smartCEM

Smart connected electro mobility

D3.2 Common Data Exchange Protocol for smartCEM

Version number	F
Main author	UNEW
Dissemination level	PU
Lead contractor	UNEW
Due date	15-09-2014
Delivery date	15-09-2014

CIP - Information and Communications Technologies Policy Support Programme (ICT PSP)
Information Society and Media Directorate-General
Grant agreement no.: 297328
Pilot type B

Revision and history sheet

Version history			
Version	**Date**	**Main author**	**Summary of changes**
0.1	Oct 2013	Simon Edwards (UNEW)	Initial template
0.2	May 2014	Simon Edwards (UNEW)	Revised table of contents and proposed partner contributions
0.3	June 2nd 2014	Simon Edwards (UNEW)	Additional content in Ch. 2
F	Sept 15th 2014	Simon Edwards (UNEW)	Input from pilot sites and data analysis leader
	Name		**Date**
Prepared	Simon Edwards		15-09-2014
Reviewed	Simon Edwards		15-09-2014
Authorised	Fernando Zubillaga		15-09-2014
Circulation			
Recipient		**Date of submission**	
European Commission		15-09-2014	
Pilot consortium		15-09-2014	

Authors (full list)

Simon Edwards (UNEW), Graeme Hill (UNEW), Txomin Rodriguez Villa (Tecnalia), Josep Laborda (RACC), Oier Iribar Ondarra (ENNERA)

Project Coordinator

Fernando Zubillaga
MLC ITS Euskadi
Clúster de Movilidad y Logística
Centro de Negocios CTVi, Oficina 1, 3-A
Lermandabide, 8. Polígono Industrial Júndiz
01015 Vitoria-Gasteiz
Tel.: +34 945 10 80 88
Email: fzubillaga@mlcluster.com

Legal Disclaimer

Table of Contents

Abbreviations

Abbreviation	Definition
CDB	Central Database
CIP	Competitiveness and Innovation Framework Programme
CS	Charging Station
CSV	Comma Separated Values
CYC	Charge Your Car
DB	Database
EV	Electric Vehicle
FTP	File Transfer Protocol
sFTP	Secure File Transfer Protocol
GB	Gigabyte
KM	kilometres
LDB	Local Database
OBU	On-Board Unit
PI	Performance Indicator
WP	Work Package

Executive Summary

D3.2 *Common Data Exchange Protocol for smartCEM* describes the structure of the central database and the data exchange mechanisms used to transfer data from local databases at each smartCEM pilot site to the central database. It also includes a short explanation of the rationale behind the central database approach. Furthermore it emphasises the importance of the link between WP3 and WP4 in terms of the utilisation of the collected data for analysis and evaluation purposes. The data must be in an appropriate format and of appropriate quality, and must be easily extractable from the CDB. Finally, the ownership of the data collected in the project and scope for its use after the project concludes is clarified.

1. Introduction

1.1. Purpose and scope of D3.2

D3.2 *Common Data Exchange Protocol for smartCEM* describes the structure and characteristics of the central database (CDB) developed in smartCEM WP3. The purpose of the CDB is to store post-processed data from local databases (LDB) at the four pilot sites (Barcelona, Gipuzkoa, Newcastle upon Tyne, and Reggio Emilia) in a usable format for analysis.

This deliverable builds on D2.4 *Logging Tools Database Definition* which defined the local systems and processes in place at each pilot site for collecting, storing, managing and processing data during the local trials. The local database architecture described in that deliverable is used for the baseline and full operational data acquisition processes.

In addition to description of the CDB structure and characteristics, this deliverable also explains the data exchange mechanisms used to transfer the data from the local databases to the CDB in a format that ensures consistent data is available in a form that all can understand and use as required.

Furthermore, D3.2 forms a strong link to D4.3 *smartCEM Experimental Design* which defined a set of measures based on each site's characteristics and operational scenarios, and proposed a methodology for deriving performance indicators from these measures. Whereas D2.4 builds on D4.3 by describing the mechanics by which the data defining the local measures is collected at pilot sites, D3.2 describes how the data from the local sites that resides in the CDB can be quality controlled, and how it can be used by the evaluation partners to obtain performance indicators for the analysis and evaluation process.

