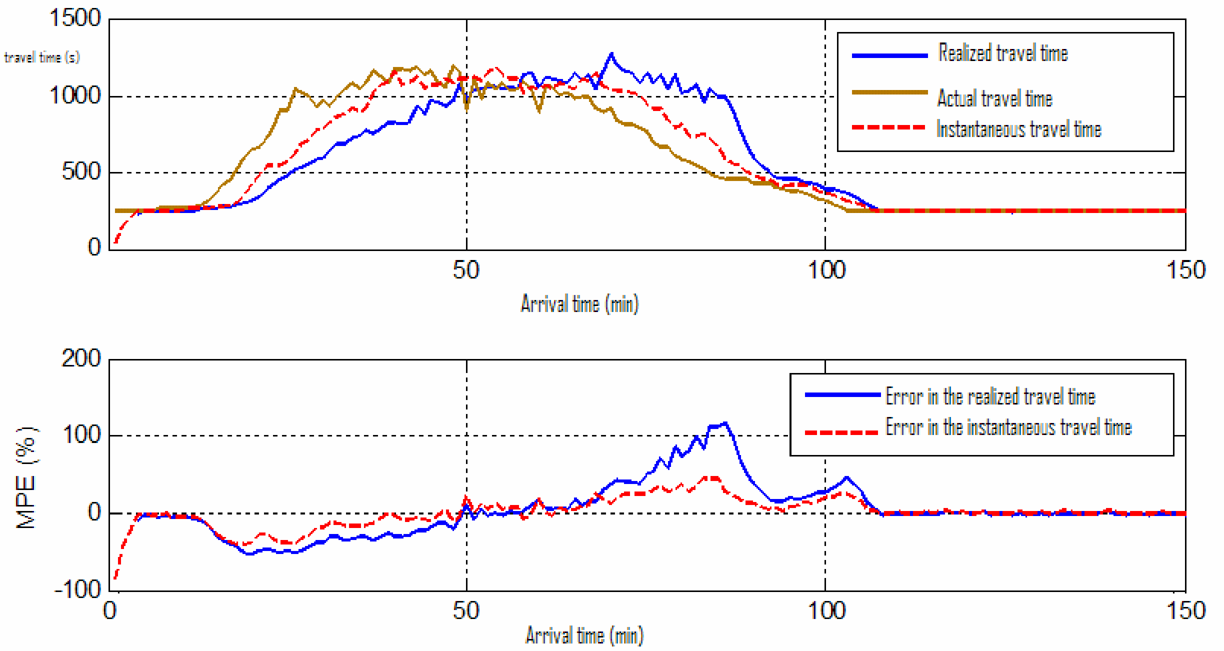

Figure 5: Error contained in the information about realized and the instantaneous travel times

Table 2: Error in the information based on realized and instantaneous travel times

	Error percentile		
Method	Minimum	Maximum	MAPE*
Realized travel time	-51,9%	117,3%	23,6%
Instantaneous travel time (500 m)	-36,9%	41,5%	10,5%
Instantaneous travel time (1500 m)	-41,4%	46,1%	12,6%
Instantaneous travel time (3000 m)	-45,3%	73,5%	15,7%

The spatial location of data points strongly affects this percentage as much as the updating rate at which data is aggregated. Especially at variable traffic conditions the quality of measures is

* Calculated over the whole time period

determinant. Figure 5 shows how the dynamics of traffic create a time delay between what travel time drivers have just finished to experience (realized travel time), what they are expected to experience while they are driving (instantaneous travel time) and what they actually experienced at the end of their trip (actual travel time) while table 2 shows how the error percentile grows when the instantaneous travel time is given at different spatial distances.

