Large Scale Collaborative Project

7th Framework Programme

INFSO-ICT 224067

Tools for Database Handling

Deliverable n.	D4.1.1	Tools for Database Handling	
Sub Project	SP 4	Evaluation and assessment	
Workpackage	WP 4.1	Database usability and accessibility	
Task n.	T 4.1.1	Pilot data retrieval and analysis	
Author(s)	Koskinen, S., Kovacs, A., Dreher, S., Welsh, R.	File name	TeleFOT_4.1.1_ToolsForDatabaseHandling_v1.0.doc
Status	Final		
Distribution	Public (PU)		
Issue date	2010/08/16	Creation date	2010/04/14
Project start and duration	1st of June, 2008 – 48 months		

European Commission
Information Society and Media

Project co-funded by the European Commission

DG-Information Society and Media

in the 7th Framework Programme

SEVENTH FRAMEWORK PROGRAMME

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBREVIATION	DESCRIPTION
ADAS	Advanced Driver Assistance Systems
CAA	Cockpit Activity Assessment Module
CAN	Controller Area Network
DAS	Data Acquisition System
ECA	Environmental Conditions Assessment Module
FMS	Fleet Management System (CAN interface for trucks and buses)
FOT	Field Operational Test
FTP	File Transfer Protocol
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
ICT	Information and Communication Technology
IMU	Inertial Measurement Unit
HTTP	Hypertext Transfer Protocol
LFOT	Large Scale FOT
OBD-II	On-Board Diagnostics (interface)
SQL	Structured Query Language

ABBREVIATION	DESCRIPTION
SYNOP	Surface synoptic observations (message)

REVISION CHART AND HISTORY LOG

REV	DATE	AUTHOR	REASON
0.1	2010-04-27	Sami Koskinen	First full draft for partners
0.2	2010-06-15	Stephane Dreher	Added Map Matching section
0.3	2010-07-28	Katia Pagle, Christina Kotsiourou	Peer reviewed version
1.0	2010-08-16	Sami Koskinen	Corrections after peer review: minor clarifications to the text, first figures better explained and status of the work added. Submitted to the EC.

EXECUTIVE SUMMARY

This deliverable D4.1.1 gives an overview of database tools used in the TeleFOT project. The test sites collect a large amount of vehicle logger, service-related and questionnaire data. This information is transferred to a central database for further processing and analysis. The data is enriched with tools such as map matching to supplement GPS coordinates with road attributes.

The database tools used in the project fall in the following categories:

- Database solutions tied to tested services or logger data collection. These tools may also provide data quality checks and driver feedback/reporting e.g. via a website.
- Project tools for extracting driving performance indicators from database and monitoring quality. These programs are run periodically in the central server.
- Project tools for data enrichment, mainly map matching.
- Generic commercial and open source database viewers and administration tools.
- Data access and analysis with mathematical and statistical tools such as SPSS and MATLAB.

The objectives of this work in WP4.1 are to complete a system for FOT data handling and to support analysts (other WPs in SP4) with database tools and scripts. The starting point of the work are the WP2.2 (Methods and tools) and WP2.3 (Data specification), especially T2.3.1 (Data acquisition) and T2.3.3 (Database structure), The database structure and tools are currently being tested in FOT pilots, preceding full-scale FOTs. The development continues as feedback is received from the pilots and analysts.

1. INTRODUCTION

TeleFOT is a Large Scale Collaborative Project under the Seventh Framework Programme, co-funded by the European Commission DG Information Society and Media within the strategic objective "ICT for Cooperative Systems".

Officially started on June 1st 2008, TeleFOT aims to test the impacts of driver support functions on the driving task with large fleets of test drivers in real-life driving conditions.

In particular, TeleFOT assesses via Field operational Tests the impacts of functions provided by aftermarket and nomadic devices, including future interactive traffic services that will become part of driving environment systems within the next five years.

Field Operational Tests developed in TeleFOT aim at a comprehensive assessment of the efficiency, quality, robustness and user friendliness of in-vehicle systems, such as ICT, for smarter, safer and cleaner driving.

Objectives

The objectives of this work described in this deliverable are to complete a system for FOT data handling and to support analysts with database tools and scripts. The tools should provide a basis for coherent evaluation of nomadic devices and services tested in the project.

The database tools include automated calculation of driving performance indicators, data quality monitoring and data enrichment. New calculation algorithms developed in the project can easily be added to these existing tools.

Scope and structure of the deliverable

This deliverable first explains the data collection during the project and in Chapter 2, it introduces the structure of the central database where all large scale FOT (LFOT) data is collected.

Further, Chapter 3 presents project partner and 3rd party database tools/services that are used for data collection. These services often provide vendor-specific indicators and e.g. web sites for reporting details about vehicle use for the drivers.

The project tools for extraction of harmonized driving performance indicators are presented in Chapter 4. Map matching and similar data enrichment tools are described in Chapter 5.

Data viewers are summarized only briefly in Chapter 6 as the partners are able to use various graphical tools to access the database.

The user guide for the tools will be described in deliverable D4.1.2, which is restricted to the project partners.

The tools described in this deliverable will be used during the pilot trials, which precede the real Field Operating Tests. The tools will be improved based on the results of the pilots and the analysis of the pilot results. The feedback and the improvements will be discussed in D4.1.3/D4.1.4.

Data collection overview

The TeleFOT project aims to collect all LFOT data to a central server. This is to harmonize data formats and analysis across test sites. There are several sources of data:

- Cost-effective (€100—500) in-vehicle loggers recording mostly GPS data. Some of these loggers can also record data from vehicle communication network, using CAN, FMS or OBD-II interfaces.
- FOTs have also chosen to modify navigation software to log GPS and usage data. In this case, vehicle-installed loggers may not be needed.
- Some of the tested services log usage data on their Internet servers and these logs are collected.
- Questionnaire results, travel diaries and supplementary weather and traffic data are also collected to the central database.

Interfaces and data transfer practises have been created to be able to conveniently upload data from mobile devices to the central server. For example a basic HTTP Post

method can be used to send a data sample from a logger directly to the central server. In most FOTs, there's an intermediate storage, a local server or hard drive or similar memory storage. In these cases, the data collection to the central server is performed periodically. Communication between servers can be as simple as using FTP to send large log files in project-specified text format. Also database dump/backup files (.sql) can be used.

The data is processed at central server using scripts that parse and save log files in harmonized format to the central database. As a next step, the database content is analysed and data is enriched automatically.

The following figure gives an overview of the data collection system architecture, where data from different sources is collected to the central server.

Data is gathered from on-board loggers (here: DAS, Data Acquisition System), usage logs and questionnaires. The TeleFOT project uses DASs from different providers which log basic motion using satellite positioning and acceleration sensors. The DAS collects GPS coordinates, velocity, heading and number of satellites used in the solution. For acceleration data, some DASs include an acceleration sensor and defined thresholds/events for when to start recording. More information on the Data Acquisition Systems is provided in D3.2.2a.