A final important issue addressed by D3.2 is that of data ownership and its availability after the termination of the smartCEM project.

1.2. Structure of the document

The structure of D3.2 is as follows:

- Chapter 2 describes the structure of the central database and data exchange mechanisms. It also includes an explanation of the rationale behind the central database approach.
- Chapter 3 describes the role of the central database from the evaluation viewpoint describing the extraction of data for evaluation purposes in terms of performance indicators, and techniques for ensuring data quality. It also summarises issues relating to the ownership and scope for future use of the data collected in smartCEM during and beyond the project lifespan.

2. Structure of Central Database and Data Exchange Mechanisms

This section describes:

- Data transfer protocols between local databases (LDBs) at pilot sites and the central database (CDB) located at University of Newcastle;
- The structure of the CDB and the rationale behind this approach;
- The mechanics of qualitative data collection in local databases.

2.1. Rationale for the central database

Through hosting a CDB University of Newcastle (UNEW) ensures that smartCEM enjoys a single system which is responsible for three main activities:

- Gathering and storing all the data collected at the individual pilot sites;
- Creating cohesive reports on the disparate data sources, which would be difficult without one system with easy access to all the data sources;
- Offering a single point of dissemination of information for all reporting by sites and any interested parties.

In particular, storing all data in the same location will allow for a much easier synthesis of possible data queries between the sites and will allow for connections between the different data sets to be more easily ascertained.

Although the creation of a single repository for data will lead to some redundancies in the storage, it will act as an additional security measure against possible catastrophic failure at the local level. The cost of additional storage space on one server is insignificant compared to the potential problems that could be caused by unforeseen events.

2.2. Structure of the central database

In general the relationships between the data in the local database and the data in the CDB will be of the same form for each of the pilot sites. At each pilot site there is a local database (in whatever form is most appropriate for that site) which will be updated using the live data from that site's baseline and operations phases. The processing steps to clean and harmonise the data will be accomplished at the local level as only the pilot sites have the requisite knowledge to manage their own data streams.

After harmonisation the data is transferred to the server hosting the CDB. Typically this is achieved using sFTP and csv files, but other options may be explored if needed. The transferred data is then uploaded into a database that replicates the structure of the local database, but in a common database format. This is currently SQLite due to the speed and lightweight nature of this database format.

Performance indicators can then be generated using a series of automatic queries on the data stored in the CDB and stored within the appropriate section of the same database.

The relationship between the local databases and the central database is shown in the following diagram (Fig 2-1).