We show the impact of different quality of data in the application to traffic estimation and prediction through another example. The application for estimation and prediction has been based on a traffic model which runs synchronously with the data. In this section the functioning principle of this model is briefly outlined. For a detailed analysis we refer to Tampère [10]. The study area consists of simply a homogeneous motorway corridor where an accident takes place at km 14. The traffic model describes the traffic flow process according to first order traffic flow theory. This means that the advancement and back-propagation of the traffic-jam area are modeled deterministically, as well as the traffic conditions in and outside the traffic-jam area. The traffic conditions are modeled on a macroscopic level, i.e. in terms of flow, intensity and average speed. Data collection points are assumed in the example to be spatially located at fixed relative distances, from 800m to 3200m. In these points traffic counts and mean speeds are collected at fixed time intervals.

Given an initial estimate of the traffic situation (e.g. an initial measurement), the macroscopic traffic model predicts how that situation evolves further. As soon as (e.g. after a minute) current measurements of the traffic situation are refreshed, possible deviations between the modeled and traffic situation are adjusted. This process is called data assimilation. Both the variables which describe the traffic situation and some important parameters of the traffic model (e.g., capacity) are adapted to the measured values. Data assimilation is operated using (extended) Kalman filtering techniques in this example. This process repeats at every refreshing time of the dataset.

The traffic model thus provides for each point in time and space an estimate of the traffic conditions, which can be used to calculate travel times. The model can also provide short-time forecasts (e.g. over 15 or 30 minutes) by making additional assumptions concerning future changes of the traffic conditions and of the management rules.
This estimate is obviously sensitive to the configuration of the measurements. In space corrections can be made only on those locations for which measurements are available, while prediction on other locations relies on the interpolation model adopted.

The analyzed scenario is an incident on a homogeneous highway of 15 kilometers long. The scenario was simulated in the microcomputer simulation model Aimsun2 over a period of 4 hours. After 45 minutes an incident blocks 1 of the 3 traffic lanes for 105 minutes, whereupon the accident area is freed. The traffic flow evolves during the first 30 minutes from 0 to 5500 veh/h, it remains stationary for 2 hours and then it decreases for 1 hour to 3000 veh/h.
The scenario has been evaluated with the traffic model using different spatial distance for the detector locations and with 1 minute delay between data acquisition and data assimilation. We compare the result of the model with the time delay with the ones run with no time delay, if for example the state of the network is immediately communicated (e.g., by means of eCall).
The model estimate has been compared with the results of the microscopic simulation program (Aimsun2 [11]).

