Change History

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Status</th>
<th>Author (Partner)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>14.04.2008</td>
<td>Draft</td>
<td>Rene Kelpin (DLR)</td>
<td>1st draft</td>
</tr>
<tr>
<td>1.0</td>
<td>09.05.2008</td>
<td>Draft</td>
<td>Rene Kelpin (DLR)</td>
<td>2nd draft</td>
</tr>
<tr>
<td>1.1</td>
<td>19.05.2008</td>
<td>Draft</td>
<td>Kelpin/Parnitzke (DLR)</td>
<td>Final draft</td>
</tr>
<tr>
<td>2.0</td>
<td>30.05.2008</td>
<td>Final</td>
<td>Kelpin (DLR)</td>
<td>Final version</td>
</tr>
<tr>
<td>2.0</td>
<td>31.05.2008</td>
<td>Final</td>
<td>Saarikivi (FORC)</td>
<td>Editorial check</td>
</tr>
<tr>
<td>2.1</td>
<td>26.05.2009</td>
<td>Final</td>
<td>Saarikivi (FORC)</td>
<td>Front page with flag</td>
</tr>
</tbody>
</table>

Distribution list

European Commission
Emilio Davila Gonzalez
Wolfgang Höfs

ROADIDEA partners
E-mail list

www.roadidea.eu
Executive summary

The main objective of the work package 2 “data collection” (WP2) is the investigation and the utilization of available and valuable data sources for the project ROADIDEA. The data collection and its results are to be regarded as a main prerequisite for further development within the project ROADIDEA. The quality of the results as well as the success of the project depends strongly on the availability of diverse and high quality data all over Europe. These data sources should be accessible in an automated way, free of charge at least for research purposes and piloting, well described, from high quality, and reliable. Data from these sources are to cover the major part of Europe to enable new innovative services along long haul routes through Europe. However, the data also is to improve local weather and traffic services by transferring research results and synergies from one European country to another.

Some innovative pilot services available local and European wide are to be developed within the project ROADIDEA.

To reach this objective first and foremost it was necessary to get an overview of data sources, which are easily accessible all over Europe. But, to avoid a limitation of the innovation level possibly caused by the availability of only a few different kinds of data, a more comprehensive data source identification was needed. This data source identification took place in the data collection period which was started with the project kick-off in December 2007. All partners have been involved in this work package. They were asked to describe the data of their every day operations as well as data from new or so far un-tapped data sources. Beside that some partners have investigated the general data availability in their countries to get a more global view of the data coverage in Europe. This general data availability investigation was to deal with weather and road traffic data availability at least.

A great amount of different data sources has been investigated and described. The data availability investigations in the participating countries have uncovered a very heterogeneous European data availability with a lot of problems which have to be tackled within the project ROADIDEA partly – namely in the following tasks of WP2.

To get a general understanding of the data and the surroundings they are from, and to discuss problems, technical details and next steps a data workshop was held in Berlin. All partners attended this workshop. They presented the data, which were described during the data collection period. Based on collected data sets and two more comprehensive data source analyses in Germany and Finland a ROADIDEA data classification was introduced, discussed and applied to the ROADIDEA data collection work.

This report describes the work that has been done and the results which have been achieved in the first task 2.1 “Data Source Identification – Present Data” of the WP2.
Table of contents

1. Introduction .................................................................................................................. 5
2. Methodology .................................................................................................................. 7
   2.1 Data Collection Survey ............................................................................................ 7
   2.2 Data Classification .................................................................................................... 10
      2.2.1 Motivation ......................................................................................................... 10
      2.2.2 German Data Source Analysis ........................................................................ 12
      2.2.3 Finnish Data Source Analysis .......................................................................... 14
      2.2.4 Data Classification Proposal – Clearing House Approach ............................... 16
   2.3 The ROADIDEA data workshop .............................................................................. 21
      2.3.1 Objectives ......................................................................................................... 21
      2.3.2 Methodology ..................................................................................................... 22
      2.3.3 Results ............................................................................................................... 23
3. Results ............................................................................................................................ 24
   3.1 Data Collection ......................................................................................................... 24
   3.2 Data Availability ....................................................................................................... 30
      3.2.1 Finnish data availability .................................................................................. 31
      3.2.2 German data availability ................................................................................. 35
      3.2.3 Swedish data availability ................................................................................. 39
      3.2.4 Dutch data availability .................................................................................... 41
      3.2.5 Italian data availability .................................................................................... 43
      3.2.6 Hungarian data availability .............................................................................. 45
      3.2.7 Croatian data availability ................................................................................ 45
      3.2.8 Slovenian data availability ............................................................................... 50
      3.2.9 Data logging from CAN-bus information ........................................................... 51
   3.3 Data Quality ............................................................................................................... 51
   3.4 Data Classification ..................................................................................................... 52
4. Conclusion/Summary ..................................................................................................... 54
References ......................................................................................................................... 56
1. Introduction

The main objective of the ROADIDEA project is to create a road map for radical innovations in a developing European transport service sector. That means accessibility to all kinds of useful background information combined with advanced data fusion methods for creation of innovative mobility services. Background information as well as mobility services have to be fed with high quality data. The quality of the services depend to a vast extend on the quality of the data.

Within the innovation processes in the ROADIDEA project it is planned to specify new transport services with respect to the analysed user and European transport system requirements. These new and innovative services are to be implemented in the ROADIDEA road map to guide transport chains through Europe. The ROADIDEA Work Package 2 “Data Collection” (WP2) supports the innovation processes and paves the way to a high quality and reliable data background.

What data might be relevant for services to be defined in a later process? That is a question nobody is able to answer. Nevertheless, the availability of a huge amount of different data sources is to be treated as a fertile field on which a broad diversity of ideas might grow. These ideas have to be fed constantly. That is the motivation to undertake large efforts in the ROADIDEA project for data source identification, data collection, data description and data classification.

WP2 includes five tasks. This report gives a detailed description of the work done in the Task2.1 “Data Source Identification – Present Data” until now.

![Figure 1: WP2 task overview](image)

After defining the pilot services within the innovation processes (WP5) Task2.2 “New Types of Data” will identify new data needs and tries to tap new data sources, which may able to feed the created services. A thorough analysis of the available data may give
the opportunity to substitute missing data or to derive further information (aggregated data) by combining different data sources in Task2.3 “Integration of Data”. Task2.4 “Definition of Uniform Data Archives” is to enable a database to host all data sets. This database is to be called by the mediation service, which is designed in Task2.5 “data mediation”. The data mediation service uses an access layer and a mediation layer to connect data providers as well as service providers with the database using the ROADIDEA uniform data archive (see Figure 1).

As shown in Figure 1 the identification of as much as possible data providers (data sources) is the most important issue in order to ensure sufficient and reliable data supply. Thus, it is necessary to get access to a huge amount of data sources as traffic data, weather data, infrastructure data, maps, schedules, etc.

Due to its data fusion tasks WP3 is closely connected to WP2. But also WP6 “Creation of Pilot Services” and WP8 “Validation and Evaluation” will have a close link to the data collection WP and will use the data mediation architecture.
2. Methodology

2.1 Data Collection Survey

From the beginning of the ROADIDEA project the data source identification had to be treated as an ongoing process. It is not finished within Task 2.1. The following Task 2.2 “New types of data needs” as well as the innovation processes in WP5 “Innovation procedures and management” might cause further data source identification and collection efforts. Especially pilot services defined in WP5 have to be fed with data. These data have to be up-to-date and of high quality. The service definition must not be limited to the already described data sets. The pilot services have to determine the necessary data.

The data collection survey referred to the data classification, which has been made in the Technical Annex1 “Descriptions of Work” to the ROADIDEA Grant Agreement [RI_GA, 2007], [RI_TA, 2007]. It distinguishes between data from vehicles, data from traffic control systems, data from infrastructure and weather data. Figure 2 shows the DoW data classification.

![Figure 2: Data classification according Description of Work](image)

It turned out soon that the data classification according the above mentioned DoW data classification was not suitable to classify all available data. The data collection survey showed that most data sets could be assigned to more than one class. A more detailed data classification was necessary. That caused the data classification work described in chapter 2.2. Nevertheless, the next paragraphs refer to the initial data classification.

The data collection questionnaire

The data source identification (data collection survey) started with the WP2 presentation at the ROADIDEA kick-off meeting in Budapest on 17-18 December 2007. After presenting the objectives of the work package and pointing out the necessity to collaborate in the data collection intensively every partner has been provided with a paper version of...
the data collection questionnaire (see Annex1). The questionnaire was also available as an online version at

http://trafficresearch.dlr.de/roadidea/wp2/

Both the offline and the online version included the same set of metadata elements that was defined based on the experience gained from describing and archiving a variety of datasets within the Clearing House of Transport Data1.

The combination of metadata elements primarily depends on multiple aspects: What kinds of datasets are to be described? How heterogeneous the data will be? Who is supposed to retrieve the metadata for what purpose? Thus, both the number of elements and the respective grade of detail may vary according to the actual objective. In case of the Clearing House, the set of metadata elements is rather comprehensive and covers not only content related information (such as thematic data category, variables etc.), but also detailed methodological specifications as well as access and contact information.

As the initial ROADIDEA data inventory was meant to get a general overview about data available within the partner's countries, the Clearing House's catalogue was reduced to a basic set of metadata elements in order to keep the questionnaire as short and clear as possible and to get a high response rate. The questionnaire asked for the following metadata elements, as short and clear as possible, and to get a high response rate:

- contact information (referring to ROADIDEA partners only)
- data ownership
- charges
- availability for third parties (in particular for ROADIDEA purposes)
- type of data (according the initial RAODIDEA Description of Work)
- abstract and objectives
- spatial reference/co-ordinates
- frequency of data collection
- availability of automated access
- format and size of the data file(s)
- availability of sample data sets
- availability of accompanying documentation (such as code books etc.)

The ROADIDEA partners were encouraged to use the questionnaire's online version. Many questions provided predefined answer categories, that way facilitating easy response and straightforward evaluation. Furthermore, most questions were indicated as 'mandatory', urging the respondent to answer. Supplementary open text fields allowed for more detailed input, if necessary. Fortunately, all partners followed the recommendation and used the online questionnaire for their dataset inventory.

---

1 The "Clearing House of Transport Data", operated by the DLR Institute of Transport Research, serves as main contact point in terms of transport related data. Its core element, a digital data archive, provides detailed metadata on available datasets such as content, scope and methodology, and access conditions. It addresses the overall transport research community including academia, public planning authorities, and private consulting engineers.

(http://www.dlr.de/cs/en/desktopdefault.aspx/1177_read-2160/)
Respondents were requested to consider existing and available data sets only. Further in-depth descriptions of particular datasets and their respective background were scheduled for the WP2 data workshop (see chapter 2.3 of this report).

According to the reduced Clearing House's catalogue the survey includes the following questions:

- Who is the contact person of the describing ROADIDEA partner?
- Who owns the data?
- What about data charges?
- Can the data be used in the ROADIDEA project?
- What kind of data?
- What is the spatial scope?
- Is the data accessible using an automated procedure?
- What are the data formats?
- What is the data set size?
- Is a sample data set available?

The ROADIDEA partners were also asked to upload additional documents related to the described data set as codebooks, format descriptions or other materials. These documents are to ease the understanding of the data and to support later analyses (Task2.3 and Task2.4). These documents should be in English.

The opportunity to upload additional materials has been a part of the data collection website. It could be done either directly after data set description or separately. The data collection web site has been hosted and managed temporarily by DLR for the ROADIDEA data collection period.