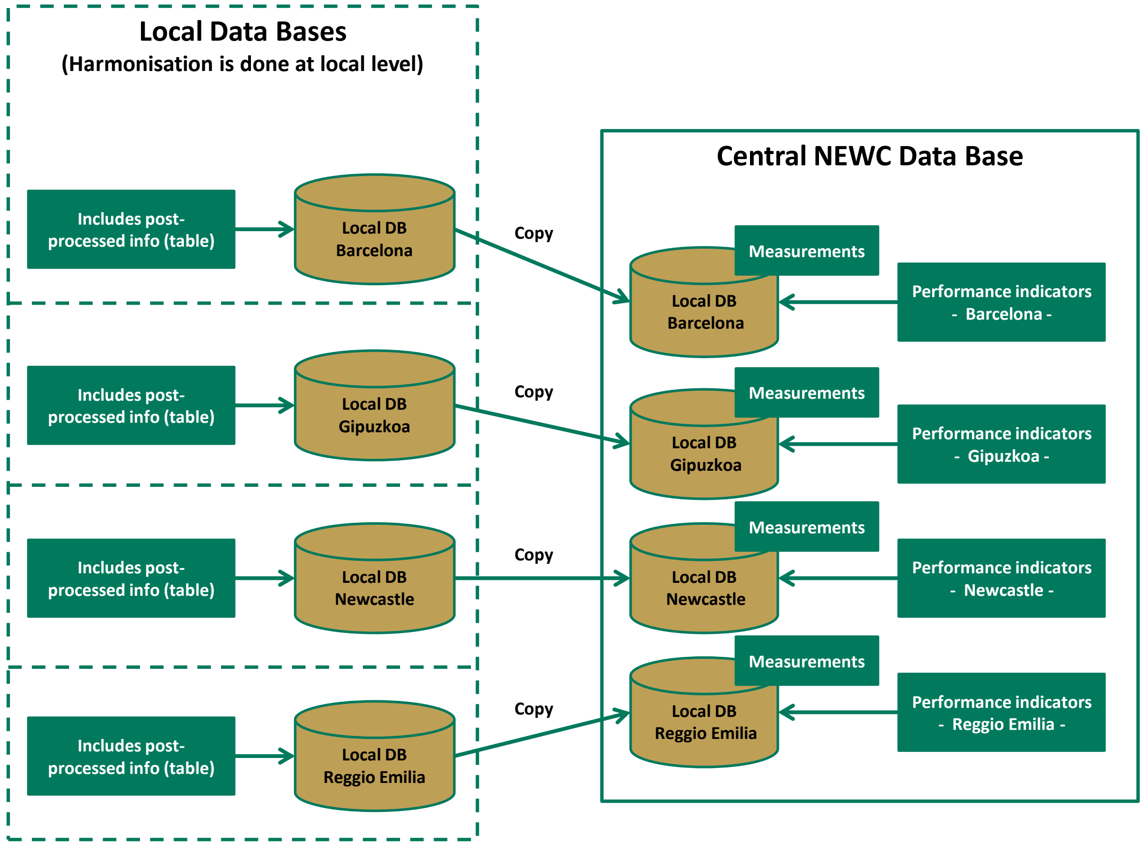

Fig. 2-1. smartCEM relationship between local databases and central database

2.3. Data exchange between local databases and central database

The process by which each of the four smartCEM pilot sites collects data into their local databases is explained in D2.4 *Logging Tools Database Definition*.

The collection of data on the CDB side is entirely automated with the data from each of the pilot sites automatically uploaded into the correct section of the CDB. This is a process requiring no user involvement.

The use of a flat csv file as the data transfer allows for the maximum variety of database systems to be used by the pilot sites whilst also strictly specifying a data format for the transfer. It is possible to extend this system by connecting the CDB and the local databases using a more automated system. This process is at the discretion of each site.

In the following sub-sections a brief summary is provided for each pilot site of quantitative data to be collected, including updated versions of the tables originally presented in D2.4. The 'MEASURE' column is a code corresponding to the table of measures previously presented in D4.3 and D2.4 annex.

2.3.1. Barcelona

Registered data is collected from the data inputs of the MOTIT service. Sources for data acquisition include data loggers (OBU) and the booking system.

NAME	TYPE	MEASURE	DESCRIPTION
BK_MODEL	INT	-	Bike model
BK_VIN	INT	ME_252	Unique vehicle identifier
BK_ID	INT		Bike ID used in relations

Table 2-1. Bikes (Barcelona)

NAME	TYPE	MEASURE	DESCRIPTION
BL_TIMESTAMP	DATETIME	ME_102, ME_103 (starting date & hour)	Date-time of the start of trip
BL_BAT1_SERIAL	DECIMAL(4,0)	-	Battery 1 serial number (not available until September-October)
BL_BAT2_SERIAL	DECIMAL(4,0)	-	Battery 2 serial number (not available until September-October)
BL_BAT3_SERIAL	DECIMAL(4,0)	-	Battery 3 serial number (not available until September-October)
BL_BAT4_SERIAL	DECIMAL(4,0)	-	Battery 4 serial number (not available until September-October)
BL_SW_MOTOR	DECIMAL(3,0)	-	Version of motor software
BL_SW_BMS	DECIMAL(3,0)	-	Version of BMS software
BL_SW_ICM	DECIMAL(3,0)	-	Version of ICM software
BL_ID_BIKE	INT	ME_252	Bike Identifier