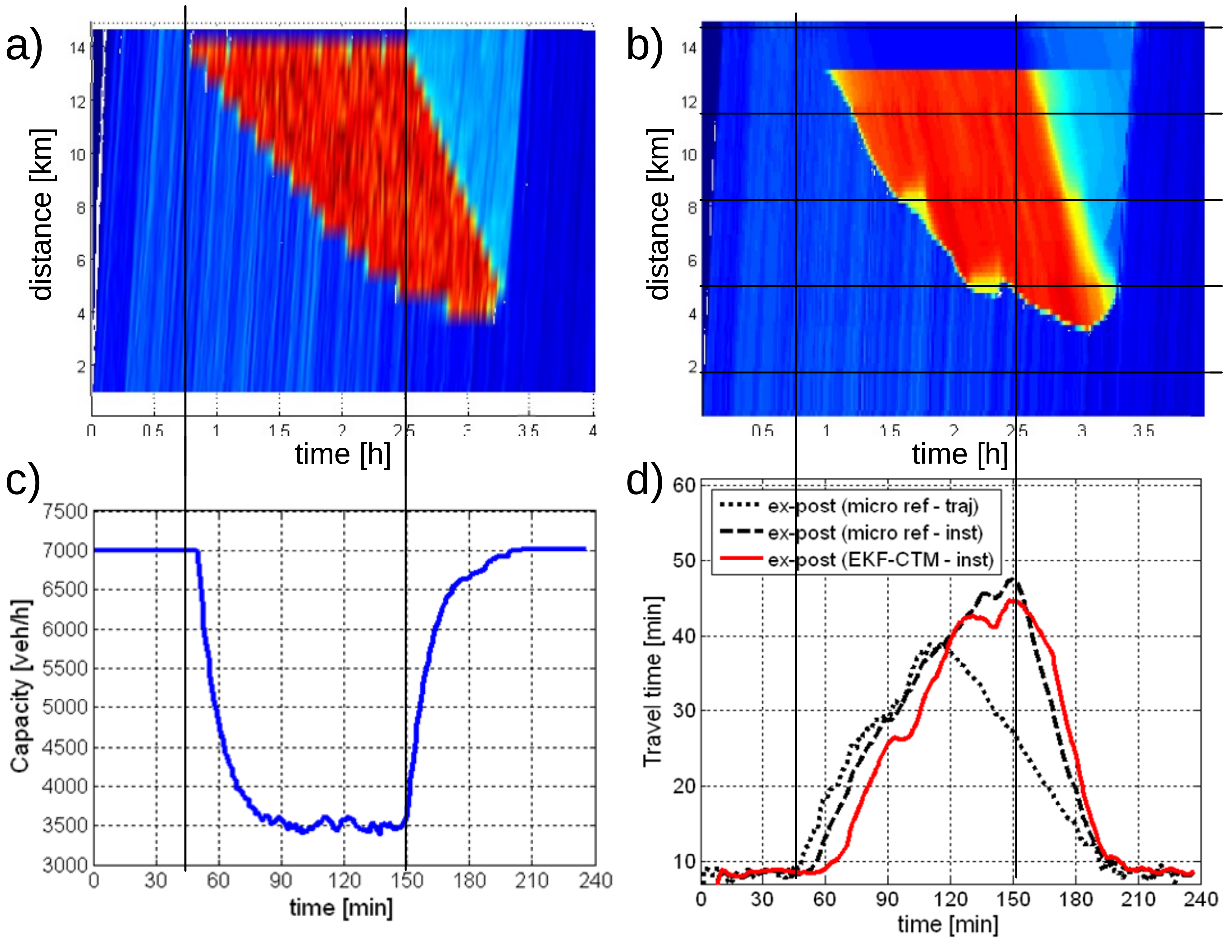

Figure 6: Evolution of the traffic density due to a road accident (right) and model prediction (left)

Fig. 6a and b represent space-time diagrams of traffic densities. The advancement of the front of the traffic-jam in the upstream direction of the flow is clearly distinguished. Also the capacity reduction is clearly shown (fig. 6c). Comparing the two top figures there is approximately 15 minutes delay before the capacity reduction is actually discovered by the estimated model. The corresponding travel times in the simulation and the model estimates are drawn in fig. 6d. The real experienced travel time is the trajectory travel time (dotted curve). However, these can be calculated just afterwards (after the ride). The best immediate estimate of the travel time is the instantaneous travel time. Note that this estimate deviates strongly in this scenario, because it cannot take into account the future variation of the inflow traffic. It should be clarified that the prediction errors in figure 6d are due for a part to the delay in the data correction phase but also to the approximation errors due to the adoption of a simple first-order model.

Figure 7 compares the performance of the prediction model with respectively 800m distance between detectors with 3200m distance. The prediction with small intervals performs almost as good as the model computed with instantaneous information (eCall). Also the location of the incident - and therefore traffic-jam length and travel time - has been better distinguished in the first case. The conclusion is that a shorter detector distance allows more accurate estimates and the predictions of a traffic model are more sensitive to sudden travel time changes.