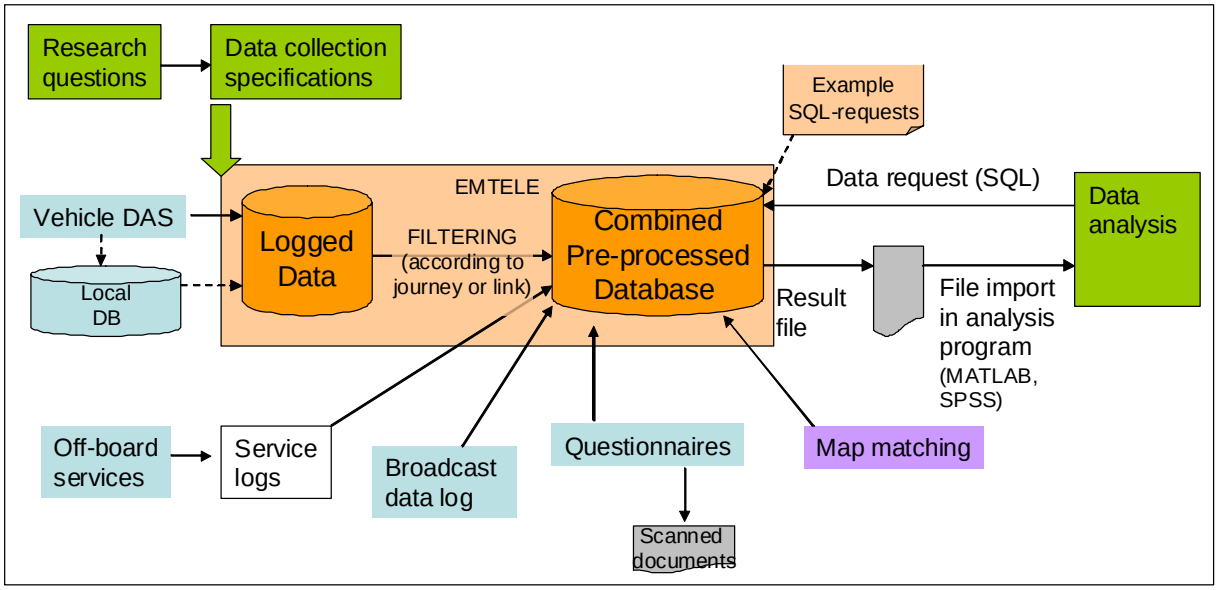


Figure 1. Large scale FOT data collection system architecture [1].

The “Combined pre-processed database” contains processed DAS data and combines the DAS data with other information sources, such as service logs and user questionnaires. TeleFOT uses LimeSurvey open source survey tool to run online questionnaires and to be able to save result data directly to a database. Service log is information stored by the services to be tested. The data collected with the DASs are processed and filtered in order to calculate a number of performance indicators, such as mean speed or journey travel time, and making the data available to the analysts. This pre-processing includes e.g. separating the location data per road segment and per journey, and calculating the relevant indicators (e.g. average speed over road segment), as well as assuring the data quality of the collected data.

During this processing phase, the position data will be (map) matched to the road network at the central server using a service provided by the map provider NAVTEQ, a partner in the TeleFOT project, which is described in this deliverable.

The detailed FOT (DFOT) data collection is not discussed in this deliverable as it currently is very vehicle specific and saved data is analysed mostly at local test sites. The central database also does not contain video data, while most DFOT trials include video files.

The central database can, however, be used to save post-processed video analysis and DFOT analysis results. This is the case especially with project tools CAA (Cockpit Activity Assessment Module) and ECA (Environmental Conditions Assessment Module) that process video and sensor data automatically to output driver vigilance and safety indicators such as ‘percentage of eyes off the road’ and ‘time to collision’.

Overview of driving performance indicators

For each studied impact area, the project has compiled key hypotheses on how nomadic devices and services affect driving [2, 3]. Various driving performance indicators are known to be linked with these hypotheses. FESTA project has compiled a rather comprehensive list of traditional performance indicators [4]. TeleFOT has analysed the potential to measure these indicators using nomadic device loggers, GPS loggers, project-specific tools such as a camera tool for measuring drivers’ eye movement and also questionnaires. Many of the indicators, such as vehicle’s position on a lane or headway, are best suitable for measurements with heavily instrumented vehicles and

studying ADAS (Advanced Driver Assistance Systems). Therefore, the project will also calculate statistical indicators such as ‘percentage of driving on motorway’ and ‘RPM distribution when driving 50-60 km/h’, to complement analysis. These indicators are widely used in fleet management systems and can be calculated from GPS data.

The following table gives an example of how performance indicators are linked with hypotheses.

Table 1. Example performance indicator for Speed Alert.

Research question	Function	Hypothesis	Performance indicator	Related parameters	Calculation example
Is speed affected?	Speed Alert / Speed Limit Information	There is likely to be a decrease in speed because the driver is warned about exceeding the speed limit	Proportion of time spent exceeding the speed limit	Speed limit accuracy in digital maps	(number of samples with speed > speed_limit) / (total number of samples)

2. LARGE SCALE FOT DATABASE STRUCTURES

The database structures in the project still expand with data coming from national FOTs, but the basic structure is in place, linking FOTs, users and vehicles to raw logger data. The current draft database structure, suggested by Emtele and VTT, is shown in Figure 2. The structure is based on the results of the work performed in WP2.2 and WP2.3. It can be used to record data from some of the first FOT pilots in Finland, Greece, UK and Spain. More data and other pilots are being imported.

Figure 2 shows the different database tables (e.g. Logger) along with their primary keys (logger_id) and other columns (country_id, vehicle_id, IMEI). The primary keys are used to identify each row and to link records to other tables.

The database structure allows for flexibility in FOT configuration: users can change loggers or even switch to another FOT. This is made possible by logging possession and participation time data. Generally the links between data types are relatively complex in this type of FOT data and this challenges the analysts. Not all analysts are experienced with SQL language and complex database structures. Therefore this document also gives short example queries on how to select pieces of data for further analysis. A tool for extracting driving performance indicators automatically is presented, making easier the use of the database.

The Limesurvey questionnaires are also stored on the central server, linked via questionnaire_id to periods of driving. The database contains traffic and weather data as well. These will be discussed in Chapter 5, Tools for data enrichment.