Table 2-2. Reservations (Barcelona)

NAME	TYPE	MEASURE	DESCRIPTION
EV_ID_BL	INT	RELATED TO ME_118	Unique book identifier
EV_TYPE	INT	-	Unique event type identifier
EV_COMMENT	VARCHAR(255)	ME 001	Event description

Table 2-3. Events (Barcelona)

NAME	TYPE	MEASURE	DESCRIPTION
SHF_TIMESTAMP	DATETIME	ME_101	Date-time at the sampling instant
SHF_ID_BL	INT	RELATED TO ME_118	Unique book identifier
SHF_BUS_VOLTAGE	DECIMAL(3,0)	ME_307	Bus voltage at the sampling time
SHF_BUS_CURRENT	DECIMAL(3,0)	ME_306	Bus current at the sampling time
SHF_SPEED	DECIMAL(3,0)	ME_251	Speed of the vehicle at the sampling time
SHF_THROTTLE	DECIMAL(3,0)	ME_712	Throttle of the vehicle at the sampling time
SHF_LONGITUDE	DECIMAL(9,6)	ME_119	GPS longitude for the location of the vehicle at the sampling time
SHF_LATITUDE	DECIMAL(9,6)	ME_119	GPS latitude for the location of the vehicle at the sampling time
SHF_ALTITUDE	DECIMAL(6,0)	ME_119	GPS altitude for the location of the vehicle at the sampling time
SHF_RPM	DECIMAL(4,0)	-	Engine RPM at the sampling time

Table 2-4. Signal HF (Barcelona)

NAME	TYPE	MEASURE	DESCRIPTION
SLF_TIMESTAMP	DATETIME	ME_101, 102	Date-time at the sampling instant
SLF_TOTAL_MILEAGE	DECIMAL(10,1)	ME_108	Accumulated mileage
SLF_STATUS	DECIMAL(4,0)	-	Bike status at the sampling time
SLF_AMBIENT_T	DECIMAL(3,0)	ME_907	Ambient temperature at the sampling time
SLF_BAT_1_SOC	DECIMAL(4,0)	ME_305	Battery 1 state of charge at the sampling time
SLF_BAT_2_SOC	DECIMAL(4,0)	ME_305	Battery 2 state of charge at the

			sampling time
SLF_BAT_3_SOC	**DECIMAL(4,0)**	ME_305	Battery 3 state of charge at the sampling time
SLF_BAT_4_SOC	**DECIMAL(4,0)**	ME_305	Battery 4 state of charge at the sampling time
SLF_BAT_STATE_FAULT	**DECIMAL(3,0)**	ME_303?	Battery fault code
SLF_BIKE_FAULT	**DECIMAL(3,0)**	-	Bike fault code
SLF_INTERLOCK_POSITION	**DECIMAL(3,0)**	-	Bike interlock position
SLF_LIGHT_STATUS	**DECIMAL(3,0)**	ME_908	The status of lights

Table 2-5. Signal LF (Barcelona)

2.3.2. Gipuzkoa

Registered data is collected from the data inputs of the car-sharing operator in Elgoibar, the hybrid bus operators and the hospital service in Donostia-San Sebastian. Sources for quantitative data acquisition include data loggers (OBU) and booking systems. The tables below relate to car-sharing in Elgoibar.