The data collection

The first data sets have been described and uploaded at mid of January 2008. In a first wave data were described which are in every day use or provided by ROADIDEA project partners.

Afterwards partners got in contact with various data providers in their countries to tap existing data sources with a varying degree of success. The objective of this work was to find, contact and convince road authorities and other data providers to collaborate in a research project context, which means to provide data free of charge for research purposes. Some partners could enter a fruitful field of accessible data sources easily, other partners run into problems with a lot of bureaucratic obstacles and very heterogeneous networks of public and private road authorities.

All described data sets have been checked and stored in an Excel spreadsheet file. This file was the base for later visualizations and analysis. After the data workshop it has been supplemented with more fields that have become necessary. The file has been used for the more detailed and technical oriented second wave of the data set description. The added fields are:

- Data set name:
  Every data set has to be named with a unique data set name, which makes it easier to get an idea of the data set contents.
- Data owner contact:
  It is to be the person (ROADIDEA Partner Company, Third Party Company, ministry, or department) that has to be called in case of data access in a more technical meaning. In terms of automated access to the data set the person is to be responsible for IT-means.

- Data classification:
  Beside the data classification according to the DoW data classification mentioned above every data set has also be assigned to one or more data class(es) of the data classification - resulting from the WP2 data workshop (mentioned in chapter 3.3 of this report)

- Data access:
  If an automated data access is possible the preferred method has to be defined. This should be discussed and prepared with the data owner contact.

The results of the data collection and the data availability investigation are listed in chapter 3 of this report.

2.2 Data Classification

2.2.1 Motivation

During the data collection partners hosting a big amount of different data sources (e.g. VTT and FMI) have had to answer the question of where to focus data collection efforts. What is the relevant data for the ROADIDEA project? This question led automatically to the more general one: What is the field the ROADIDEA track will go through? To keep the example, FMI and VTT were not able to describe all the data they are hosting, using, providing and in contact with.

Unfortunately nobody could answer these questions straight away. Pilot services demonstrating the new and innovative approaches of the ROADIDEA project are to be designed in the innovation processes. These innovation processes are to be targeted even in the WP5 “Innovation procedures and management” starting at a later project phase. Only if these pilot services are defined it is possible to concentrate on a focussed data collection. In the meantime one crucial question remains: What can be done to prepare a focussed data collection without knowing future innovations or future research fields?

Another problem was that data sets could not be assigned clearly to one data class defined in the DoW. A clear data assignment is urgently needed. Especially in a research project like ROADIDEA with different topical fields (e.g. road traffic and weather) it has to be possible to emphasize or define special links between data classes. This can only be done by using a reasonable data classification, into which all data sets have to be assigned clearly. Beside that, a data set assignment will facilitate the search for data for empty or sparsely filled data classes.

This can hardly be done with the initial data classification introduced in the DoW. Furthermore, data from traffic control systems inevitably are data from infrastructure, because traffic control systems are also parts of the infrastructure. Thus, the initial data classification includes only three data classes instead of four. This fact complicated the data class assignment additionally.
The assignment induced another discussion, which also could not be finished clearly with the available data classification. It was not able to distinguish between data from and data about a unit. Taking for instance a car as the unit, a classification assignment has to clarify the differences between data from the CAN-bus of a car (data from cars) and technical descriptions of this car (data about cars). Another example might me the differences between data from detectors included in a street stretch (that is "data from") and data about street capacities and other attributes (that is "data about"). It is crucial to clarify these issues during data collection to avoid later misleading interpretation and inadequate usage.

These aspects made it necessary to think about a data classification beyond the initial data classes stated in the Description of Work (DoW). Accordingly, data source analyses in the two participating countries Germany and Finland have been established to find state-of-the-art, research and visionary services and applications and corresponding data sources. Both analyses followed the innovation perspective shown in Figure 3.

![Figure 3: Innovation perspective (Source: Jörg Dubbert, PÖYRY)](image)

The perspective shows the different innovation levels and aims at the ROADIDEA innovation horizon, which lies beyond the state-of-the-art applications. The first level is only interesting as a basic reference. The second level focuses also on available technologies, but filters out the most advanced available solutions in daily operation, which have been implemented here and there. Examples: traffic-adaptive signal control, floating-car data (travel times). This level needs to be borne in mind. The third level is the level where possible future applications have been identified and are being tested. An everyday operation, however, has not yet been achieved. This level becomes interesting for ROADIDEA, and here the pilots will probably be located. Finally, the fourth level represents the visionary stage. Here, the brainstorming and conceptual designs on all kinds of conceivable and helpful future applications which tackle existing and future problems related to traffic and transport have to be opened. These ideas will probably not be realised during the project runtime, but they might appear in the roadmap.

These ROADIDEA pilots, to be defined in the innovation seminars (WP5), have to be fed by data described in the WP2 data collection.
Due to the participation of VTT (Finland) and DLR (Germany) in the project ROADIDEA it has been decided to apply the data source analyses in Finland and Germany. VTT as the largest Finnish research institution is providing a huge amount of data. VTT is also involved in a lot of national projects and initiatives dealing with various data sources. This also applies for DLR in Germany. The DLR Clearing House for Transport Data is the data provider for public founded data in Germany and has gained great experiences in data description, data classification and data handling. In co-operation with PÖYRY a comprehensive overview of available data sources in Germany could be obtained.

The results of these analyses were to be compared to derive a common data classification draft framework, which may be relevant for all European countries with respect to the ROADIDEA project objectives.

Finally, this data classification draft framework was to be analysed by the Clearing House for Transport Data in order to develop a synoptic data classification to be introduced and discussed during a data workshop.

**2.2.2 German Data Source Analysis**

The German data source analysis considered both state-of-the-art and innovative approaches and services. State-of-the-art systems are modern implemented systems which are in everyday operation. They are, however, not necessarily innovative as such. For ROADIDEA, they are the reference basis, a data source and a starting point of discussion. They are also a good basis for more innovative approaches which are supposed to be found in ROADIDEA.

Innovative approaches are most relevant for the development of ROADIDEA pilots. They are located at the cutting edge between everyday use and trial, and they are often content of research and development projects. They are only sometimes implemented as pilots, not very widespread or only in temporary use. However, in order to draw a roadmap for the future development they might be very relevant.

PÖYRY has investigated a huge amount of already available as well as prospectively conceivable data sources, which might be valuable for the ROADIDEA project. Figure 4 shows the fields where the investigation took place.
Within these fields about 20 systems have been identified. Every system includes on its part different data sources and data sets. For instance the Police TICs (Traffic Information Centre) collect information on the current traffic situation from a variety of data sources and distribute traffic related messages after verification to several recipients. Police TICs collect information about incidents (e.g. accidents, obstructions to road traffic, congestion or slow traffic flow) as well as road conditions (e.g. road weather warnings, road works).

During the data source analysis neither the comprehensive description of nor the direct access to these data sets were planned. The objective of the analysis was to identify data sources and to bring them into a systematic order (classification draft). Furthermore, all available information concerning the data sources (provider, system description, and data formats) was investigated. This investigation includes a huge amount of valuable information, which can be accessed easily after defining the pilot fields the ROADIDEA project has to operate on.

In a next step all identified systems and data sources were analysed and brought in a systematic order.

Three main classes were derived from the set of investigated data: Road Traffic, Road Weather and Public Transport. If applicable these main classes were subclassified (e.g. class “Road Traffic” was divided into subclasses traffic management, road observations, vehicle observations, misc). Figure 5 shows a first data classification approach. All systems and data sources from the German data source analysis have been assigned to only one data class clearly.
### German Data Source Analysis

#### Road Traffic

**Traffic Management**
- Network of Police TIC
- TIC of the road administrations
- Traffic-adaptive and traffic-dependent signal control
- Traffic Management Systems
- Motorway section, network and intersection control
- Electronic Parking Guidance Systems
- Tunnel Control Systems

**Vehicle Observations**
- Taxi FCD
- FCD (GATS)
- Floating Phone Data Trials
- X-FCD (cars & busses)
- Bus-FCD D

**Road Observations**
- Private radar detectors
- Permanent Counting Points
- Airborne traffic data detection

**Misc**
- Truck Tolling System (TollCollect)
- C2C communication/ C2I communication
- Traffic control based on pollution data

#### Road Weather

**In-Situ Measurements**
- Road Weather Data (SWIS)

**Airborne Observations**
- Road Weather Data

#### Public Transport

**Misc**
- AVL (Vehicle Location Systems) of PT companies
- Timetable ticketing systems of PT companies
- Ferry Data Pool (VIKING)

---

*Figure 5: Assignment of German data sources to an initial data classification*

As already mentioned above some systems utilize and refer to a wide range of different data sets which might need a division into further subclasses.

The shown approach is to be elaborated in the final data classification generation. The data source analysis in Germany was the base for the data classification work to be done by the Clearing House for Transport Data. The outcome of this investigation has been merged with the results from the Finnish data source investigation described in the following chapter.

### 2.2.3 Finnish Data Source Analysis

Finnish data source analysis was started by dividing the collection of data source descriptions among Finnish partners in ROADIDEA (VTT, Destia and FMI; Foreca was not participating this task). FMI provided information about weather related data sources and Destia about real-time traffic related data sources available in their Traffic Information Platform (TIP). VTT covered other available data sources in Finland. From innovation
perspective FMI and Destia provided information about data sources in every-day operation and leading state-of-the-art, VTT covered also some data sources in research level.

The FMI and Destia task in data source descriptions in Finland was clear but VTT task was more ambiguous. The focus of the ROADIDEA project is very wide covering weather, road transport with all its sectors, including multi-modality and other forms of transport. It was soon realised that there is a huge number of different data sources in Finland available. In order to match the resources in ROADIDEA for data source collection work and the number of possible data sources in Finland, it was decided to concentrate to the most important data sources (e.g. data sources in every-day operation) and most potential data sources in research level. However, VTT wanted to provide information about data sources with wide scope (including e.g. environmental, public transport, railway, etc. data sources) to support the broad focus of the ROADIDEA project and pending innovation seminar.

Information about available data sources in Finland was collected by a group of experts from different sectors including road infrastructure, telematics and traffic services, GIS/maps, public transport, road safety and environmental impact researchers. After the list of available data sources in Finland was generated current information services and applications utilised by different user groups was utilised to cross-reference that all important data sources (in every-day operation) were listed.

The description of available data sources in Finland was followed by the classification of data sources. From the start it was evident that the data classification in ROADIDEA DOW (vehicle, traffic control system, infrastructure and weather monitoring) was not enough. A large part of data sources could not be classified into these. From previous projects there were several data classifications available e.g. for traffic data, but no classification was available for the large scope of ROADIDEA. A new classification of data sources was needed. The classification of Finnish data sources was done in a brainstorming session in which selected experts from VTT participated. The classification was intended to be quite simple – with only five classes and limited number of sub-classes. Furthermore, classification was designed to be general - not specifically for Finnish data sources. The following list describes the VTT classification proposal for ROADIDEA data sources, which was presented in the ROADIDEA Berlin workshop.