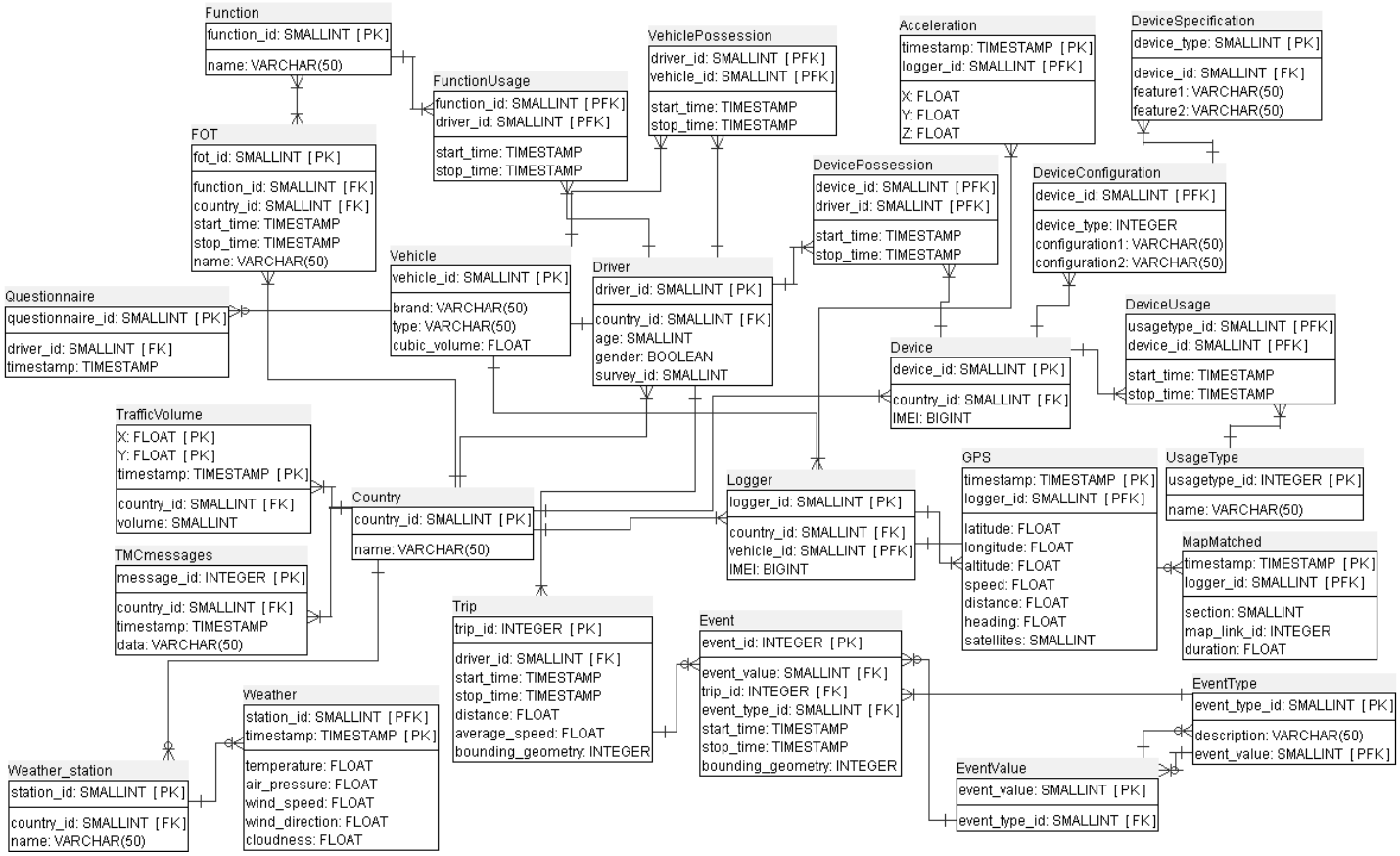


Figure 2. Draft database structure by Emtele and VTT.

3. TOOLS INTEGRATED WITH DATA COLLECTION

The project's central server

Emtele operates the TeleFOT project central server, where all large-scale FOT data is collected. The server platform is based on Comptel EventLink product, which is preferred by many telecom operators. EventLink ensures a robust and widely scalable basis for data collection. A more detailed description of the platform can be found from TeleFOT deliverable D3.7.2 Data and User Management.

Emtele has implemented FTP and SFTP collector "nodes" to the platform for enabling test site data to be transferred over FTP. Similar collector nodes have been implemented also for various GPS loggers. As a next step in the data collection, various sanity checks are performed for the data fields and warnings are triggered for the operator, if faulty data is detected.

The system also includes a FTP distributor node for sending files to a remote location. The database can be exported and copied also using traditional methods, i.e. SQL backup files.

BroadBit's database tools

BroadBit's data loggers capture the following types of vehicle data:

- Position and time-stamp data of periodic way-points.
- Detailed acceleration log of +/- 5 seconds around traffic incidents. This data means 3-axis acceleration values sampled at 50 Hz, and the incident's geographic position and time-stamp.
- Daily statistics of the vehicle utilization.

Vehicle utilization data is first transferred to BroadBit's database where it becomes the basis for automated statistical analysis and incident lists. For those vehicles, which are equipped with detailed field trial equipment, such as drive camera recording device, more detailed information can be manually obtained about incidents of interest.

As the following figure illustrates, the logged vehicle data is reported from the data loggers through a local reporting client application. Logged data can be accessed at any point within this data collection architecture; through local Bluetooth connection, through a receiving HTTP server or through an SQL database management application. Therefore the integration process can be adjusted to specific test site needs.

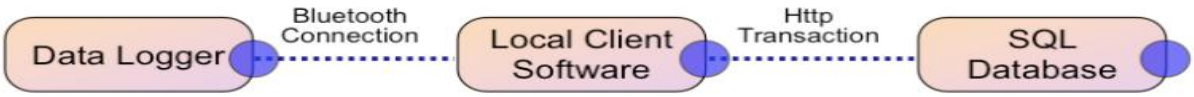


Figure 3. BroadBit's data reporting steps.

The structure of the reported log data has been defined during the project. There are separate data tables for the above-listed types of log data, for the reporting events, and for the list of field trial vehicles. This data of the vehicle utilization and detected incidents can be easily managed over a web-based SQL administrator interface such as phpMyAdmin, which is illustrated in the following figure:

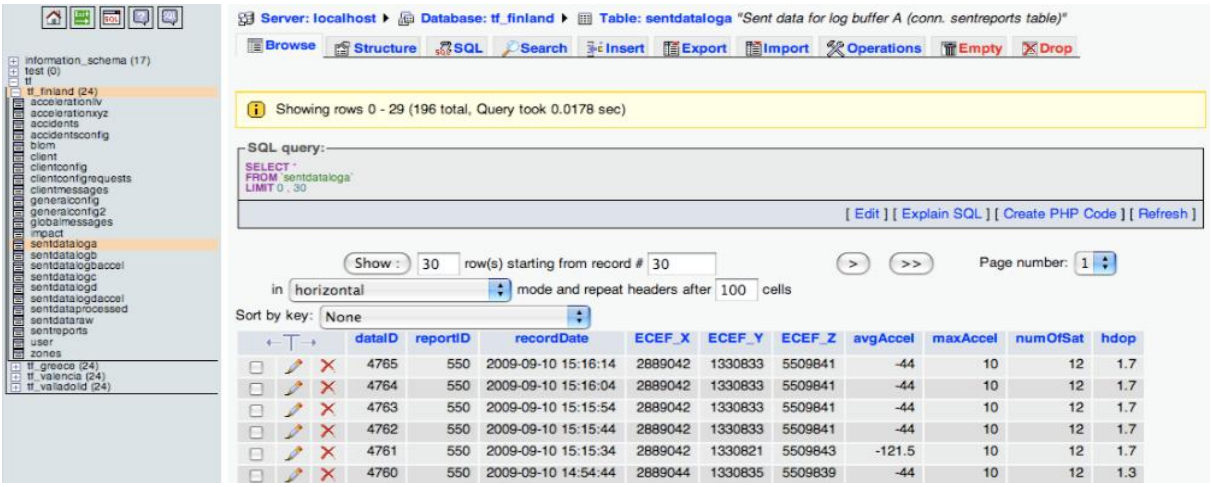
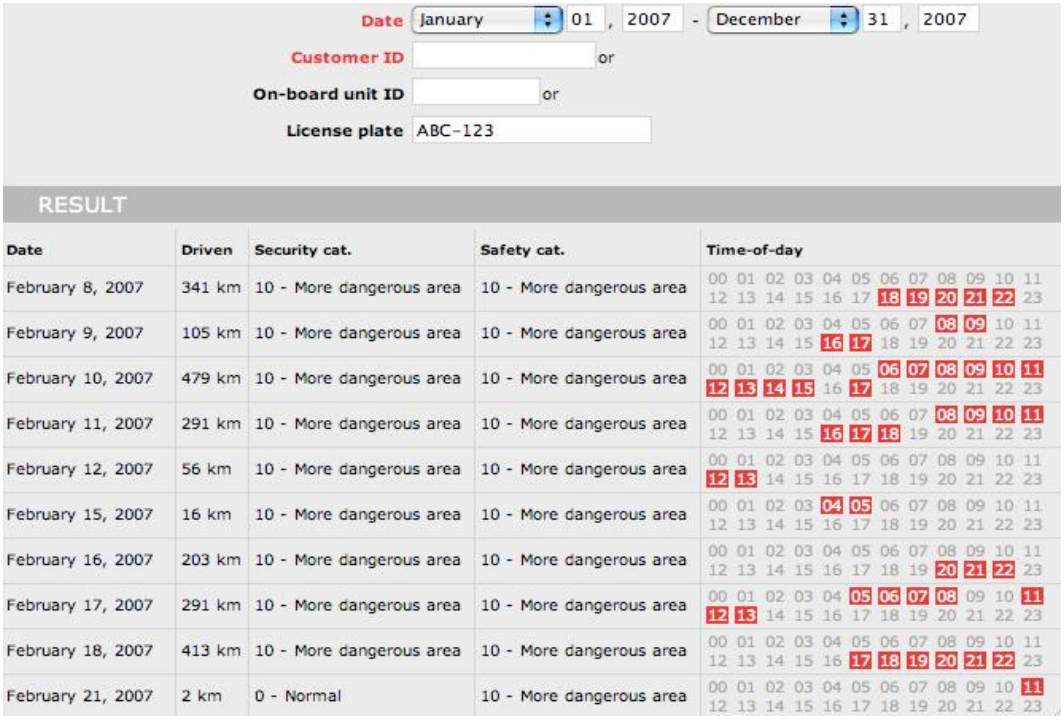


Figure 4. Screenshot of phpMyAdmin open source web-based SQL interface.

Additionally, BroadBit provides a web-based tool for the query of reported vehicle utilization statistics and incidents. A daily breakdown or yearly summary can be produced. This display can be adjusted to show statistical data such as:

- Total driven distance
- The hours of the day when the vehicle was used
- Driven distance by road category (urban/rural) or by time category (daytime/night-time)
- Value and location of top speed
- Starting and ending locations
- Average acceleration values



Date January 01 , 2007 - December 31 , 2007
Customer ID or
On-board unit ID or
License plate ABC-123

RESULT

Date	Driven	Security cat.	Safety cat.	Time-of-day
February 8, 2007	341 km	10 - More dangerous area	10 - More dangerous area	00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 **18** **19** **20** **21** **22** 23
February 9, 2007	105 km	10 - More dangerous area	10 - More dangerous area	00 01 02 03 04 05 06 07 **08** **09** 10 11 12 13 14 15 **16** **17** 18 19 20 21 22 23
February 10, 2007	479 km	10 - More dangerous area	10 - More dangerous area	00 01 02 03 04 05 **06** **07** **08** **09** **10** **11** **12** **13** **14** **15** 16 **17** 18 19 20 21 22 23
February 11, 2007	291 km	10 - More dangerous area	10 - More dangerous area	00 01 02 03 04 05 06 07 **08** **09** **10** **11** 12 13 14 15 **16** **17** **18** 19 20 21 22 23
February 12, 2007	56 km	10 - More dangerous area	10 - More dangerous area	00 01 02 03 04 05 06 07 08 09 10 11 **12** **13** 14 15 16 17 18 19 20 21 22 23
February 15, 2007	16 km	10 - More dangerous area	10 - More dangerous area	00 01 02 03 **04** **05** 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23
February 16, 2007	203 km	10 - More dangerous area	10 - More dangerous area	00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 **20** **21** **22** 23
February 17, 2007	291 km	10 - More dangerous area	10 - More dangerous area	00 01 02 03 04 **05** **06** **07** **08** 09 10 **11** **12** **13** 14 15 16 17 18 19 20 21 22 23
February 18, 2007	413 km	10 - More dangerous area	10 - More dangerous area	00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 **17** **18** **19** **20** **21** **22** 23
February 21, 2007	2 km	0 - Normal	10 - More dangerous area	00 01 02 03 04 05 06 07 08 09 10 **11** 12 13 14 15 16 17 18 19 20 21 22 23

Figure 5. BroadBit's statistics viewer.

The list of detected incidents can be displayed by date and/or by vehicle identity. A graphical output of the incident data is shown on a map, as illustrated on the figure below. Each car icon represents the vehicle's position around the incident, at 1 second snapshots. Clicking on any car icon brings up the vehicle dynamics data:

- Time-stamp
- Position co-ordinate values
- Velocity value and direction (blue arrow)
- Longitudinal and lateral acceleration charts of +/- 1 second around the snapshot
- Direction of the acceleration impact (red arrow).