NAME	TYPE	MEASURE	DESCRIPTION
RESERVATION_ID	**INT**	ME_112	A unique identifier to define each reservation (unique for all operators and reservations)
OPERATOR	**VARCHAR**	-	Car-sharing operator of the reservation
OP_RESERVATION_ID	**INT**	-	A subjective unique identifier for reservations for each operator. This identifier can overlap between two reservations of two different operators, that is why RESERVATION_ID (unique for all the companies) has been defined as the key identifier and first column of the reservations table
USER_ID	**INT**	ME_002	Unique identifier (could be the User ID registered for the user) for the user that makes the reservation
VEHICLE_ID	**VARCHAR**	ME_252	Number on the plate of the car that has been reserved

RESERVATION DATE	**DATE**	-	Refers to the date when the user has made the booking process
RESERVATION TIME	**TIME**	-	Refers to the time when the user has made the booking process
STATUS	**VARCHAR**	-	Defines the status of the reservation. If it has been completed (vehicle picked up and trip done), not completed (vehicle not picked up and trip not done) or cancelled
EST_INIT_DATE	**DATE**	-	Estimated initial date for the reservation. Estimation done by the user when making the reservation.
EST_INIT_TIME	**TIME**	-	Estimated initial time for the reservation. Estimation done by the user when making the reservation.
EST_FINAL_DATE	**DATE**	-	Estimated final date for the reservation. Estimation done by the user when making the reservation.
EST_FINAL_TIME	**TIME**	-	Estimated final time for the reservation. Estimation done by the user when making the reservation.
EST_KM	**INT**	-	Estimated amount of KM. Estimation done by the user when making the reservation
REAL_INIT_DATE	**DATE**	ME_103	Real initial date for the reservation. Registered when the user picks up the previously reserved vehicle
REAL_INIT_TIME	**TIME**	ME_103	Real initial time for the reservation. Registered when the user picks up the previously reserved vehicle
REAL_FINAL_DATE	**DATE**	ME_104	Real final date for the reservation. Registered when the user returns the vehicle at the end of the reservation process
REAL_FINAL_TIME	**TIME**	ME_104	Real final time for the reservation. Registered when the user returns the vehicle at the end of the reservation process
REAL_KM	**DOUBLE**	ME_120	Real amount of KM travelled by the user, during the reservation
BASE_ID	**INT**	-	Identification of the base where the user has picked up the vehicle
CP_ID	**INT**	ME_505	Identification for the charging station from which the user has

			picked up the vehicle
INITIAL_CHARGE_PERCENT	**INT**	ME_301	The battery level of the car, when the user picked up the vehicle (%)
INITIAL_CHARGE_KWH	**DOUBLE**	ME_301	The battery level of the car, when the user picked up the vehicle(kwh)
FINAL_CHARGE_PERCENT	**INT**	ME_302	The battery level of the car, when the user returned the car after the reservation (%)
FINAL_CHARGE_KWH	**DOUBLE**	ME_302	The battery level of the car, when the user returned the car after the reservation(kwh)
CONSUMPTION_PERCENT	**INT**	ME_404	The battery level difference between initial and final (%)
CONSUMPTION_KWH	**DOUBLE**	ME_404	The battery level difference between initial and final (kwh)
TEMP_MIN	**INT**	ME_901	Minimum temperature for the region when the user picked up the car
TEMP_MAX	**INT**	ME_901	Maximum temperature for the region when the user picked up the car
TEMP_FORECAST	**VARCHAR**	ME_901	The temperature forecast for the region, when the user picked up the car
WIND_FORECAST	**VARCHAR**	ME_901	The wind forecast for the region, when the user picked up the car
WEATHER_FORECAST	**VARCHAR**	ME_901	The weather forecast for the region, when the user picked up the car. These measures come from an external API that gives weather forecast for the region

Table 2-6. Reservations (Gipuzkoa)

NAME	TYPE	MEASURE	DESCRIPTION
ID	**INT**	-	Unique identifier for event register
REGISTER_DATE	**DATE**	ME_502	Date when the event is registered
REGISTER_TIME	**TIME**	ME_502	Time when the event is registered
RESERVATION_ID	**INT**	RELATED TO ME_118	The reservation id to which the event corresponds
OPERATOR	**VARCHAR**	-	The operator to which the register

			corresponds, for example, EMUGI
OP_RESERVATION_ID	**INT**	RELATED TO ME_118	The reservation id for the operator to which the event corresponds
TYPE	**VARCHAR**	-	Event type
GPS_LAT	**DOUBLE**	-	GPS latitude for the location of the vehicle at the event
GPS_LONG	**DOUBLE**	-	GPS longitude for the location of the vehicle at the event
SPEED	**INT**	-	Speed of the vehicle at the event instant