- **Weather and environment**
  - Weather monitoring
  - Climate monitoring
  - Weather forecasts and warnings
  - Air quality
  - Noise
  - Pollution
  - Energy and environmental accounting
- **Infrastructure**
  - Geographical information (GIS/Maps)
  - Address databases
  - Point Of Interests (POIs)
  - Road registers
  - Road works
• Road condition
• Traffic
  • Real time traffic information
  • Traffic history information
  • Travel times
  • Cameras
  • Vehicle fleets -> Floating Car Data (FCD)
  • Emergency response centre data
  • Disturbances
  • Incidents
  • Accident databases
  • Availability of parking places
• Public transport and passenger information
  • Real time information trains/busses
  • Disturbances on public transportation
  • Operators, routes, stations/stops and timetables
  • Public transport usage/payment data
  • Trip planning
• Other data
  • Events
  • Vehicular and Driver data registers
  • Freight transportation schedules/routes
  • Terminal information
  • Vehicle bus data (CAN via FMS/OBD) and On-board unit data

### 2.2.4 Data Classification Proposal – Clearing House Approach

The outcomes of both the German and Finnish data source analyses were used to derive a consolidated data classification scheme applicable to all European countries with respect to the ROADIDEA project objectives. Furthermore, the thematic structure developed for archiving a variety of datasets within the Clearing House of Transport Data was taken as basis for the future ROADIDEA classification.

Main entry point to the Clearing House's digital archive is a hierarchical structure containing a dozen overlapping thematic categories (see Figure 6).

The main objective of this thematic classification is to support a contextual search. Depending on its respective scope, each data set hold within the archive is assigned to at least one category. If appropriate, a data set might be assigned to more than one category (see Figure 7). Resulting redundancies are intended, as they facilitate different search strategies. Additionally, an extended search mask is available that allows for targeted keyword searches within most meta-data elements.
The 12 main categories cover a broad range of topics and heterogeneous data sets. Therefore, a direct transfer of the present classification scheme to ROADIDEA is not useful. Some categories might be dropped while others might be changed or even added such as a particular category for data related to climate and weather. Figure 8 shows a preliminary adjustment of the classification scheme for ROADIDEA purposes.
In order to refine the new classification scheme, the following issues are to be considered. The classification could be structured according to:

1. the data content (e.g. air temperature, traffic volume, address/POI databases, etc.),
2. the data provider/data source (e.g. Road Administration TIC, Floating Car Data, etc.),
3. tasks and services derived from data (e.g. real time traffic information, road weather warnings, etc.)

In a next step – according to the data sources identified by the German and Finnish inventories – a comprehensive classification scheme consisting of 11 main categories was developed (see Tables 1-11 to 1-11).

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>mobility / travel behaviour</td>
<td>A1</td>
<td>TIC (police, road administrations)</td>
</tr>
<tr>
<td>passenger transport</td>
<td>A2</td>
<td>police departments</td>
</tr>
<tr>
<td>commercial transport</td>
<td>A3</td>
<td>emergency response centres</td>
</tr>
<tr>
<td>traffic participants/target groups</td>
<td>A4</td>
<td>inter-/innerurban traffic control centres</td>
</tr>
<tr>
<td>trip/transport purpose</td>
<td>A5</td>
<td>radar detector networks</td>
</tr>
<tr>
<td>transport mode</td>
<td>A6</td>
<td>permanent counting points (long-distance road network)</td>
</tr>
<tr>
<td>infrastructure [transport (management) system]</td>
<td>A7</td>
<td>camera networks</td>
</tr>
<tr>
<td></td>
<td>A8</td>
<td>maps (geographical information)</td>
</tr>
<tr>
<td></td>
<td>A9</td>
<td>TMC location codes</td>
</tr>
<tr>
<td></td>
<td>A10</td>
<td>address databases (POI, car parks, event locations etc.)</td>
</tr>
<tr>
<td></td>
<td>A11</td>
<td>road registries</td>
</tr>
</tbody>
</table>

Table 1-1 Category A "Infrastructure (Geodata, Transport System)"
<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>B. Infrastructure (Transport Management System)</strong></td>
<td>B1</td>
<td>road condition and weather information system</td>
</tr>
<tr>
<td></td>
<td>B2</td>
<td>electronic parking guidance systems</td>
</tr>
<tr>
<td></td>
<td>B3</td>
<td>automatic vehicle location systems (AVL)</td>
</tr>
<tr>
<td></td>
<td>B4</td>
<td>traffic-adaptive/ traffic-dependent signal control systems</td>
</tr>
<tr>
<td></td>
<td>B5</td>
<td>tunnel control systems</td>
</tr>
<tr>
<td></td>
<td>B6</td>
<td>truck tolling systems</td>
</tr>
<tr>
<td></td>
<td>B7</td>
<td>vehicle fleets (Floating Car Data)</td>
</tr>
<tr>
<td></td>
<td>B8</td>
<td>phone &quot;fleets&quot; (Floating Phone Data)</td>
</tr>
</tbody>
</table>

Table 1-2 Category B "Infrastructure (Transport Management System)"

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C. Infrastructure (Public Transport Management)</strong></td>
<td>C1</td>
<td>route networks</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>bus stop/ station networks</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>timetable/ ticketing systems</td>
</tr>
</tbody>
</table>

Table 1-3 Category C "Infrastructure (Public Transport Management)"

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>D. Public Transport and Passenger Information</strong></td>
<td>D1</td>
<td>real time information on trains/ busses (e.g. scheduled vs. actual arriving times)</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>disturbances on public transportation</td>
</tr>
<tr>
<td></td>
<td>D3</td>
<td>timetables</td>
</tr>
<tr>
<td></td>
<td>D4</td>
<td>routing tools (trip planning)</td>
</tr>
<tr>
<td></td>
<td>D5</td>
<td>public transport usage data</td>
</tr>
</tbody>
</table>

Table 1-4 Category D "Public Transport and Passenger Information"

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E. Traffic Monitoring (Road Observation)</strong></td>
<td>E1</td>
<td>traffic volume</td>
</tr>
<tr>
<td></td>
<td>E2</td>
<td>traffic density</td>
</tr>
<tr>
<td></td>
<td>E3</td>
<td>speed/ average speed, travel times</td>
</tr>
<tr>
<td></td>
<td>E4</td>
<td>traffic flow, congestion, level of service</td>
</tr>
<tr>
<td></td>
<td>E5</td>
<td>sudden obstructions to road traffic (e.g. incidents, accidents)</td>
</tr>
<tr>
<td></td>
<td>E6</td>
<td>road works</td>
</tr>
<tr>
<td></td>
<td>E7</td>
<td>road condition</td>
</tr>
</tbody>
</table>

Table 1-5 Category E "Traffic Monitoring (Road Observation)"
<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F. Traffic Monitoring (Vehicle Observation)</strong></td>
<td>F1</td>
<td>vehicle detection (type)</td>
</tr>
<tr>
<td></td>
<td>F2</td>
<td>vehicle detection (position)</td>
</tr>
<tr>
<td></td>
<td>F3</td>
<td>vehicle activity status</td>
</tr>
</tbody>
</table>

Table 1-6 Category F "Traffic Monitoring (Vehicle Observation)"

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>G. Traffic Management</strong></td>
<td>G1</td>
<td>messages about current traffic situation</td>
</tr>
<tr>
<td></td>
<td>G2</td>
<td>priority requests at traffic lights</td>
</tr>
<tr>
<td></td>
<td>G3</td>
<td>car parc occupancy, availability of roadside parking places</td>
</tr>
<tr>
<td></td>
<td>G4</td>
<td>routing/ evacuation/ parking management data</td>
</tr>
</tbody>
</table>

Table 1-7 Category G "Traffic Management"

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H. Freight/ Service Transport</strong></td>
<td>H1</td>
<td>freight transportation routes</td>
</tr>
<tr>
<td></td>
<td>H2</td>
<td>freight transportation schedules</td>
</tr>
</tbody>
</table>

Table 1-8 Category H "Freight/ Service Transport"

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>W. Climate and Weather Conditions</strong></td>
<td>W1</td>
<td>air temperature</td>
</tr>
<tr>
<td></td>
<td>W2</td>
<td>air pressure</td>
</tr>
<tr>
<td></td>
<td>W3</td>
<td>wind speed</td>
</tr>
<tr>
<td></td>
<td>W4</td>
<td>wind direction</td>
</tr>
<tr>
<td></td>
<td>W5</td>
<td>relative humidity</td>
</tr>
<tr>
<td></td>
<td>W6</td>
<td>intensity of precipitation (rain, snow, hail)</td>
</tr>
<tr>
<td></td>
<td>W7</td>
<td>black ice</td>
</tr>
<tr>
<td></td>
<td>W8</td>
<td>depth of snow</td>
</tr>
<tr>
<td></td>
<td>W9</td>
<td>visibility, brightness, fog</td>
</tr>
<tr>
<td></td>
<td>W10</td>
<td>road condition</td>
</tr>
<tr>
<td></td>
<td>W11</td>
<td>road surface temperature</td>
</tr>
<tr>
<td></td>
<td>W12</td>
<td>road humidity</td>
</tr>
<tr>
<td></td>
<td>W13</td>
<td>remaining salt on road surface</td>
</tr>
<tr>
<td></td>
<td>W14</td>
<td>climate and weather monitoring</td>
</tr>
<tr>
<td></td>
<td>W15</td>
<td>weather forecasts and warnings</td>
</tr>
<tr>
<td></td>
<td>W16</td>
<td>road weather monitoring</td>
</tr>
<tr>
<td></td>
<td>W17</td>
<td>road weather forecasts and warnings</td>
</tr>
</tbody>
</table>

Table 1-9 Category W "Climate and Weather Conditions"

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>J. Environmental Impact of Traffic</strong></td>
<td>J1</td>
<td>air quality</td>
</tr>
<tr>
<td></td>
<td>J2</td>
<td>emissions</td>
</tr>
<tr>
<td></td>
<td>J3</td>
<td>noise</td>
</tr>
<tr>
<td></td>
<td>J4</td>
<td>radiation</td>
</tr>
<tr>
<td></td>
<td>J5</td>
<td>energy and environmental accounting</td>
</tr>
<tr>
<td></td>
<td>J6</td>
<td>pollution reports</td>
</tr>
</tbody>
</table>
Table 1-10 Category J "Environmental Impact of Traffic"

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Class ID</th>
<th>Type of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other Data</td>
<td>O1</td>
<td>vehicle registries</td>
</tr>
<tr>
<td></td>
<td>O2</td>
<td>driver registries</td>
</tr>
<tr>
<td></td>
<td>O3</td>
<td>vehicle bus data (CAN via FMS/ OBD)</td>
</tr>
<tr>
<td></td>
<td>O4</td>
<td>C2C, C2I</td>
</tr>
<tr>
<td></td>
<td>O5</td>
<td>Metadata Platform</td>
</tr>
</tbody>
</table>

Table 1-11 Category O "Other Data"

However, some open questions still remain and were not answered during the workshop. First, it has to be clarified, which are the main objectives for the classification. Should it be primarily serve user needs or are technical requirements related to the implementation of the database more important? In this context, another question arises: Are redundancies desirable, necessary or do they contravene basic principles of database design?

In terms of the data subject: Should it focus on the source of data – the data provider – or rather the content (data "from" vs. data "about")?

Finally, with respect of describing meta-data elements: How to handle real time data vs. historic data covering the same subject, given that they are available in different formats? Or: How to handle data sets of the same category but of different format due to different origins?

These issues need to be discussed in the course of the project.

**2.3 The ROADIDEA data workshop**

The ROADIDEA data workshop was held in Berlin in March 2008. Every project partner participated with at least one person. The workshop was organized and chaired by DLR.

To save travel expenses the workshop was followed by a WP3 “Method and Model Development” workshop. Because of the close connection between WP2 and WP3 it was very useful to combine these workshops. This chapter only describes the part of the workshop dealing with data collection means.