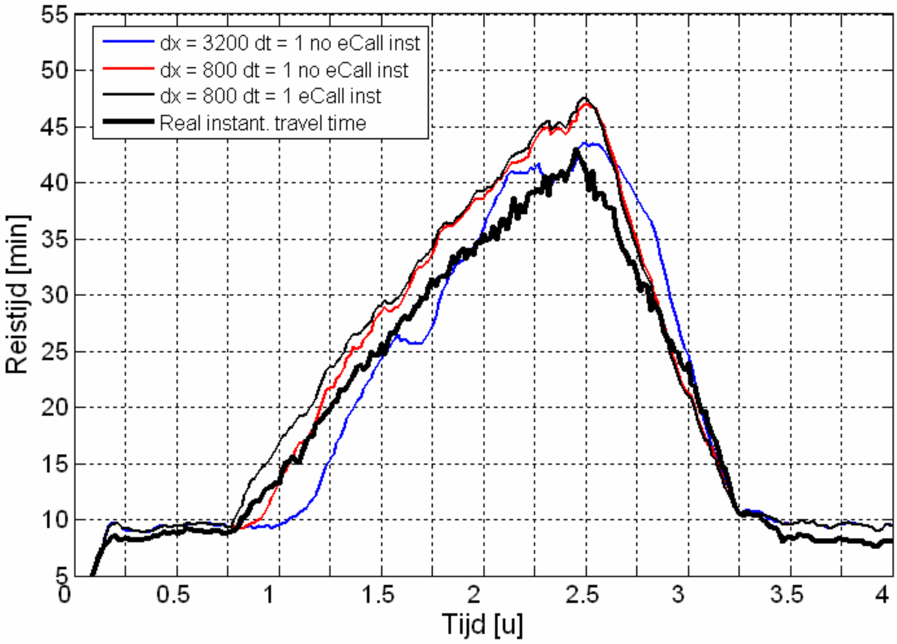

Figure 7: Instantaneous travel time vs. predicted ones

The conclusion we have drawn strongly depends on the network chosen as example and the traffic model adopted. Other traffic models may show lower sensitivity to aggregation of data and interpolation, as well as other networks may behave more rigidly to changes in the inflow. It is therefore good policy to determine the minimum aggregation time and distance between data collection points depending on the network context they are being installed.

NATIONAL DATA WAREHOUSE: THE DUTCH CASE

The observations made in this paper have motivated the implementation of a National Data Warehouse in the Netherlands. The NDW was started up as an initiative of the Ministry of Transport, Public Works and Water Management, but is now a broadly supported project by a group of leading authorities and service providers. These authorities recognise the importance of a good information system and want to create the NDW as a joint effort.
The group of ‘leaders’ in the NDW consists of the provinces of Noord-Holland, Zuid-Holland, Noord-Brabant and Utrecht, the municipalities of Amsterdam, Utrecht and Rotterdam, the city regions of Arnhem-Nijmegen, Eindhoven, Rotterdam and Amsterdam, and the Ministry of Transport, Public Works and Water Management. Discussions with the Northern Provinces (SNN), aimed at involving them among the leaders, are currently in progress. This means that in the future the group of leaders will cover the most traffic problem-sensitive areas in the Netherlands.

A large number of these leaders have signed a joint declaration of intent, in which they have confirmed their desire to implement the NDW with each other. The leaders will enter into a collaboration agreement in which they will document the NDW structure and how it will be financed and managed.