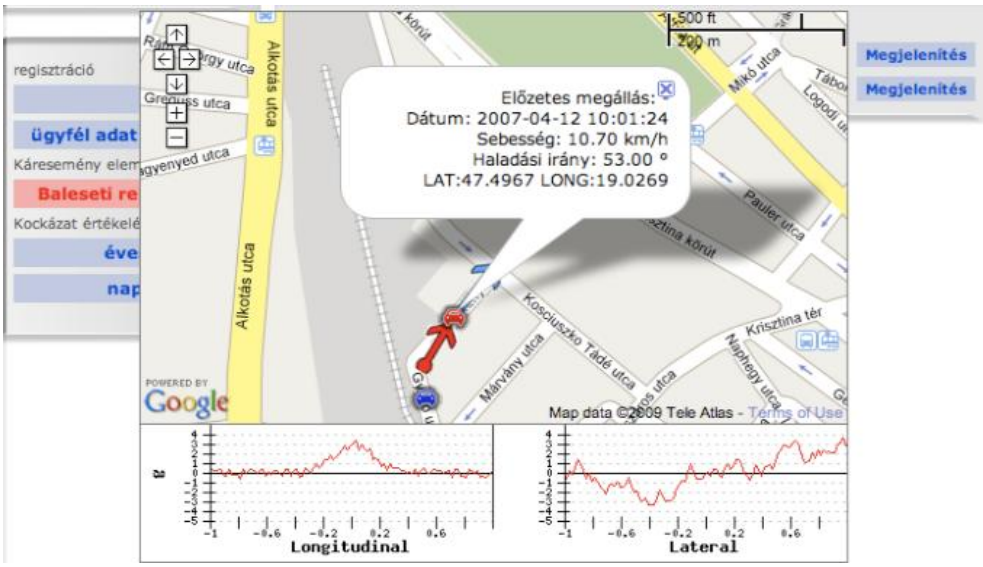


Figure 6. BroadBit's incident viewer.

Based on this identified incident data, incidents of interest can be manually further analyzed on those vehicles, which are equipped also with a drive recording camera.

After fetching the (optional) drive recording camera's SD card, this manual analysis can be performed through a provided PC-based tool, which is illustrated on the following figure. The data in the drive camera recording unit can be queried by the GPS-derived time, so that the recording of the incident of interest can be found by the matching of time-stamp.

The BlackBox analysis tool plays back the video and audio recording around the incident time, so that the circumstances and type of the incident can be analysed in detail.

Figure 7. BroadBit's drive recording camera data viewer.

Logica & Mediamobile LATIS

Logica operates in collaboration with Mediamobile[1] a commercial traffic and weather information service called LATIS. In the TeleFOT project, their platforms and mobile phone application are tailored to also record user GPS data along with usage statistics. Map matching is performed on the server as the service also includes features such as speed alert, where road segment speed limit information is used. The data collected to LATIS database is transferred automatically to Emtele's central server. The quality checks and user administration are mostly performed with LATIS.

[1] Mediamobile bought Destia Traffic in June 2010.

LimeSurvey

The TeleFOT project is using LimeSurvey (www.limesurvey.org), an open source survey tool for online questionnaires. Also paper questionnaires are formatted using LimeSurvey tools and printed out from the system. This approach ensures that all data can be conveniently saved to the central/local databases and it can be linked with collected logger data. Theoretically this easily configurable survey tool would even enable using logger data summaries when creating questionnaires: this would mean that the user would be asked about his logged driving. This is an optional possibility that the local test sites may use.

The LimeSurvey administration tools enable easy creation of user surveys without programming knowledge. The software is controlled via a web interface. The output data can also be exported in various formats (text, Excel etc), if the data is to be analysed using separate tools and not accessed directly from the central server.

4. EXTRACTION OF PERFORMANCE INDICATORS

Driving performance indicators are generated and calculated at least in three ways in the project:

1. FOT-specific indicators that are directly available through tested services or 3rd party systems. When used for analysis, the calculation methods of these indicators have to be well known in order to match them with the set of project indicators. Even 'average speed' may mean different things in different systems.
2. Indicators post-processed using project tools and scripts running at the central database server. These tools calculate a documented set of indicators and simultaneously sort and monitor data.
3. Analysts study and process their advanced indicators manually and using personal tools in connection with the central database and local data.

The first category of indicators includes 3rd party summaries of e.g. total kilometres driven and fuel consumption. These won't be covered in this document as the project's aim is to harmonize and recalculate these indicators with project tools. If these indicators are, however, well documented and don't exist in project's list of indicators, they may be used to support analysis.

For the second category of generated indicators, VTT has during the project created a server-side Java software for calculating driving performance indicators from data collected at the central database. This software is to be run periodically (e.g. daily). The process is automated and only database passwords need be given in a configuration file. With each execution round, it searches for new logger data from the database. If new data is found and it already contains complete journeys, decided mainly by time difference from previous driving, the software extracts these new journeys and calculates tens of indicators for each. New algorithms producing more indicators are added to the software as the project continues.

As a result, the software provides summary tables for collected data and saves them to the central database, making it easier for analysts to continue with their advanced

indicator analysis with raw data, and on the other hand, to harmonize calculations across test sites. The summary data should already provide support for answering several research hypotheses.

The summary table consists of detected journeys and their parameters. For each journey e.g. the following summary data (indicators) are saved:

Table 2. Example summary data per journey calculated from FMS data

leg_id	Assigned ID for a leg
logger_id	Data logger ID
timestamp_start	Time for leg start
timestamp_stop	Time for leg end
total_distance_driven	Total distance driven
odometer_start	Vehicle odometer reading (FMS)
odometer_end	Vehicle odometer reading (FMS)
avg_velocity	Average velocity
total_fuel_consumption	Total fuel consumption (FMS)
time_speed_0_when_engine_on	Time stopped when engine on
variance_of_acceleration_when_driving	Variance of acceleration when speed > 0
number_acceleration_over_5	Times of “hard” acceleration/deceleration
time_enginetemp_under61	Time running with “cold” engine
time_rpm_between1200_1400	Time for RPM being inside a class (several)
avg_rpm_withvelocity_between60_70	Avg. RPM when driving certain speed
time_coasting	Time spent coasting
time_cruise_control	Time when cruise control is activated
std_velocity	Standard deviation for velocity
median_velocity	Median for velocity
nr_brakings	Number of brakings
p_rush_hours	Percentage driven during rush hours
…	…

This example set is produced by FMS (truck CAN interface) data processing modules. Similar summary tables are calculated for GPS data and OBD-II (On-Board Diagnostics

interface) and then these calculations are combined as a final step. The first steps of processing are partly logger-specific if the logger data structures have required changes in the generic data structures in the central database. This is the case mainly for other than GPS data (e.g. logger events such as ignition on/off).

After the journeys have been identified and saved to a new table, they are classified based on their type and origin/destination. Origins and destinations basically form groups of *x, y* coordinates. Common locations include e.g. home, work, summer cottage and regular shopping places. These coordinate pairs are classified using partly traditional algorithms and partly project-specific methods that are still being further researched. The goal is to classify the different journeys so that meaningful comparisons can be made between the same types of journeys. The classifications should, when possible, also match the purposes and classifications used in the project's travelling diaries filled periodically by the drivers.

The software is programmed using Eclipse and Java language. It can run both in Windows and Linux servers of the project and locally in analysts' computers. Currently it supports MySQL and PostgreSQL databases. The software structure and modularity makes it reasonably easy for developers to add more indicators to calculation. However, it doesn't provide any graphical information, so results graphs and visual analysis must be made with other tools such as MATLAB or Excel.