Table 2-7. Events (Gipuzkoa)

2.3.3. Newcastle upon Tyne

Registered data is collected from the data inputs of the participating vehicles. Data is collected directly in the CDB which is hosted at the University of Newcastle. Sources of quantitative data acquisition are on board data loggers (OBU).

NAME	TYPE	MEASURE	DESCRIPTION
ID	**INT**	-	Unique identifier for event register
USERID	**INT**	ME_002	Individual id for each user
CHARGE_ID	**INT**	ME_505	The id corresponding to the charge point
CHARGE_TRANS	**INT**	ME_506	Charging transactions id
TIMESTAMP	**DATETIME**	-	Date-time when the event is registered
LOGGERID	**VARCHAR**	RELATED TO ME_118	The reservation id to which the event corresponds
TYPE	**VARCHAR**	-	Event type (Charge or Drive, possibly Park)
GPS_LAT_START	**DOUBLE**	ME_119	GPS latitude for the location of the vehicle at the event's moment (start of journey)
GPS_LONG_START	**DOUBLE**	ME_119	GPS longitude for the location of the vehicle at the event's moment (start of journey)
GPS_LAT_END	**DOUBLE**	ME_119	GPS latitude for the location of the vehicle at the event's moment (end of journey)
GPS_LONG_END	**DOUBLE**	ME_119	GPS longitude for the location of the

			vehicle at the event's moment (end of journey)
SPEED	**INT**	-	Speed of the vehicle over the event, if applicable
EVENTDURATION	**DATETIME (delta)**	-	Duration of the event
EVENTDISTANCE	**DOUBLE**	-	Event distance, if applicable
ENERGY	**DOUBLE**	-	Total Energy either consumed or used in the charge/drive event
REGEN	**DOUBLE**	ME_402	Total energy regenerated
TEMPERATURE	**DOUBLE**	ME_907	Ambient Temperature of the event
HARD_BRAKE	**INT**	ME_716	Number of hard braking events
HARD_ACCEL	**INT**	ME_718	Number of hard acceleration events

Table 2-8. Events table (Newcastle)

NAME	TYPE	MEASURE	DESCRIPTION
LOGGERID	**VARCHAR**	-	Unique identifier for logger
USERID	**INT**	ME_002	Individual id for each user
VEHICLEID	**VARCHAR**	ME_252	Unique identifier for the vehicle the logger is on
DATE_INSTALLED	**DATETIME**	-	Date the logger was installed
DATE_REMOVED	**DATETIME**	-	Date the logger was removed, if applicable

Table 2-9. OBU/ vehicle table (Newcastle)

2.3.4. Reggio Emilia

Registered data is collected from the data inputs of the participating vehicle operator (Municipality of Reggio Emilia). Sources for quantitative data acquisition include data loggers (OBU) and the key management system.

NAME	TYPE	MEASURE	DESCRIPTION
ID	**INT**	-	Unique identifier for event register

TIMESTAMP	**DATETIME**	ME_102, ME_103	Date-time when the event is registered
TYPE	**VARCHAR**		Event type
GPS_LAT	**DOUBLE**	-	GPS latitude for the location of the vehicle at the event's moment
GPS_LONG	**DOUBLE**	-	GPS longitude for the location of the vehicle at the event's moment
SPEED	**INT**	-	Speed of the vehicle at the event instant
TRAFFIC (Optional)	**VARCHAR**	ME_904	Traffic incidents to which the event could be related. In cases where traffic incidents are not registered around the GPS position this value could be defined as NULL
STARTING TIME CHARGING EVENT	**DATETIME**	ME_502	Starting time of the charging session
END TIME CHARGING EVENT	**DATETIME**	ME_501	Ending time of the charging session