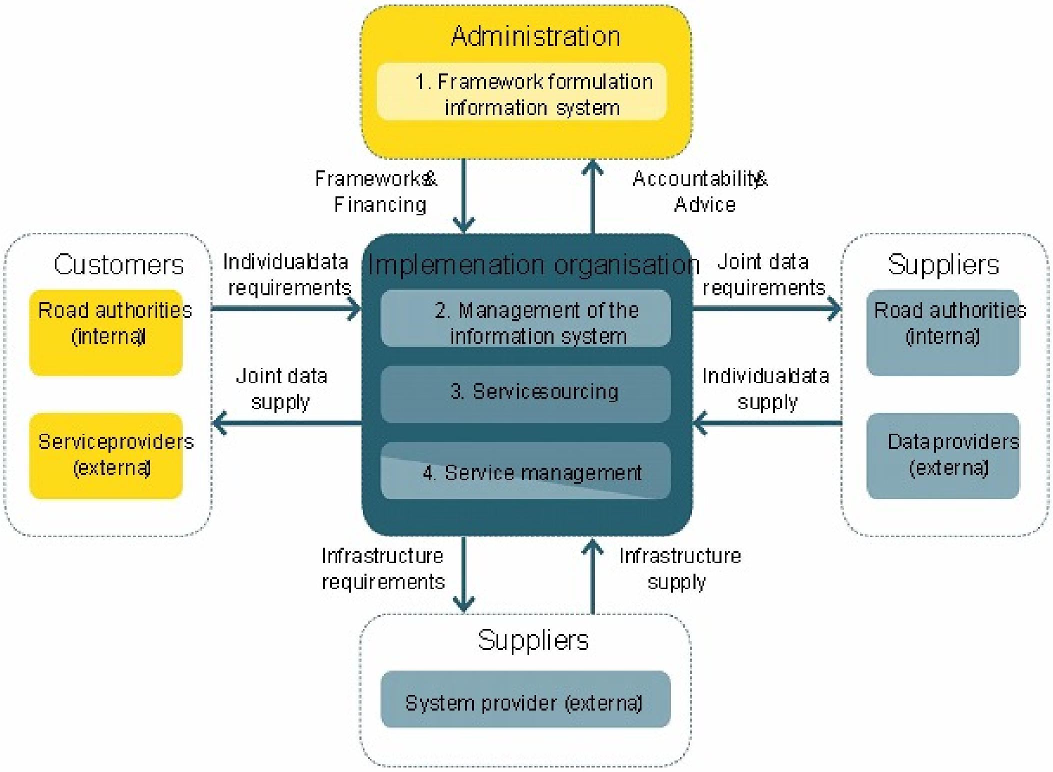

Figure 8: Organization chart of the NDW project

Figure 8 shows the organization diagram of the NDW project and the role that each party has been assigned. The NDW is still in the preparatory phase. The National Mobility Council (NMB) is involved in the decision-making process and approves all important steps in the implementation process. On 29 March 2007, the NMB once again underlined the benefits and necessity of the National Data Warehouse and asked for it to be implemented as soon as possible. Once the leaders come to an agreement about the cost distribution, the collaboration agreement can be concluded. This will end the preparatory phase of the NDW and a start will be made on the tender for data collection and the setup of the implementation organization. The collaboration agreement is expected to be signed in September 2007. The subsequent tender and the establishment of the organization will take approximately six months. In the first half of 2008, the first data from the functional NDW will become available to the users.

The NDW will contain minute-by-minute data. This information will be provided to service providers (e.g. traffic authorities, traffic information providers, etc.). The various parties involved in the project have agreed to provide travelers with no different data than what is provided to DRIPs (Dynamic Route Information Panels). The National Data Warehouse will compare the raw data with the transferring costs (which are relatively low). It is then responsibility of service providers to process and enhance the data, and to make it available to the road users (via, e.g., internet, radio, television, sms, etc.). Private information providers (such as TomTom and the like navigation systems) can acquire this information and use it to improve their own service. Even if relatively very timely with respect to traditional data warehouses, this information will still need to be improved to get actual or expected future traffic conditions. Providers can request 1-minute data from a certain area and use it for their prediction models. The NDW acts also as historical database and historical data of a certain area can also be used by service providers to feed their models. There will be also some tools to visualize the data, for example for policy making. These tools will be particularly useful for municipalities and provinces, which often do not have such types of tools.