**2.3.1 Objectives**

A clear understanding of the described data with respect to the systems and applications they are included in is a basic prerequisite for discussion of further developments of innovative ROADIDEA services and applications. During the workshop every partner have had the opportunity to describe not only the data set attributes but also the projects, applications or research fields in which the data sets are involved. The intention to do this was to initialize a process, which might enable the partners to see connections and synergies between data sets that have not been compared yet. Data quality issues were to be discussed too.

Beside that data sets were to be assigned to a data classification. As described in the previous chapters a more detailed data classification has been developed based on the data source analyses in Germany and Finland. This data classification was to be intro-
duced and discussed with the project members. This part of the workshop aimed at a general agreement to the proposed data classification. Furthermore the WP2 task allocation of work collaboration was to be introduced. In the run-up of the workshop every partner was asked how to collaborate in the WP2. These task allocations have been settled with the work plan (DoW tasks and more detailed sub-tasks) of WP2 by the work package leader DLR. The balanced task allocation was to be discussed and agreed at the workshop.

### 2.3.2 Methodology

To illustrate the data workshop methodology the agenda of the workshop is attached to this report (see Annex2).

The data workshop started with the presentation of the data sets. Every partner presented the data sets described during the data collection campaign. The focus of the presentation was on the data set attributes as well as on the contextual background. Partners described projects, applications and services with respect to the data utilised or created by these purposes. The following discussion of open questions deepened the general understanding for this data set.

Escorting the presentations the data sets have been assigned to the DoW data classification on a whiteboard to prepare the later data classification discussion.

The workshop continued with the presentation of the results of the German and the Finnish data source analysis. The results are described in detail in the chapter 2.2.2 and 2.2.3 of this report. With the following presentation of the less successful Croatian data source investigation started the discussion on the experiences in the other European countries. The Croatian examples pictured the obstacles and difficulties in the developing European countries very clearly. Maintenance of motorways and main roads are seldom in a public hand. Several public and private road authorities are responsible for parts of the national road network. They have different policies and levels of willingness to collaborate in European research projects and also different business models. Reliable structures and responsibilities are only in development. But not only are developing countries in the south-eastern part of Europe facing these problems of unclear or complex responsibilities and bureaucratic restraints. These experiences have been discussed and analysed in detail. Next steps and actions have been identified and its implementation has been set.

Originating in the results of the data source analyses a data classification suitable for the described data as well as for those, which are to be taken into further consideration during the ROADIDEA project was proposed. This data classification proposal was derived on the one hand from the data classification of the Clearing House of transport data with special respect to the needs and objectives of the ROADIDEA project. On the other hand it factors in the results of the data source analysis in different European countries described in chapter 2.2.2 and 2.2.3. This data classification and its applicability for the means of ROADIDEA have been discussed and adopted.

At the end of the workshop the overall WP2 task allocation table was introduced. It takes the work package collaborations of the partners (their person month and skills) into consideration and assures a reasonable distribution to all WP2 tasks and subtasks. This task
allocation table was discussed. All partners finally agreed on this table and their qualitative and quantitative WP2 collaboration.

### 2.3.3 Results

In general, the workshop was very valuable for all participants. The data presentations and very active discussions improved a general understanding of the data and their purposes. A lot of new considerations and ideas came up and have been discussed and borne in mind for later assessment.

The results of the workshop have been noted with a deadline on an action item list. It was included in the workshop minutes that have been distributed to the project members.

**Data collection**

The data set presentations showed the described data. It has been pointed out, that the data collection has to be treated as an ongoing process which has to be continued with new data acquisitions and with more detailed descriptions of the present data sets. In order to prepare (automated) data access the necessity arose to get more information about data contact persons. These persons are responsible for the data storage and handling and have to be called in terms of direct (automated) data access. More detailed information about data charges and data sizes are needed too.

It was decided to add fields to the data collection table (see Annex3) accordingly. This table was to be edited by the DLR. DLR sent it to the project members for updating with the needed information.

**General data availability**

The discussion about data availability in the participating European countries has been very comprehensive and valuable for all partners. In some countries the bureaucratic hurdles and its heights are well known. For other countries an investigation of these issues are to be initiated. Therefore it has been decided to investigate the general availability of public or private data sources with respect to the ROADIDEA project needs in the participating European countries. Beside VTT for Finland and PÖYRY/DLR for Germany the following partners have been tasked with these investigations:

- KLIMATOR (Sweden),
- DEMIS (Netherlands),
- ARPAV (Italy),
- ROADSAFETY (Hungary),
- METEO (Croatia, continuing their efforts) and
- AMANOVA (Slovenia).

The results of these investigations are listed in the chapter 3.2 of this report.

**Data classification**

After discussion of the data classification proposal all project partners have agreed to this classification. This data classification is not to be treated as carved in stone. It will be adapted to the ROADIDEA needs constantly.
3. Results

3.1 Data Collection

As mentioned before the survey started with the WP2 presentation at the ROADIDEA kick-off meeting in Budapest at the end of 2007. The survey will be an ongoing process, so no end date has yet been defined. This means that, if the ROADIDEA project partner have additional data sets they still can use the questionnaire to describe there data. Until now 48 questionnaire responses have been received. The Chart 1 shows the number of responses from each project partner.

![Chart 1: Number of questionnaire responses](chart1.png)

As visible in the chart VTT has used the questionnaire quite a lot. But, almost every project partner has described some data.

The Chart 2 represent the question “Who owns the data?” with the following results: 20 data sets are owned by the government, 8 data sets belong to project partners, 11 data sets are owned by a third private party and the other are owned by more than one party.
A very important question is: Can the data be used in the ROADIDEA project? In the responses 42 data sets can be used in the project and 5 data sets cannot. The results are shown in Chart 3.

Furthermore it has been asked what about data charges? The Chart 4 shows the results. For two data sets charges are required, 37 data sets are free of charge and for remaining 9 data sets a determination is needed if charges are required or not. This further investigation was decided in the data workshop described in chapter 2.3.
Chart 4: What about data charges?

Also very interesting to know is: What kinds of data are available? The results are shown in Chart 5. The question referred to the data classification, which was defined in the ROADIDEA description of work. 15 data sets belong to “data from infrastructure”, 7 to “data from traffic control systems”, 14 to “data from vehicles” and 16 data sets belong to “weather monitoring”. Some of the data belongs to more than one category.

Chart 5: What kinds of data are available?
Furthermore, there are 13 data sets, which are not suitable for any one of these categories. The following list shows received answers:

- accident information
- background information
- aggregated or calculated data
- direct communication with officers
- information from rescue calls
- road weather analysis, forecasts & warnings
- transport information

This issue caused the data classification efforts, where these answers flow into the data classification proposal, which is described in chapter 3.4.

The Chart 6 represents the results of the question about the spatial scope. There is at least one data set from every country in the ROADIDEA project. Some of the data belongs to more than one country. In 11 of the data sets other countries were mentioned. Some described data sets referred to regions listed as follows:

- Austria,
- the Baltic Sea,
- Denmark,
- Estonia,
- Norway

Some of the data set belongs to whole Europe.

![Chart 6: What are the spatial scopes?](chart6.png)

Another question of the survey was: Is the data accessible using an automated procedure? The Chart 7 shows the results. For 32 data sets an automated procedure is possible and for the other 16 data sets it is not possible. In this context it has also been asked for the access method. Chart 8 displays the results. Seven data sets can be retrieved us-
ing a database, 10 data sets can be retrieved by using the file transport protocol and 18 data sets can be retrieved via http. For 8 data sets were other methods mentioned. Some of the data sets are accessible using more than one method.

![Chart 7: Is the data accessible using an automated procedure?](image)

![Chart 8: Access method](image)

The Chart 9 shows the results of the question: What are the data formats? 20 data sets are regular flat text files (ASCII format), 5 data sets are images, 9 data sets are in XML and for 13 data sets other formats were mentioned for example gme, pdf, or xls.
Finally the survey included a question about sample data sets and a question about related information to the data sets. The Chart 10 shows the results for the sample data sets. For 28 data sets we can retrieve sample data sets and for 20 data sets there are no sample data sets available.

The Chart 11 shows the results for related information. For 8 data sets English codebooks are available, 3 data sets have a non-English codebook. Furthermore, 2 XML-schemas, 8 other documents in Finnish and 25 other documents are reported.
3.2 Data Availability

At the data workshop it was decided to investigate the general data availability of data from public and private road networks. Early in the data collection period the Croatian partner was confronted with a very complex network of public and private road authorities, special responsibilities and different willingnesses to collaborate. Also in Italy it is not possible to name only one road authority to establish a general contact to get data access to all segments of a nationwide road network. In some regions private road authorities manage maintenance issues of the complete road network, other are completely managed by public road administrations. Sometimes only one important transit road is private administrated. To specify these issues in all participating European countries some partners have been asked to get in contact with the relevant road authorities. The objective of these activities was to find general responsibilities, persons, able to establish reliable access to well-specified data and to convince these authorities and persons to collaborate in the research project ROADIDEA.

It is also important to name countries or regions where a data access cannot be established.

During the first data collection initiative partners from some participating countries have already described the general weather data availability. Within the following data availability investigation also weather data had been taken into consideration to obtain a good coverage all over Europe.

The table in Annex4 gives an overview about the data availability in the ROADIDEA member states.
3.2.1 Finnish data availability

Data from road networks

At present, there are approximately 350 permanent traffic counting stations on the Finnish road network. These counting stations are basically loop detectors, which have been placed on major interurban roads and on major roads in urban areas. A map of their locations can be found on the web site of Finnish Road Administration (Finnish Road Administration 2008). The counting stations have communication links to the Finnish Road Administration. Real-time data is available from some counting stations while historic data base is available of them all.

In addition to permanent traffic counting stations, Finnish Road Administration has also a travel time information system, which collects information on travel times in urban areas and on interurban main road network. Historical data of travel times is available via http download as compressed text files. Real-time data is also available to information service providers. At present, the system covers Helsinki urban area, but there are plans to extend it to interurban main road network. Finnish Road Administration provides information about traffic, road weather and road works via their web-pages (www.tiehallinto.fi).

Floating Car Data (FCD) is collected by Infotripla and a local taxi company (Tampereen aluetaksi) in Tampere region. Real time traffic information is calculated from FCD data combined with traffic light and fixed measurement data. Traffic information is provided via web page (http://www.liikennetampereella.fi). Destia collects FCD from taxis in Oulu.

Figure 9: Travel time information system in Helsinki metropolitan area (Finnra 2008d).
region and provides traffic information via web page (http://www.oulunliikenne.fi/). Destia will provide FCD-based traffic information also in Helsinki area in near future.

Destia Traffic offers dynamic traffic information in Finland, Sweden, Norway, Denmark and Estonia. Destia uses several information sources to validate traffic data national road administrators (induction loops, traffic cameras, road weather data, etc.), radio stations, road users and several other partnership companies. Destia Traffic provides information on traffic flows, disruptions and weather conditions. At the www.destiatraffic.com website, a map service is available for Sweden, Norway, Denmark and Estonia. This service incorporates a route and address search function, and shows road works and other traffic disruptions. Destia also provides traffic information to navigators and mobile phones (via TMC or GPRS).

Weather data

An extensive network of road weather monitoring stations is operated by the Finnish road authority. In the end of year 2006 there were 330 road weather monitoring stations on the Finnish road network (Kulmala, Karhumäki 2007).

Figure 10: Locations of road weather monitoring stations (Finnish Road Administration 2008c)
Data produced by road weather monitoring stations is available online. For example, real-time temperature information is published on the web site of Finnish Road Administration.