The following figure shows the basic output of this console application (development version).

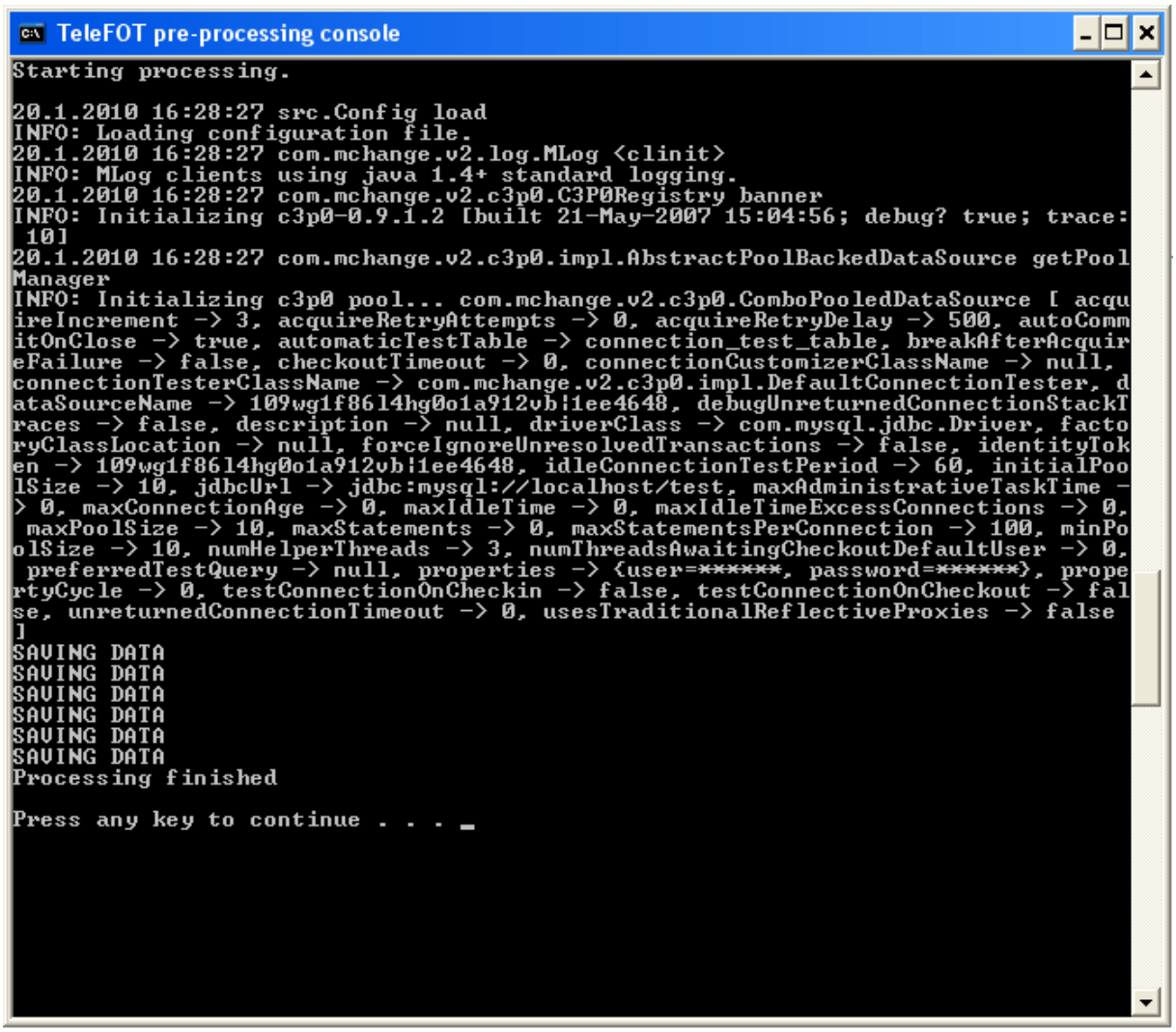


Figure 8. TeleFOT console application (development version) screenshot by VTT.

The software also keeps a separate error log in case logger data doesn't meet expected value ranges or contains a lot of logger-specific error codes. This log can be reviewed by the database administrators (e.g. weekly) to detect if certain loggers raise a lot of errors. Alternatively the errors will be instantly visible for the analysts running the software. No data is lost and in the case of a software bug, the analysis could be easily run again.

As another "quality" check, the software fills the indicator table with codes representing unavailability or calculation errors (e.g. -1, -1000), when an indicator cannot be calculated for a reason. These values can be easily handled / excluded in result prints.

5. TOOLS FOR DATA ENRICHMENT

The TeleFOT project uses server-side tools and accesses external information services to supplement in-vehicle logger data. The main method of data enrichment is map matching, with a tool provided by NAVTEQ. Other information sources are national traffic information services and weather data. The data enrichment also refers to Emtele's work in the project, where the central database structure is created so that collected information can be easily linked.

NAVTEQ's map matching

The 'map matching' process consists of projecting an estimate of the geographical position of a vehicle on a representation of the road network stored in a map database. This process generally involves the use of GPS and dead reckoning sensors like gyroscopes, odometers and compasses.

The representation of the road network in the database constitutes a key element of the map matching process. A digital road map for in-vehicle applications consists of geometry and related attributes. The core geometry provides a connected node and link representation of the road network in which roads are represented by their centreline and are connected together by nodes with a position. The shape of a link, if it is not a straight line, may be represented by one or more shape points. Attributes are referenced to links, nodes and shape points. Typical attributes associated to links are access restrictions, speed limits or road type and category. Up to 300 attributes are available in the NAVTEQ database.

Inertial sensors, such as an Inertial Measurement Unit (IMU), or the more simple combination of the vehicle odometer output and a gyroscope, provide relative positioning, which is fused with GNSS (Global Navigation Satellite System) positioning to provide an estimated absolute position, which in turn is fused with map information in the map matching process to generate an estimated position on the centreline of a road

in the map. Once the most probably matched position, consisting of a link identification number and a position along this link (called “offset”) measured from its start point, is found, the values of attributes associated to it can be extracted. The characteristics of the link traversed by the vehicles can consequently be known and used in the analysis of Field Operational Test data.

The Java map matching libraries developed within TeleFOT are integrated in a Java software (similar to the indicator calculation software presented earlier) and provided for demonstration purposes of the library usage. This separate map matching tool is intended to be run at the central server and basically requires only the NAVTEQ map database to run. The implementation is based on the MMF (Mobile Map Format) binary map format.