Data in NDW will be gathered in two ways:
data concerning the underlying road network will be given to service providers, municipalities, etc. through aggregated measures like traffic average speeds and intensities (via, e.g., double loop detectors) and
(realized) travel times (via, e.g., camera detection systems) will be obtained by external traffic data providers. A formal competition for these providers to participate to the NDW project will be made after Christmas 2007.
The underlying road network data will be already available from the beginning of 2008, while external service providers will be selected based on the formal competition in mid-2008 and during the second half of the same year they are requested to install their system in the network. By the start of 2009 these systems will have to provide the NDW with their data. Thus in principle the NDW will be fully operational in the beginning of 2009.
For the first phase of the NDW (first 4-6 years) only conventional techniques will be used (e.g., double loop detectors, camera detection systems, etc.). This is because upcoming techniques are yet not able to meet the needed quality requirements. In some cities they already use static traffic models to enhance the data gathered from conventional collection systems.
NDW will provide, apart from intensities and speeds, 7 other types of information:
1. information about road works;
2. state information –actual and planned- concerning bridges, peak-lanes ;
3. actual road blockages;
4. events (festivals, football matches, etc.);
5. control strategies and control scenarios currently applied;
6. occurrence and clearing time of incidents;
7. actual information about queues, traffic management messages (re-routing, closings, etc.).

The NDW uses a growth model. This means that at the beginning not all roads in the Netherlands will be included in the database. The aim is to cover the entire relevant road network in the Netherlands within a few years (target year 2009). Figure 9 shows which principal roads are currently monitored in the country together with a (preliminary) overview of the future NDW basic network (with the current forerunners). The NDW project is expected to cover a basic network of around 6600 kilometers of motorway, provincial and urban roads within 3 years.

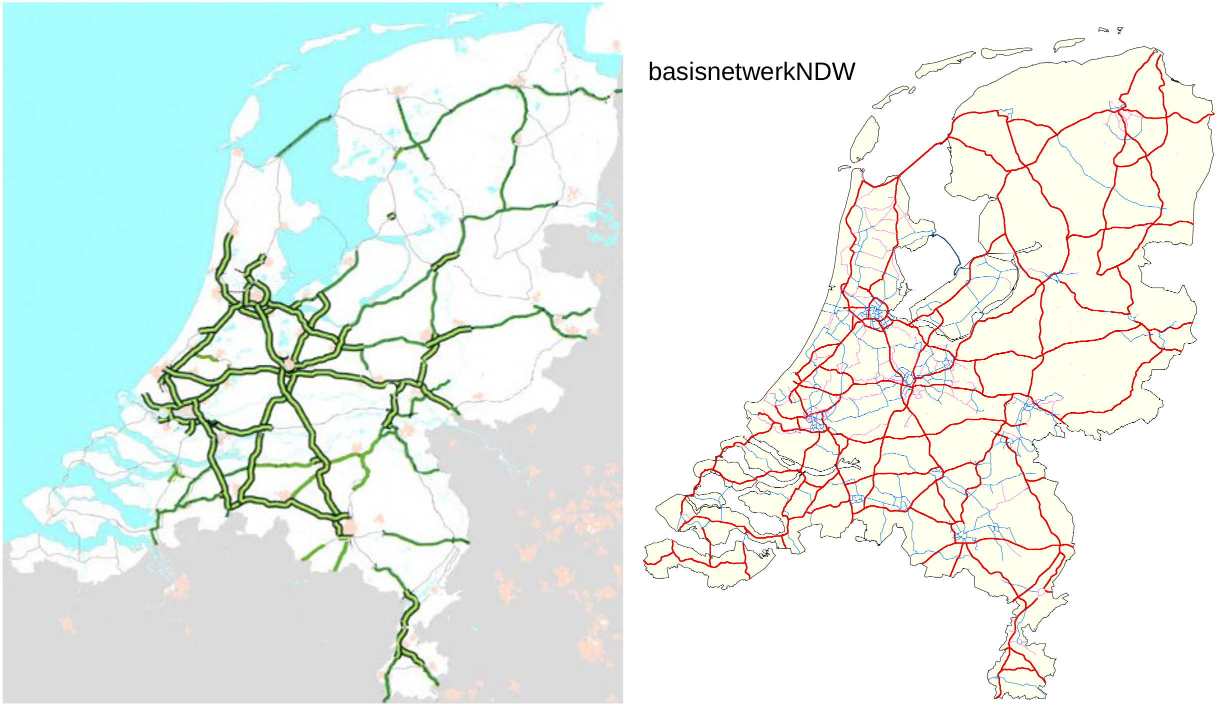

Figure 9: Current network coverage (left) and basic network targeted by the NDW (right)