Finnish Meteorological Institute (FMI) has its own network of automatic weather observation stations. At the end of year 2007 FMI had 150 automatic weather observation stations. Most of them – 90 stations – produce real-time information by sending data once in ten minutes to FMI. For road users FMI provides weather forecasts and road weather warnings e.g. via web pages (www.fmi.fi).

From FMI data VTT has calculated freezing index and thawing index data. The surface freezing index represents the amount of heat loss from the road surface occurring over a year. Respectively the surface thawing index reflects the heat entering into pavement. Freezing index is used to estimate frost penetration during winter and thawing index is used to estimate thawing during spring and summer. Statistical data is available for (air) freezing and thawing index. Mapping of weight limitation and thaw damage has been carried out on low-volume roads more than last 10 years in Finland. Using this data, risk of thaw weakening can be estimated on local levels.

Other data

There are several data sources in Finland related to road infrastructure. Digiroad is the national street and road database in Finland. The whole database can be ordered as a CD or DVD-ROM from Finnish Road Administration, and some sample data is available at the Digiroad website (http://www.digiroad.fi).

Finnish Road Administration has a reporting data base called AURA, which is used internally. It collects service and maintenance (both summer and winter) information (volume, quality, cost, timing) from regional authorities and contractors. Finnish Road Administration has also a road Condition Data Bank, which contains information of surface condition of all paved and public Finnish roads. Information contains measured information of longitudinal and transverse evenness, amount of damages, and level of bearing capacity. This information is used in planning of maintenance and reconstruction works if future. Another road specific register is called the Road Data Bank. It contains all information about roads maintained (owned) by Finnish Road Administration. Information included is: location, road junctions, road structure, maintenance actions, etc. Theme data base about properties of public roads to be used in maintenance planning, maintenance and project level preliminary design.

There are also other data sources with historical data. For example, statistics on road accidents are available from traffic insurers (http://www.valt.fi) and Statistics Finland (http://www.stat.fi). Finnish Road Administration also has a database of all road accidents reported by the police.

VTT together with Helsinki City Planning Department have utilised videos acquired from cameras mounted on buildings or aboard helicopters to assist road user safety studies and traffic counts. The TrafMON system, developed by VTT, processes these videos to estimate vehicle position and speeds along the sequence. The software has been used by Helsinki City Planning Department for example to count pedestrians crossing the streets outside the zebra crossings (from helicopter) and to study the behaviour of drivers at zebra crossings (from ground base). Data is available for ROADIDEA project.
There are number of public transport related data sources available. Electronic ticketing systems used in public transport gather information on travel times, passenger count information, fare income information and information on the movements of public transport vehicles. These systems are in operation in several cities. Helsinki Metropolitan Area Council operates the system in Helsinki. Data from the Helsinki area public transport ticketing system is available upon request for research purposes.

Helsinki Metropolitan Area Council also has public transport register, which contains timetables of public transport services in Helsinki region, time used on links, bus stops and information on the operator of a particular public transport service. This register covers geographically only Helsinki metropolitan area. Public transport register data is available for download via ftp on request.

Finnish Rail Administration has a train timetable database which contains information on timetables, cancellations and modifications to already published timetables and information on wagons and locomotives of a particular train. The database is updated regularly.

As traffic related environmental impacts have increasing importance emissions of vehicles are monitored and reported more and more. Unit emissions such as CO, NOx and PM per kilometre for transport vehicles are available from VTT (http://lipasto.vtt.fi). Statistical data on emissions of transport vehicles is available from EMISTRA – Energy and Environmental Accounting and Reporting System for Transport and Logistics (http://www.emistra.fi).
3.2.2 German data availability

In Germany, there is a wide range of systems which could be used as data sources. This chapter is supposed to provide an overview of the potential sources. In the Annex, there is a detailed list of systems together with indications on the type of data and the contact details.

Interesting systems could be:

- the traffic information systems of the police

  The system is organisation according to the German administrative structure. Each Federal State has an own TIC, and there is on overarching centre for the whole country. The System mainly provides online data on incidents and congestion on the major roads. It is the basis for the RDS-TMC service in Germany and provides many service providers and media with information on the current traffic situation.

- the traffic information systems of the road administrations in some federal states

  In some federal states, usually in those with the higher traffic loads the road authorities decided to have own traffic information centres where the functionality and the data pool is extended in comparison with the TIC of the police.

- traffic-adaptive urban traffic control systems

  A relatively new tendency in Germany is the implementation of traffic-adaptive signal control in Cities. In Germany most BALANCE and MOTION are in use. The “advantage” of these systems is that they need complex online measurements and they do modelling and prediction. So data on the traffic flow is available in those systems.
• regional and urban traffic management systems

In larger urban agglomerations, traffic management systems can be found which try to bring together information from a variety of subsystems and make an attempt to influence the traffic situation in the City or region based on strategies. The may be the source of a variety of different data.

• the inter-urban traffic control system

Interurban traffic control systems are mainly responsible for section and network control on motorway. They usually work automatically, and they use online measures data for the detection of traffic flow and road weather. They do short-term prognoses, and they also provide aggregated traffic massages on the road sections which they control. These systems are in the hands of the road administrations in the respective federal states.

• the permanent counting points

Permanent counting points exist in every Federal State. They were once installed for statistical purposes in order to measure the long-term development of the road traffic in Germany. Nowadays, the counting points are more and more brought online as valuable data source for traffic control and information. The counting points are in the hands of the road administrations in the federal states. Collected statistical data could also be obtained from the German Federal Highways Institute (BASt).

• the private traffic information service providers

Private traffic information service providers use the publicly available traffic data, however they also operate their own additional data collection systems in order to upgrade their services. T-Systems is one on the most important service providers. The daughter DDG is responsible for data collection and generation of content. DDG operate a private radar detector system and a floating-car fleet in order to obtain improved traffic flow data.

• Taxi FCD schemes in urban areas

Taxi FCD is a method to obtain travel time data from urban areas. There we have a high concentration of taxis reporting their travel times for the generation of a traffic situation presentation. The advantage of taxi FCD is that the data can be produced as an additional functionality of taxi fleet management systems. There are implementations in several German Cities such as Berlin, Hamburg, Nuremberg, Hanover, and Dresden.

• The German truck tolling system TollCollect

Although the system is by law currently not allowed to be used for other purposes than toll collection (it would create a monopoly), the system has a large potential for the use of truck-related floating-car data in future.

• Electronic parking guidance systems
Practically in any larger City parking guidance systems can be found which can deliver information in free parking lots. They are in the hands of the municipalities.

- Road tunnel control systems

Road tunnel control systems can provide traffic flow information. As they are also relevant for road safety, some tunnels have also incident detection and warning systems.

- the SWIS road weather monitoring system

SWIS is the German road weather information system. Stationary detectors provide all relevant road weather data (e.g. temperatures, visibility, wetness, black ice warnings...). SWIS is operated by the German road administration in the Federal State and the DWD.

![Figure 13: Distribution of SWIS Station in Northern Germany (Source: VIKING/ Pöyry Infra Traffic)](image)

- Public Transport AVL systems

Practically in any larger city there are AVL systems which monitor the position and the status of public transport vehicles. They are operated by the PT companies.

- the passenger information system DELFI
DELFI is the interesting attempt to set up a PT traveller information system across all major PT operators in Germany with the aim of providing seamless traveller information for the public transport user.

- the VIKING ferry data pool

The Ferry Data Pool is a project which is supposed to provide timetable information in the Baltic Sea Region for all ferries, no matter to which company they belong.

- Floating phone data trials

There are trials which attempt the determination of travel times based on the localisation of mobile phones. One current example is the project on a motorway in Lower Saxony.

- Traffic monitoring for airborne sources

There have also been trials to determine the traffic situation from airborne sources (e.g. helicopters, balloons, airplanes) and from space (e.g. via radar). The DLR has experience with this.

- Bus FCD

Also travel times and other data from busses floating in the urban traffic are seen as an interesting data source. One example is the City of Hanover, where floating-bus data have been merged with floating-car data from other sources.

- Co-operative Systems

Currently, there is a modern research topic also in Germany: the co-operative systems which are built on the assumption that road traffic management can be improved by the inter-linkage of in-car computers of different vehicles (Car-to-car) and with intelligent roadside infrastructure (Car-to-Infrastructure). There are two larger EU projects (COOPERS, CVIS) with German participation which support these ideas. There is furthermore the German AKTIV consortium with major players from the automobile industry involved.

- Monitoring systems for pollutants

One modern tendency in Germany is also traffic control depending on the measurement of pollutants (e.g. NOx). There are respective products on the market.

PÖYRY has generated a comprehensive table which includes contact information to these data sources. With respect to the data privacy of the listed persons and institutions this table is not added to the report.
3.2.3 Swedish data availability

Road Weather Data

The Swedish road weather information system (RWIS) has 720 field stations located along the roads in Sweden. These field stations are located at sites along the roads to be able to give early warning for risk of slipperiness and severe weather conditions during the winter season. The siting of the field stations reflect the local conditions which is often an extreme climate determined by the surrounding.

The field stations are basically equipped with sensors for recording of following parameters:

- Air temperature (°C)
- Road surface temperature (°C)
- Relative humidity (%)
- Wind speed (m/s)
- Wind direction (degrees)
- Precipitation type (snow, rain, sleet)
- Precipitation amount (mm)

The data is recorded and sampled with 30 minutes interval. All data is stored at the Swedish road administration in Borlänge. The Swedish road administration is the owner of the data. The availability of data is restricted to maintenance organisations decided on by the road administration. It is also possible to apply for data for research projects.

Data from the road network

Data about accidents are also kept by the Swedish road administration the database STRADA is based on police reported accidents in different parts of Sweden.

The NVDB is the main data base for the road net work in Sweden.

The Swedish National Road Data base – NVDB: There is a great need for digital road information. Several companies are already initiating programs to introduce advanced electronic in-car navigation systems. Transport companies and the forestry industry are building systems for making transport more efficient.

Road traffic safety and road management are other areas of major development potential. The Swedish government has therefore issued directives to create a nationwide road database, containing up-to-date information that fulfils particular quality standards. The Swedish National Road Database, NVDB, is unique in its scope and content even on an international scale.

The aim of NVDB is to meet the immediate and long-term need for fundamental road information, and to be accessible both to the public and private sectors. The purpose is to create the right conditions for a breakthrough of ITS.

The NVDB contains different types of features such as:
• Types of administration
• Types of traffic regulations
• Types of technical features
• Special types of features for transportation on forestry road network
• Other NVDB features
• SRA features – compulsory road data for the state road network
• Traffic data features
• Bridge and tunnel features.

The NVDB data base is available as at the Swedish road administration www.vv.se/nvdb. See also appendix: SRA_NVDB_RoadData.pdf. The cost for this data is calculated depending on the amount of data needed.

STRADA - Swedish TRaffic Accident Data Acquisition

STRADA (Swedish Traffic Accident Data Acquisition) is an information system concerning data about injuries and accidents within the entire road transportation system. The data is based on reports from the police as well as from the hospitals and medical care. STRADA take-out of data may be done either as an Access database or by a special client tool.

In the STRADA Take-out you can choose a take-out on a national level or defined to a certain area such as a county or a specific street. It is also possible to make a selection by use of different criteria. It is possible to search for accidents or (injured) persons and if these are known by the police and/or the medical care. Data is distributed by the SRA. More information can be found on SRA home page for information in English. http://www.vv.se/templates/page2_2____13172.aspx?epslanguage=EN

Example of data: http://www.vv.se/templates/page3____13201.aspx

See also appendix: SRA_STRADA.ppt.