The Map Matching Algorithm works as follows:

- The match probability of a position point on a road link is a linear combination of weighted distances :
    - Distance to the GPS coordinate.
    - The heading difference from the GPS and the link section is turned into a “distance”, if the vehicle is driving faster than 3 km/h.
- Latitude/longitude GPS drifts are corrected by exponentially smoothed application of previous corrections.
- Links adjacent to previously matched links get higher probability. This ensures that the vehicle stays positioned on a road.
- Ghost driving is prohibited and results in a matching with zero probability.
- The map matcher can use following data: latitude, longitude, heading, current route path, road name/number.

The map matching demonstration application works with data files. Using the library, different implementations based on the example program can however be realized in order to work for example directly on a relational database.

Following data is provided as input to the demonstration map matcher:

- Latitude
- Longitude

- Heading

For each input file, a new session is started and a separate output file is written. The extracted attributes are added in additional columns at the end of each line in the data file.

The following attributes are extracted by the Map Matching Tool. This list of attributes has been defined based on the requirements provided by TeleFOT and euroFOT partners (euroFOT has been in discussions with NAVTEQ about using the same tool):

- Matched link id
- Direction of travel along the link
- Indicator if the vehicle entered a new link
- WGS84 map-matched coordinates
- Map matched heading against North
- Distance between input and map-matched position
- Legal speed limit on the link in km/h
- Length of link in meters
- Road type (functional calls and type)
- Urban flag
- Official road number
- Official road name
- Number of lanes in driving direction
- Allowed driving directions
- Allowed vehicle types
- Tunnel flag
- Bridge flag
- Average driving speed on link in km/h
- Form of way
- Functional road class
- The angle of turn when entering a new link at an intersection
- Traffic light flag
- Stop sign
- Priority sign
- Junction type (ramp or intersection)
- Overtaking restrictions
- Accident spots
- TMC code
- Variable speed limit
- Truck restrictions

The map matching library is currently being tested in FOT pilots and new software versions can be used to re-run the data if necessary.

Integration of environmental and traffic data

The project attempts to acquire weather and traffic data from the test sites. This information can be used as metadata for interpreting the logger measurements.

At certain test sites, such as in the Finnish FOT, the partners have access to national information services which provide traffic and weather measurements at road segment level. Depending on the test site setup, this type of information can be collected to the project's central database during the trials. This requires agreements on the data interfaces, transmission periods and later on use of the data.

Some local weather conditions like heavy rainfall, cause traffic disturbances. They can be difficult to get accurate information from. Accurate detection of traffic jams is a similar factor.

Weather information sources

Weather information can be retrieved from national meteorological institutes or international web sites. The information is usually transmitted in standardised SYNOP (surface synoptic observations) messages.

The website ogimet.com contains a collected database of SYNOP messages. Searches can be performed on the databases and stored in text files. The text files can easily be parsed and saved to database.

The SYNOP messages are a standard way of sharing weather information from manned and automated weather stations. The use of the messages, according to World Meteorological Organization (WMO) Resolution 40 (http://www.nws.noaa.gov/im/wmocovr.htm), is free and unrestricted for research purposes.

Traffic information sources

As an example of national traffic information to be collected in the central database, Mediamobile will provide the following information in the Finnish FOT:

- TMC messages on road works and traffic disturbances.
- Periodical updates for traffic volumes / travel times for hundreds of roads.
- Road weather information from measurement stations.

The information transfer is arranged between servers as XML format push packets, being sent at appropriate intervals.

6. TOOLS FOR ACCESSING DATABASE

The project uses the two most common open source databases currently, MySQL and PostgreSQL. Free or relatively inexpensive database viewers are available for both. The following figure shows the user interface of MySQL Query Browser, when selecting data from “FMS” data table.

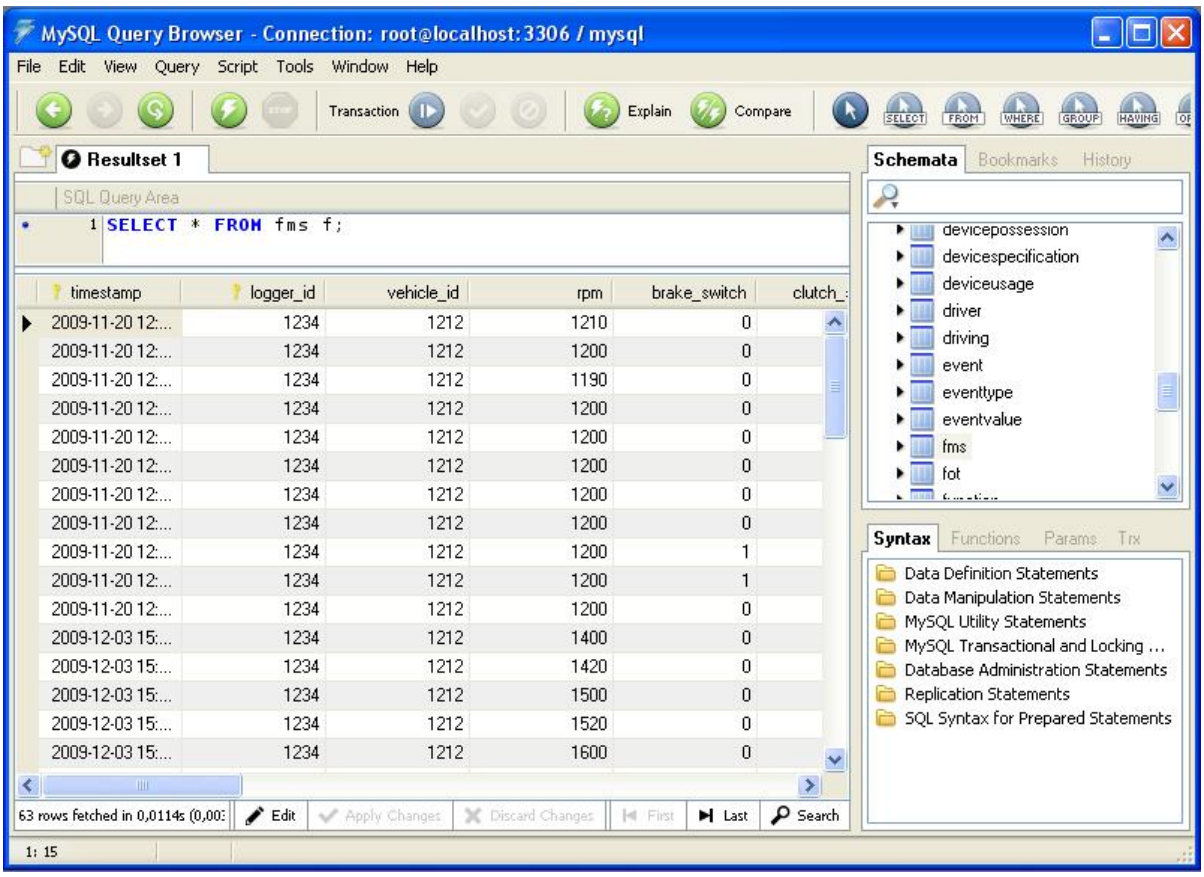


Figure 9. MySQL Query Browser screenshot, a freely available database viewer.