The improved utilization of the existing infrastructure through traffic management and traffic information will result in a 5 to 10% decrease in traffic jams [12]. This is expected to result in considerable social benefits.

CONCLUSIONS

Many road users are damaged daily by an inaccurate information system, if any exists. Many areas in fact, in particular provincial and urban roads, are often not covered by any traffic monitoring system, and lack at this coverage level often reflects also to the predictability of the monitored part. Lack of complete, reliable, and timely information affects the travelers' choices which are bound to be often non-optimal and the use of the overall system is far from being efficient.

Furthermore, road authorities can currently benefit from the available traffic management measures up to certain limits. Apart from limitations due to the incompleteness of the overall data monitoring system, the tasks of problem prognosis and prevention strongly depend on how quickly and precisely it is possible to gather all the relevant data to have an estimate of the current and the future state of the roads. Only with fast and efficient measures to redirect traffic or modify the available capacity it will be possible to prevent traffic from facing critical congestion phenomena such as blocked junctions, gridlock etc., which, once initiated, it is improbable to be manageable.

This paper has investigated the minimum requirements that traffic data should satisfy in order to achieve accurate and reliable information. Theoretical as well as numerical examples have been given to show how different spatial distances and different updating times can reflect into different quality of the information given to the travelers and, as consequence, to the performance of the network itself.

The need for accurate and timely information for ITS systems has motivated the establishment of the Dutch National Data Warehouse, which is a project developed with the scope of collecting information from many sources and bring them to a minimum level of quality on a basic network level. The goal is to cover large part of the highway network as well as the underlining urban network, resulting in about 6600Km of roads. This unique data bank will provide road authorities with accurate, reliable and timely information and will allow them to provide the road users with high levels of information and to manage the road network more efficiently.

ACKNOWLEDGEMENT

The authors would like to thank Andre Solinger and Frank Zuurbier for contributing to parts of this paper.

REFERENCE

[1] Bovy, P. H. L., 2001. Traffic flooding the Low Countries: how the Dutch cope with motorway congestion. *Transport Reviews* **21**(1): 89-116.

[2] Srinivasan. K. K., Mahmassani, H. S. , 2002. Trip Time Perception and Judgment Processes in Tripmaker Decisions under Real-time Traffic Information. Presented at Transportation Research Board, 81st Annual Meeting. Washington, DC, CD-ROM

[3] http://www.calccit.org/projects/pems.html

[4] http://www.vics.or.jp/english/vics/structure.html

[5] http://www.regiolab-delft.nl/

[6] http://www.inrix.com/accuracy.asp

[7] http://www.be-mobile.be/

[8] www.nationaaldatawarehouse.nl

[9] PTV, 2003. Vissim 3.70 User's Manual.

[10] Tampère C.M.J., Immers L.H. (2007). An extended Kalman filter application for traffic state estimation using CTM with implicit mode switching and dynamic parameters. Proceedings of the 10th IEEE-ITSC Conference on Intelligent Transportation Systems, September 2007, Seattle.

[11] Barceló, J., Ferrer, J.L., Montero, L., Perarnau, J., 2004. Microscopic Simulation With Aimsun For The Assessment Of Incident Management Strategies. *European Transport Conference, ETC2004,* Strasbourg, France.

[12] McKinsey and the Transport Research Centre of the Ministry of Transport, Public Works and Water Management, 2007. Technical Report.\