Floating car data

The technique for data acquisition, described below, was tested winter 2006-2007 (20 cars) and winter 2007-2008 (102 cars), with good results.

Technique:
1) Data is read from the cars sensors via the CAN-bus.
2) The data is marked with current time and current GPS-position.
3) The data is continuously sent to a central database.

Structure:

Background data (sent every 30s)
• Position (long, lat)
• Time
• Speed (not derived from position per time)
• Course
• Identity of individual car
• Air temperature
Event data (sent on activation of cars safety system)

- Position (long, lat)
- Time
- Speed (not derived from position per time)
- Course
- Identity of individual car
- Information of which one of the safety systems, activated the event report.
- Friction estimation (on favourable road conditions)

Data for floating cars are available only as test samples by Caran.

Airport weather contains:

- Observation METAR
- Forecast TAF

All airports have while its open manually observations with automatic tools two times every hour. Most airports are open 24h, a reason why this information is reliable. Airports not open 24h have automatic observations. The observations are always made the same time around the world, xx:20 and xx:50.

Airport observation contains data about:

- Wind direction
- Wind speed
- Wind gust
- Significant weather; e.g. Fog, rain type, snow, rain snow, thunder, ice, sand-storm, smog, etc.
- Temperature
- Dewpoint
- Air pressure
- Visibility
- Sky condition, cloud types and level
- More information for pilot as runway condition etc

The format is a global standard (since 1968) and the data is available at a lot of places close to aviation. For example http://www.aviationweather.gov

The data is also free because it is seen as safety information.

For more information, please see the document “Airport weather observations combined with satellite images” uploaded under Road Idea / WP2.

### 3.2.4 Dutch data availability

Data from road networks

The road network in the Netherlands consists of motorways and regional roads managed by the Ministry of Public Works, regional and local roads managed by the Provincial authorities and local roads managed by communities.
The motorways are monitored by about different traffic control systems, such as inductive loop detection, GPRS stations and Astrid MTM2 stations. Inductive loop detectors are used to detect the traffic flow data, volume and speed. The density of traffic is calculated in the centre. Web cameras are installed for internet use and special camera systems check the vignettes.

An overview of the measurement locations is given in Figure 14:

![Figure 14: Dutch road traffic measurement locations](image)

The MTR+ Wegwerk data bank contains information about the main technical (infrastructural) road parameters, and traffic counts by hour, day, month, year and type of vehicle.

Actual speed measurements are also made available via the internet; see for example a screenshot of the ANWB web site in Figure 15 (note the traffic jam in the top left corner from 20:00 on a Sunday!):
Figure 15: Speed measurements are also made available via the internet.

Weather data

Weather data are provided by the Dutch Royal Meteorological Institute KNMI. They provide historical, actual and forecast datasets:

1. Historical data by weather station
2. Actual data by weather station
3. Actual and predicted (2 hours ahead) Radar precipitation data on a 1 by 1 km grid
4. XHIRLAM model weather forecasts for 9 days on a 11 by 11 km grid.

The KNMI data are usually only available at a charge, unless it’s for research that leads public publications.

Other data

The road, bicycle path and footpaths network from OpenStreetMap.org (provided by AND www.and.nl) is available at no cost under the Creative Commons License. This network can be used in case a Dutch pilot is defined.

3.2.5 Italian data availability

In this section a summary of weather and other data for the north-eastern Italian region Veneto and some of its surroundings is presented. Unlike in other European countries,
meteorological services in Italy are fragmented and quite, but not completely, independent. One consequence, which is relevant for Roadidea is that weather data are not handled on a national level but are managed on a regional level on distinct and independent systems. Accordingly, there is no uniform data policy, which would govern the access to these data. Data access for general purpose is relatively restrictive, possible for research, yet subject to bilateral request procedures with every individual data producer.

For transport and traffic related data the situation is somewhat similar, in that there is not a single managing authority for the entire motorway network. ANAS, a joint-stock corporation with the "Ministero dell'Economia e delle Finanze" as the sole shareholder, manages the road and motorway networks of national importance. The actual construction and management of most toll motorways and tunnels is committed to more than twenty private companies or concessionaires. The largest concessionaire is "Autostrade per l'Italia S.p.a" which is also responsible for two stretches on the Veneto territory. In Veneto there are four more private concessionaires, which are responsible for four more stretches. Consequently, there is no simple and overall entry point for traffic related data. Moreover, data access with the private companies will have to be negotiated bilaterally. On a positive note, the European effort DATEX II on homogenization of the exchange of traffic and road weather data is having a beneficial impact for the Italian situation.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Owner/provider</th>
<th>Notes/availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weather data Veneto</td>
<td>ARPAV</td>
<td>Available for project</td>
</tr>
<tr>
<td>Weather data: boundary layer profilers Veneto</td>
<td>ARPAV</td>
<td>ditto</td>
</tr>
<tr>
<td>Weather data: meteorological satellite</td>
<td>EUMETSAT</td>
<td>available via EUMETCAST (<a href="http://www.eumetsat.int">www.eumetsat.int</a>)</td>
</tr>
<tr>
<td>Weather data: visibilimeter network Veneto</td>
<td>ARPAV Private motorway concessionaire</td>
<td>Arpav visibilimeter network: to be established within Roadidea; fully available for project Motorway concessionaires network: 1 concessionaire data available for ARPAV Roaddea activity, but not in real time</td>
</tr>
<tr>
<td>Webcams</td>
<td>Various owners</td>
<td>Usefulness to be evaluated</td>
</tr>
<tr>
<td>Weather data rest of Italy</td>
<td>National and regional Met. Services</td>
<td>ECOMET; bilateral agreements; use not planned for Roadidea</td>
</tr>
<tr>
<td>Traffic-related data Veneto (traffic conditions, mean traffic velocity, accident occurrence, presence of road maintenance work)</td>
<td>5 privately managed motorway concessionaires</td>
<td>Bilateral negotiation needed; (contact with 1 concessionaire for Roadidea, to be further developed) Partially available on internet</td>
</tr>
</tbody>
</table>

In summary, any data collection exercise at the national level in Italy will be faced with the fragmented nature of both the meteorological services as well as the managing authorities for the management of the national motorway network. What is reported here is, therefore, not exhaustive and reflects mainly the situation in the political region Veneto. As far as other regional meteorological services are concerned, however, the data types available are very similar to what is listed for Veneto.
3.2.6 Hungarian data availability

Data from road networks

The road network in Hungary consists of 1000 km motorways, 30 000 km public roads and 100 000 km streets in build up areas.

3 different traffic control systems are in operation on the motorway network for the time being. Inductive loop detectors are used to detect the traffic flow data, volume and speed. The density of traffic is calculated in the centre. On the motorway network 25 local weather observation stations (UTMET) are working and transferring the information of weather situation to a centre, mainly for the winter maintenance service. The motorway accident data are collected by police and partly by road managers. Web cameras are installed for internet use and special camera systems check the vignettes.

On the public road network are 300 permanent traffic counting stations some of them with direct data transfer to the centre. The Road Data Bank (OKA) contains information about the main technical (infrastructural) road parameters, and traffic and safety data of national roads.

Weather data

The other part of the mentioned UTMET system consists of 250 weather stations on the public network. Very detailed information is available from the UTMET centre about the microclimatic situations for maintenance activity but not for other road users at the moment.

Other data

Data sources in the urban areas are the traffic signal centres, and the control stations of public transport. No data from private cars, but a trial is going on in Budapest using floating cars.

The majority of data are historical and free of charge. Real time data are available from control centres, except the police system.

3.2.7 Croatian data availability

There are five concessionaries for all roads in Croatia. The biggest one is Hrvatske ceste Ltd. who is in charge for all ordinary roads in Croatia. Other four are in charge for highways. The biggest highway concessionaries is Hrvatske autoceste Ltd, second is Autocesta Rijeka Zagreb s.c. (stock company), third Autocesta Zagreb-Macelj Ltd and fourth Bina-Istra s.c.

Till today all of mentioned companies have not been wiling to give any sample of their data. Attempt to get data samples begun in Ministry Of Transportation to which a contact has been established. There a kind of Letter of support for all concessionaries Managing boards was created. After contacting all concessionaries names of persons in charge, which should fulfill the ROADIDEA requests were known. But problems always occurred on some lower level in company structure.
In all contacts, at least, a general knowledge of the way of data collecting could be obtained. Even though no samples of data sets were available, real time data access for ROADIDEA seems to be feasible to realize.

**HRVATSKE CESTE Ltd** ([www.hrvatske-ceste.hr](http://www.hrvatske-ceste.hr))

Hrvatske ceste Ltd (Croatian Roads Ltd) maintenance all ordinary roads in Croatia and they were non-cooperative in highest level of management. It was not able to manage to organize any meeting with them. They have contract with National Meteorological Service (NMS) about Automatic Meteo Stations (AMS) data handling and providing with weather forecasts.

Figure 16 shows the highway network in Croatia.

**HRVATSKE AUTOCESTE Ltd** ([www.hac.hr](http://www.hac.hr))

The Zagreb-Split highway route includes a part of Zagreb-Rijeka highway which is in charge of Autocesta Rijeka Zagreb Company. Part from Bosiljevo to Split is 313 km. On this route should be about 40 AMS.
The 306 km long Bregana-Lipovac route of highways (see Figure 17) is connecting Slovenian and Serbian borders. The Number of AMS is unknown. Figure 18 shows the 97 km long Zagreb-Gorican route from Zagreb to the Hungarian border. This route is mostly in project.

Traffic data from these routes could be defined as numerical and descriptive. Numerical data include meteo data and traffic count data. Descriptive data include video cameras images/movies, road accident reports and road maintenance reports. Numerical data are collected. These data can be used for some analyses. It is not known how much AMS they have, but they are standard Vaisala road AMS measuring wind, pressure, air and road temperature, moisture, precipitation and road surface condition.

This concessionaire has contract and very close cooperation with NMS about meteo data handling and weather forecast supply. Croatian laws force state companies to deal only with other state companies. Only if no one of state companies gives needed service or product, state companies may deal with private sector. Price and other conditions are not important.
AUTOCESTA RIJEKA ZAGREB s.c. (www.arz.hr)

Autocesta Rijeka Zagreb is charged for highway routes between Zagreb and Rijeka, the Slovenian border and Rijeka, and the bridge to the island Krk (see Figure 19 and Figure 20).

![Figure 19: Routes Rijeka Zagreb](image)

Data can be selected in two categories: numerical data and descriptive data. Numerical data include AMS data and traffic count data. Descriptive data include accident reports and road maintenance information. Numerical data are stored in databases, due to last information in Oracle database.

Significant data (air temperature, road surface condition, wind warning, maintenance or accident warnings, etc) are displayed on several electric road displays, very densely covering whole route. There are at least 20 AMS on route, mostly standard Vaisala road AMS. AMS are located on bridges and viaducts (see Figure 21).
Figure 21: Vaisala automatic meteo station

Measured types of data are air and road surface temperature, pressure, humidity, wind speed and direction, precipitation and road surface condition. AMS are not calibrated by some official institution. Company has a contract with NMS for meteo data handling and weather forecast supply.

All critical points are covered with video cameras, but only in visual spectrum (no IR). Most of the cameras are in tunnels and on pay points.