These generic viewers provide database experts tools to view and modify data. They are reasonably simple to use but this requires a command of the SQL language.

Conversion of data

As a common feature of many web-based database viewer tools (e.g. Figure 4 about phpMyAdmin), they offer methods for exporting certain data sets at least in Excel, web page, PDF and CSV (comma-separated values) formats. These data sets can be again imported in tools such as Matlab or Excel, without any database experience.

MATLAB direct database access

The following gives an example of accessing a MySQL database with MATLAB. It first requires installing MySQL ODBC (Oben Database Connectivity) Connector. Similar drivers exist for all common databases. These ODBC connectors provide standard access to databases for applications. Next, a new ODBC data source has to be configured using Windows' Control Panel's Administrative Tools.

This example requires the MATLAB DB Toolbox. For easy access and not having to study DB Toolbox commands, user Jozef Rudy has provided a helpful script for MATLAB user community at http://www.mathworks.com/matlabcentral/fileexchange/24703-mysql.

In this short script only one line has to be modified for connecting to a pre-configured ODBC data source:

```
obj.dbconn = database('user defined ODBC source name','user','password');
```

After this short configuration, any SQL commands can be executed using the following command format:

```
result = test.execute('SELECT * FROM driving');
```

The result holds the data that has been selected from the database for further analysis.

7. ANALYSIS AND QUERY EXAMPLES

The two examples given here draw upon logged data from the central TeleFOT database. An assumption is that all cases (or journeys) will be complete. That is, all of the background information relating to the driver characteristics will be available for data mining and GPS information relating to the date and time of the journey are complete.

Example 1

Hypothesis: There is a change in the proportion of road type driven on.

In particular we are looking for a *shift to or away from motorway travel.*

Data selection from the central database would require;

- Data for both the 'before' and 'after' phases of a within subjects design LFOT.
- Journeys where the device was in use in the 'after' phase.
- Data from the whole collection period.
- Longer journeys, over 25 km.
- Journeys where map-matching to determine road type has been undertaken in the post processing.
- Ideally journeys where there would have been the potential to use a motorway during the course of the journey.

Query

As a first step, all legs are selected from a separately produced summary table Driving (see Table 2), where logger_id is of a logger that the user has been in possession with during a certain time period:

```
SELECT * FROM driving d where d.logger_id = (select device_id from devicepossession p where p.start_time < "2009-05-01 00:00" and p.stop_time > "2009-06-01 00:00" and p.driver_id = 1);
```

The result table has the same columns and driving performance indicators as the table Driving. These include e.g. the leg ID, logger ID, timestamps, total distance driven and total fuel consumption.

The user ID could also come from a longer selection from FOT users in a country, but this example uses it directly.

The time stamps dividing before/after would be used to first select everything "before" and calculate further indicators for those retrieved legs (e.g. total distance driven during a time period as a sum of legs). This would be repeated for "after".

The table Driving would give a list of legs, each leg having a value (or indicator) for %_on_motorway (and for other road classes as well).

If we would want to select only legs > 25 km, the example would continue like

```
SELECT * FROM driving d where d.logger_id = (select device_id from devicepossession p where p.start_time < "2009-05-01 00:00" and p.stop_time > "2009-06-01 00:00" and p.driver_id = 1) and d.total_distance_driven > 25000;
```

The possibility of using motorways would come from leg classification based on coordinate bounding box, total distance, number of turns etc. The classifications will be studied during the project. There could for example be a classification that says "on this type of route it is possible to use motorways or not". Currently, the classification of common starting and ending points is included in the driving table as an indicator.

Example 2

Hypothesis: There is a change in the length of journeys with comparable origins and destinations

Data selection would supposedly require:

- Data for both the ‘before’ and ‘after’ phases of a within subjects design LFOT.
- Journeys where the device was in use in the ‘after’ phase.
- Data from the whole collection period.
- Journeys where the total distance in km is known.

- Journeys with comparable origins and destinations – for each journey selected from the 'before' phase for a particular subject there needs to be a matched journey in the 'after' phase. This constitutes a pair of comparison journeys. There should be a least 10 of these pairs for each origin/destination.
- The journeys must not take in deviation for a purpose other than re-routing by system / following traffic information given by system.

Query

Firstly, the same type of legs are selected from the Driving table, based on common starting points, journey length, number of turns, time of day and other preferred indicators. It might be practical to produce a new separate table "common_routes" as a result from the classification, listing e.g.

- common_route_category ID
- distance avg XXX km
- frequency X/month
- number_of_times_used XX

We would have to repeat the selects above (the before/after a time stamp is already included) for each of these suitable categories (e.g. where the number_of_times_used > X). We would also have to add a new line to ensure that the device has been used during the leg. That would mean another search from deviceusage (assuming here logger = device!):

```
SELECT * FROM driving d where exists (select start_time from deviceusage u where u.start_time > d.timestamp_start and u.stop_time < d.timestamp_stop) and d.total_distance_driven > 1000 and d.category = "a1" and d.logger_id = (select device_id from devicepossession p where p.start_time < "2009-05-01 00:00" and p.stop_time > "2009-06-01 00:00" and p.driver_id = 1);
```

The request for covering only routes with no uncommon deviation would have to be included in the classification of common routes based on e.g. their area and length.

CONCLUSIONS

This document presented the first database tools to be used in the TeleFOT project. Many of the tools come with the tested services, but in these cases the data is eventually transferred to the project's central server for analysis. Data quality and user administration is however controlled mainly during data collection.

The project develops tools for extracting driving performance indicators from raw data and for map matching. These programs are run periodically at the central server but can also be used at analysts' personal computers. These Java language tools require programming experience to develop further but single performance indicators have been made relatively easy to add to the code by providing ready templates.

The automatic extraction of driving performance indicators makes it easier for analysts to quickly get an overview of data and answer many of the initial hypotheses. Further analysis will be performed accessing data sets with analyst-preferred tools such as SPSS or MATLAB.

This document gives examples of accessing the database with MATLAB and generic database viewers, along with queries to select data to study hypotheses.

The tools are being further developed based on feedback from analysts and first FOT pilots.

REFERENCES

[1] Scholliers, J., Koskinen, S., Vasama, J. and Martikainen, H. 2009. Data Collection in Field Operational Tests of Telematic Functions for Nomadic Devices in Vehicles. Proceedings of ITS World 2009 Stockholm.

[2] Franzén, S. and Karlsson, I.C.M. 2009. TeleFOT project deliverable D1.5, Testing and Evaluation Strategy.

[3] Karlsson, I.C.M. et al. 2009. TeleFOT project deliverable D.2.2.1 Testing and Evaluation Strategy.

[4] KIRCHER, K ET AL. 2008. FESTA D2_1_PI_MATRIX_FINAL.XLS HTTP://WWW.ITS.LEEDS.AC.UK/FESTA/DOWNLOADS/FESTA%20D2_1_PI_MATRIX_FINAL.XLS