**BINA-ISTRA s.c.** ([www.bina-istra.hr](http://www.bina-istra.hr))

Bina-Istra Operation and Maintenance Ltd is a company for the road maintenance and is completely in Bina-Istra s.c ownership.

Actually, Bina-Istra route shown in Figure 22 is not real highway because it has only one roadway, so it is rather a road for motor vehicles. In some future it is planned to make it real highway. It has three main parts: first from Pula on the south to Kanfanar where it splits in west and east parts. West part ends on Slovenian border and east one in tunnel Ucka.

Figure 22: Bina-Istra-Route
There are only two AMS on the whole route and unknown number of traffic counters. Most information on route, except from tunnel Ucka, is collected by road maintenance patrols. Tunnel Ucka (5000m) is well equipped with video cameras and fire/smoke sensors. Web page offers information about maintenance works, while in the tunnel Ucka most of the domestic radio broadcasts are interrupted with brief information about tunnel.

Radar information signs are positioned at six selected locations along the Y to show drivers their current speed.

**AUTOCESTA ZAGREB-MACELJ Ltd** ([www.azm.hr](http://www.azm.hr))

Autocesta Zagreb-Macelj is charged for one 60 km long highway route from Zagreb to Slovenian border.

### 3.2.8 Slovenian data availability

The Slovenia partner Amanova in the Roadidea project works on statistical modelling of traffic activity in changeable weather conditions, prediction of severe conditions in winter traffic, and modeling influences of salting on road conditions and degradation of structures. For this purpose contacts have been established with the main providers of data in Slovenia, which are the Slovenian Agencies for Roads and Environment. Their web page addresses are: [www.dc.gov.si/en/](http://www.dc.gov.si/en/) and [www.arso.gov.si/en/](http://www.arso.gov.si/en/). Both web pages are structured similarly with boxes representing the general properties of agencies as well as their specific characteristics. Many data of interest are represented in specific boxes, such as roads, traffic, vehicles, and air, weather, water, etc. the data are represented either as on-line or processed mean data.

In Slovenia there exist three types of traffic and weather data:

- free available on internet or obtainable directly from both agencies
- commercially accessible over internet, phone or CD, and
- hidden data that can be extracted from internal data bases as determined by specific contracts with agencies.

Contacts with data holders can be established by using information from web pages of agencies. Complete information is mainly in Slovenian language, but some basic information is also in English. Data are mainly provided in .txt, .xls and .mht formats.

Both agencies provide data specific for their fields, however, there exists a network of weather stations along roads providing weather data specifically adapted for application in traffic. To inform both agencies about the research carried out in Roadidea project several contacts with their representatives have been established and a complete structure of work plan has presented. Several members from agencies have been invited to cooperate at preparation of traffic and weather data. This cooperation has made feasible non-parametric statistical modeling and forecasting of traffic flow rate on Slovenian highways as well as first joining of data from traffic and weather data bases.
3.2.9 Data logging from CAN-bus information

Although not yet described by one of the ROADIDEA partners the data logging from vehicle CAN-bus is to be born in mind for future innovative ideas within the project ROADIDEA.

During the year 2001 major truck manufacturers agreed to give third parties access to vehicle data used by Fleet Management Systems (FMS). They have designed a common interface as an open standard (www.fms-standard.com). Therefore, data logging from heavy vehicle CAN-bus is rather simple and straightforward. Furthermore, FMS-standard has enabled the development of advanced fleet management systems for haulage and logistics companies by utilising vehicle data also from a mixed fleet of vehicles. Fleet management systems utilise vehicle data e.g. for fuel consumption, driving style, and vehicle condition monitoring.

CAN-busses are utilised also in passenger cars and light vehicles. There is a standardised On-Board Diagnostics (OBD) connector in cars and light vehicles, which has been installed for the maintenance and inspection purposes only. It provides only a fraction of the data available in the CAN-bus. Passenger cars and light vehicles have CAN-bus connectors, but no standard has been agreed for a common interface and all manufacturers have their own places for the connectors and procedures. This is a great obstacle for the exploitation of the passenger and light vehicle data. So far the car manufacturers have not had a common interest to open the CAN-bus information to third parties as truck manufacturers have done. Although truck manufacturers have released the FMS-standard, a direct connection to the vehicles' internal CAN bus is unwanted by the manufacturers. The truck manufacturers have stated that they reserve right to withdraw any warranty on the product and shall not be subject to product liability arising from any direct CAN bus connection made by a third party (http://www.fms-standard.com/down_load/letter_acea.pdf).

Large scale utilisation of the vehicle data has numerable advantages and applications such as collection of road weather information with vehicles (Extended Floating Car Data). Utilisation of vehicle (CAN-bus) data has had a great interest in the research domain and also in the ICT-industry. For example, there are some products in the market that can log limited data set from CAN-bus of a certain passenger car brands. Overall the current situation in the vehicle data access is complicated. The standardised and secured access for third party developers to the vehicle data should be the target in order to generate new innovative services. ROADIDEA project could clarify the current restrictions and possibilities; and show the potential advantages and future services based on open interfaces.

3.3 Data Quality

Data quality issues have not been investigated very detailed during the initial data collection period described in this report. However, data quality has been discussed several times. It is indisputable, that data quality issues have to be considered while ROADIDEA is going to provide data for third party service providers and public or private end users. How can data quality been assessed? How do data quality issues effect service quality levels?
Several research projects are dealing with these questions. Public and private data and service providers are engaging itself to apply quality aspects to every part of the entire process chain from the installation of the sensor up to the provision of aggregated information to the end user. Until now, no standards or measurement instruments have been established to apply quality aspects standardized or to assess quality.

In Finland the Finnish Road Administration (FINNRA) has developed a quality management system for the road weather information. Road weather data quality depend on several factors including properties of the road weather measuring device in use, as well as placement, assembly, and maintenance of the device. The development started in 2002. The quality standard ISO 9001 was taken as a frame of reference for the work, as it is a generally accepted framework for quality. The quality system was developed by FINNRA, winter maintenance organizations, measuring equipment assembly and maintenance service providers, and the Finnish Meteorological Institute. In 2004 the quality management system was operational. The Finnish Road Administration ISO 9001 quality system for the road weather information production process is the first one in the world. [Iivanainen et.al, 2004]

Data quality becomes considered more deeply in the following tasks of the WP2. Task2.2 “New types of data needs” as well as task2.3 “Integration of data” includes data quality considerations.

3.4 Data Classification

Chapter 2.2.4 described the ROADIDEA data classification. This data classification has been introduced, discussed and applied to the ROADIDEA data collection work within the data workshop in Berlin. As a result of the workshop, all described data sets were reconsidered with respect to the new data classification. The data sets were assigned to at least one data class. The result of the assignment is shown in Figure 23. This figure has to be compared with the Chart 5 of the data collection results in chapter 3.1, where no data classes following the initial data classification of the DoW are empty.
The figure emphasizes data classes, which have to be reconsidered within the task 2.2 “New types of data needs”. These data classes are
- G. Traffic Management,
- H. Freight service and transport
but also the barely filled classes
- A. and C. Infrastructure as well as
- F. Traffic Monitoring.

Depending on the ideas to be created in the innovation process (WP5) and to be developed and piloted in the project ROADIDEA existing data have to be taken from the well filled classes or new data have to be ascertained for the needed data classes.
4. Conclusion/Summary

During the data collection period 46 data sets were described. The results chapter of this report shows different charts, which illustrate the content and the coverage of the collected data. In addition, the general availability of several data sources all over Europe has been investigated. Initial contacts with providers of these data sets have been established. Participating countries can be more or less covered with almost the same data sources. General data availability is guaranteed all over Europe. However, the data source identification has to be regarded as an ongoing process; the data collection has to be proceeded. The ROADIDEA data classification points out classes where further data collection efforts are necessary.

But also the starting innovation process (WP5) might cause new data needs, which lead automatically to new data collection efforts. The innovations must not be limited to available data. The innovations rather have to determine the necessary data. The data source identification has to ensure a provision with actual and high quality data. The innovation process will also clarify the thematic field where the development of ROADIDEA services will take place. The objective of the future work in WP2 must be the support of the services with data as well as the securing of data exchanges between data providers and service providers as users.

During the investigation of data availability in the participating European countries some partners got in contact with different obstacles and problems. These problems are rather bureaucratic than technical. To convince road authorities or other public and private data providers to collaborate in a research project a lot of work is remaining to do. The most sustainable way to find new collaborators and data providers is to show the success of innovative ROADIDEA services, which might be achieved with data the companies are providing. A win-win-situation for both the data provider and the service provider has to be obtained. The possible positive benefit of participating in a service process as data provider has to be emphasized. New business models should not regard only the profit of the service provider; it should also take the data provider as a relevant part of the process chain into consideration directly and visibly. If such business models succeed in the market, new data provider may be convinced to collaborate easily.

Thus, an offensive dissemination of project results is necessary. New services pilots, the underlying data and its quality assessment have to be published to a community, which includes desired data providers as well as successful service providers.

The generation of a data classification for the project ROADIDEA is to ease future data collection efforts. If a new road idea is developed and its data needs are specified the data classification is able to determine easily whether data are available or new data have to be ascertained. Beyond that, the data classification is to be used for data integration, which means the derivation of additional content by combining different data sources. This might be done by comparison and joining of data sources from different countries with similar weather conditions. Together with partner AMANOVA first ideas have been developed on this field. A detailed data classification and a clear assignment of all known data sets to this data classification is a prerequisite for this approach.

Data quality has to be considered thoroughly. Until now no unitary data quality assessment methods are known, which are able to assess quality of services based on the qual-
ity of underlying data. Corresponding methods are to be developed. To provide a high quality service data are needed which are both accurate and precise. Precision and accuracy are data attributes, which have to be quantified for data quality purposes. These attributes may depend on the quality of the sensor, the quality of the data acquisition procedure as well as on the algorithms of data aggregation. All these issues cannot be treated in the project ROADIDEA.

However, some criteria might be implemented which allow a derivation of service reliability measures, for instance from the available amount of sensors in comparison to the total amount of existing sensors.

With the data source identification of the work package 2 the way for the innovations for a new European road map is paved.
References

[RI_GA, 2007] ROADIDEA Grant Agreement – date 05/12/2007
Annex1 Data collection Questionnaire

ROADIDEA WP2 Data collection

Task 2.1: Data source identification - present data

Questionnaire

Mandatory fields are marked with an asterisk *.

You can also fill in this form online at: http://trafficresearch.dlr.de/roadidea/wp2/

<table>
<thead>
<tr>
<th>Contact information</th>
</tr>
</thead>
<tbody>
<tr>
<td>company (ROADIDEA-Partner) *</td>
</tr>
<tr>
<td>given name *</td>
</tr>
<tr>
<td>family name *</td>
</tr>
<tr>
<td>phone (incl. area code)</td>
</tr>
<tr>
<td>fax (incl. area code)</td>
</tr>
<tr>
<td>e-mail *</td>
</tr>
</tbody>
</table>

| Who owns the data set? *     |
| company (as indicated above) |
| third private party          |
| government                   |
| other                        |
| copyright                    |

| Data charges *               |
| free of charge              |
| charges required            |
| one-time delivery charges (€)|
| one-time usage charges (€)   |
| license fee (€)              |
| other (€)                    |
| unknown                      |

| Data availability *          |

57/64
<table>
<thead>
<tr>
<th>Can the data be used in the ROADIDEA project? *</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ yes  ☐ no, because [-]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data description * (classification see: ROADIDEA Description of Work, WP2, Task 2.1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ data from vehicles  ☐ data from traffic control systems</td>
</tr>
<tr>
<td>☐ data from infrastructure  ☐ weather monitoring</td>
</tr>
<tr>
<td>☐ other [-]</td>
</tr>
</tbody>
</table>

**abstract, objectives of data collection** *
The abstract serves as a brief general description of the data set. Please specify the purpose for which the data were collected.

<table>
<thead>
<tr>
<th>Spatial reference *</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Europe  ☐ country [-]</td>
</tr>
<tr>
<td>☐ region [-]</td>
</tr>
<tr>
<td>☐ city [-]</td>
</tr>
<tr>
<td>☐ other [-]</td>
</tr>
</tbody>
</table>

**coordinate system** * (e.g. WGS84, UTM, ETRF 89, ...)

<table>
<thead>
<tr>
<th>Frequency of data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ once  ☐ several times (regular)  ☐ several times (irregular)</td>
</tr>
<tr>
<td>☐ permanent (regular)  ☐ permanent (irregular)</td>
</tr>
<tr>
<td>☐ other [-]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Are the data accessible in an automated procedure? *</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ yes  ☐ no</td>
</tr>
<tr>
<td>☐ ftp (file download)  ☐ http (file download)</td>
</tr>
<tr>
<td>☐ database access  ☐ other [-]</td>
</tr>
</tbody>
</table>
### Data format
- [ ] ASCII
- [ ] XML
- [ ] other

### Data size
- [ ] < 1 MB
- [ ] 1 MB < 10 MB
- [ ] 10 MB < 100MB
- [ ] 100 MB < 1 GB
- [ ] > 1 GB
- [ ] per hour
- [ ] per day
- [ ] per month
- [ ] other

### Is a sample data set available? *
- [ ] yes
- [ ] no

If yes: Is the sample data set attached?  
- [ ] yes
- [ ] no

### Related information *
- [ ] codebook (in English)
- [ ] other documents

If yes: Is the codebook attached?  
- [ ] yes
- [ ] no

If yes: Are the other documents attached?  
- [ ] yes
- [ ] no

---

**Deadline: Friday 18.01.2008**

If you have any questions, please contact:

- **René Kelpin**  
  German Aerospace Center  
  Institute of Transport Research  
  Rutherfordstrasse 2  
  12489 Berlin-Adlershof  
  Telephone: +49 (30) 67055-217  
  Fax: +49 (30) 67055-8217  
  E-Mail: rene.kelpin@dlr.de  
  Internet: www.dlr.de/vf

- **Antje Parnitzke**  
  German Aerospace Center  
  Institute of Transport Research  
  Rutherfordstrasse 2  
  12489 Berlin-Adlershof  
  Telephone: +49 (30) 67055-178  
  Fax: +49 (30) 67055-283  
  E-Mail: antje.parnitzke@dlr.de  
  Internet: www.dlr.de/vf
Annex2 Data Workshop Agenda

Data Workshop WP2 & WP3

Berlin 10.-12.03.2008
Wissenschaftsforum am Gendarmenmarkt, Markgrafenstraße 37, 10117 Berlin

Agenda version: 07.03.2008

Participants:

- Pirkko Saarikivi, FORECA Finland
- Mikko Tarkiainen, VTT Finland
- Ville Könönen, VTT Finland
- Risto Öörni, VTT Finland
- Pertti Nurmi, FMI Finland
- Ilkka Juga, FMI Finland
- Marjo Hippi, FMI Finland
- Jussi Kiuru, Destia Finland
- Jörgen Bogren, Klimator Sweden
- Torbjörn Gustavsson, Klimator Sweden
- Lina Nordin, Klimator Sweden
- Poul Grashoff, Demis Netherlands
- Rene Kelpin, DLR Germany
- Antje Parnitzke, DLR Germany
- Angelika Schulz, DLR Germany
- Jörg Dubbert, Pöyry Germany
- Andrea Rossa, ARPAV Italy
- Franco Zardini, ARPAV Italy
- Janko Domokos, Road Safety Hungary
- Nebojša Subanovic, Meteo-Info Croatia
- Djurdica Markovic, Meteo-Info Croatia
- Pår Ekström, CARAN Sweden
- Olle Wedin, CARAN Sweden
- Igor Grabec, AMANOVA Slovenia
### Monday, March 10th

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>13:00-14:00</td>
<td>Get together</td>
<td></td>
</tr>
<tr>
<td>14:00-14:15</td>
<td>Introduction by Prof. Barbara Lenz, Head of the institute for transport research</td>
<td>DLR</td>
</tr>
<tr>
<td>14:15-14:30</td>
<td>The data collection survey: Results</td>
<td>Antje/DLR</td>
</tr>
</tbody>
</table>

**WP2: Introduction of the data in detail**

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30-14:50</td>
<td>Finish data</td>
<td>Mikko/VTT</td>
</tr>
<tr>
<td>14:50-15:10</td>
<td>Weather and other data from Finland</td>
<td>Marjo/FMI</td>
</tr>
<tr>
<td>15:10-15:30</td>
<td>More Finish data</td>
<td>Jussi/DEST</td>
</tr>
<tr>
<td>15:30-15:45</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>15:45-16:05</td>
<td>Swedish road weather data</td>
<td>Jörgen/KLIM</td>
</tr>
<tr>
<td>16:05-16:25</td>
<td>FEWS software and architecture</td>
<td>Poul/DEMI</td>
</tr>
<tr>
<td>16:25-16:45</td>
<td>FCD and more</td>
<td>Rene/DLR</td>
</tr>
<tr>
<td>16:45-17:05</td>
<td>Veneto data</td>
<td>Andrea/ARPV</td>
</tr>
<tr>
<td>17:05-17:25</td>
<td>Hungarian data</td>
<td>Domokos/RODS</td>
</tr>
<tr>
<td>17:25-17:35</td>
<td>Coffee Break</td>
<td></td>
</tr>
<tr>
<td>17:35-17:55</td>
<td>Croatian data</td>
<td>Nebojsa/METI</td>
</tr>
<tr>
<td>17:55-18:15</td>
<td>FCD – slippery road data</td>
<td>Par/CAR</td>
</tr>
<tr>
<td>18:15-18:30</td>
<td>Slovenian data</td>
<td>Igor/AMA</td>
</tr>
<tr>
<td>18:30-18:40</td>
<td>Summary and outlook</td>
<td>Rene/DLR</td>
</tr>
</tbody>
</table>

### Thursday, March 11th

**WP2: Data classification issues**

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00-09:10</td>
<td>Welcome and Outlook</td>
<td>Rene/DLR</td>
</tr>
<tr>
<td>09:10-09:40</td>
<td>The VTT data – a classification approach</td>
<td>Mikko/VTT</td>
</tr>
<tr>
<td>09:40-10:10</td>
<td>German data sources and its availability</td>
<td>Joerg/POY</td>
</tr>
<tr>
<td>10:10-10:40</td>
<td>The Clearingshouse for traffic data</td>
<td>Angelika/DLR</td>
</tr>
<tr>
<td>Time</td>
<td>Event</td>
<td>Organizer(s)</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>10:40-11:00</td>
<td>Does it fit to all European countries – The Croatian Example</td>
<td>Nebojsa/METI</td>
</tr>
<tr>
<td></td>
<td>Possibilities &amp; obstacles - discussion</td>
<td>Rene/DLR</td>
</tr>
<tr>
<td>11:00-11:30</td>
<td>Coffee break with discussion</td>
<td>Rene/DLR</td>
</tr>
<tr>
<td>11:30-12:30</td>
<td>Further discussion, results and list of actions:</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>• Next data collection steps</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• New sources</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• ...</td>
<td></td>
</tr>
<tr>
<td>12:30-13:00</td>
<td>Partner task allocations and the WP2 schedule</td>
<td>Rene/DLR</td>
</tr>
<tr>
<td>13:00-14:00</td>
<td>Lunch break at Hotel Hilton</td>
<td></td>
</tr>
<tr>
<td>14:00-14:30</td>
<td>Project news and dates</td>
<td>Pirkko/FORC</td>
</tr>
</tbody>
</table>

**WP3: Methods and models**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Organizer(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>14:30-16:00</td>
<td>Start of workshop WP3 with presentations from each participants in WP3</td>
<td>Jörgen/KLIM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
</tr>
<tr>
<td>16:00-16:15</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>16:15-17:45</td>
<td>Work and discussions in groups according to interests and task allocation</td>
<td>Jörgen/KLIM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
</tr>
<tr>
<td>18:00-19:00</td>
<td>Boattrip or Sightseeing-Tour</td>
<td></td>
</tr>
<tr>
<td>19:30</td>
<td>Dinner at &quot;Mutter Hoppe&quot;</td>
<td></td>
</tr>
</tbody>
</table>

**Wednesday, March 12th**

<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Organizer(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00-10:30</td>
<td>Group discussions and presentations of the group work</td>
<td>Jörgen/KLIM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>all</td>
</tr>
<tr>
<td>10:30-10:45</td>
<td>Coffee break</td>
<td></td>
</tr>
<tr>
<td>10:45-12:30</td>
<td>WP3 Results from workshop, conclusion and list of actions</td>
<td>Jörgen/KLIM</td>
</tr>
<tr>
<td>12:30-12:45</td>
<td>Conclusion and closing of the meeting</td>
<td>Rene/DLR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pirkko/FORC</td>
</tr>
<tr>
<td>12:45-</td>
<td>Light lunch and Farewell</td>
<td></td>
</tr>
</tbody>
</table>
Annex3 Data Set Table

The data set table can be found as WP2_task2.1_data_set_table.xls at the ROADIDEA web page: http://www.roadidea.eu/documents/default.aspx
### Annex4 Data Availability Overview Table

<table>
<thead>
<tr>
<th>Country</th>
<th>Data availability from/about Road networks</th>
<th>Weather</th>
<th>Other Data</th>
</tr>
</thead>
</table>
| Finland     | - 350 traffic counting stations (loop detectors)  
- Urban traveltime information in Helsinki  
- FCD (Tampere and Oulu, Helsinki planned) | - Finnish road authority: 330 road weather stations  
- FMI: 150 automatic weather observation stations | - Digiroad (infrastructural street and road database)  
- Accident statistics  
- Video traffic monitoring in Helsinki  
- Public transport register |
| Germany     | - Police TIC in every federal state  
- TIS of some federal road administrations  
- Permanent counting points  
- FCD (car and phone) in different cities | - SWIW road weather monitoring systems | - traffic-adaptive traffic control systems  
- truck tolling systems  
- parking guidance systems in large cities  
- public transport  
- VIKING ferry data |
| Sweden      | - Swedish national road database  
- FCD with Can-bus data trials | - 720 field stations  
- Airport weather observations (available all over Europe) | - accident database STRADA |
| Netherlands | - loop detectors  
- GPRS stations  
- Astrid MTM2 stations | - historical, actual and forecast weather | - road, bicycle path and footpath network data |
| Italy       | - 5 private motorway concessionaires | - Veneto region data  
- Boundary layer profilers  
- Meteorological satellite data  
- Visibilimeter network region Veneto | - traffic webcams |
| Hungary     | - 3 different traffic control systems in the motorway network  
- 300 permanent counting stations  
- FCD trial | - 25 local weather observation stations | - Traffic webcams  
- Motorway accident data  
- Technical road parameter database OKA  
- Data from public transport control stations |
| Croatia     | - 5 motorway concessionaires with unknown Data availability | - unknown amount of road weather stations  
- access unknown | |
| Slovenia    | - 2 road traffic agencies with free data availability | - network of road weather stations | |