Table 2-10. Events (Reggio Emilia)

2.4. *Qualitative data acquisition*

2.4.1. Core (common) surveys

As reported in D3.1 *Operational Plans for smartCEM Platforms*, core baseline and operations phase questionnaires have been designed by DLR (see D4.4 *smartCEM Assessment Tools*). The core questionnaires form the basis of the qualitative data collection to be performed at each pilot site, and have been designed in collaboration with other CIP projects in order to ensure a harmonised approach. The core questionnaire is built around the four subcategories identified for evaluation: EV acceptance, smartCEM services acceptance, range anxiety, and willingness to pay. However, pilot sites consider only the questions corresponding to the services implemented; thus, for some sites certain questions or even whole subcategories are not relevant. This is particularly the case regarding willingness to pay, which for commercial reasons is not covered in the Newcastle site or the Barcelona site, or for operational reasons in the Reggio Emilia site where participants are fleet drivers from the local municipality.

It should also be noted that questions about smartCEM acceptance are only performed on the operations phase questionnaire; as smartCEM is not implemented at the baseline phase these questions are not relevant to the baseline.

Table 2-11, reproduced from D3.1, summarises the subcategories surveyed at each site.

	BARCELONA	GIPUZKOA			NEWCASTLE	REGGIO EMILIA
		Car-sharing	Hybrid bus	Hospital		
EV acceptance	Y	Y			Y	
smartCEM acceptance	Y	Y	Y	Y	Y	Y
Range anxiety	Y	Y			Y	Y
Willingness to pay		Y				

Table 2-11. Application of questionnaires per pilot site

Finally, supplementary questions can be designed by each pilot site relevant to specific attributes of that site. This is particularly important in recognising the locally-specific characteristics of EV usage in different regions of Europe, and also serves to enhance stakeholder engagement in each of the four pilot sites.

2.4.2. Non-EV user survey

A further survey performed in smartCEM is the non-EV user questionnaire. It is highly relevant to survey non-EV users as this is the future untapped market for EV uptake and it is important to understand what barriers or perceived barriers exist to achieving this. A questionnaire has been designed by Tecnalia and is being implemented at all sites in online format.

2.4.3. Status of qualitative data acquisition

The following tables summarise the vital statistics at each pilot site at August 2014. At the present time only the baseline surveys have been completed. Dates of completion, method of circulation, and final response figures for the operations phase and non-EV user survey will be provided in D3.3 *Final Operation Report*, along with annexes presenting all questionnaires performed at each site. Analysed results will be presented in D4.5 *Results of the evaluation.*

NB in the tables below TBC = to be confirmed

Type of survey	Date administered	How administered	Number returned	Analysis
Baseline	-	-	-	Not applicable
Operational (core)	June 2014	Online (Eval & Go) to smartCEM participants (users of MOTIT e-	7	WP4

		scooter sharing service)		
Operational (supplementary)	-	-	-	Not applicable
Non EV users	May 2014	Online (Eval & Go) to RACC employees; dissemination through Creafutur Twitter (@creafuturcom), RACC (@ClubRACC), Barcelona Municipality mobility dept @BCN_Mobilitat, and Live Office (@LiveprojectBCN)	725 (all sites)	WP4

Table 2-12. Summary of survey implementation (Barcelona)

Type of survey	Date administered	How administered	Number returned	Analysis
Baseline	March 2014	Online questionnaire (Eval&Go)	23	WP4
Operational (core)	TBC	Online questionnaire (Eval&Go)	TBC	WP4
Operational (supplementary)	TBC	Online questionnaire (Eval&Go)	TBC	WP4
Non EV users	April 2014	Online questionnaire (Eval&Go)	725 (all sites)	WP4

Table 2-13. Summary of survey implementation (Gipuzkoa)

Type of survey	Date administered	How administered	Number returned	Analysis
Baseline	May 2014	Email WORD survey to smartCEM participants with extended circulation to members of Charge Your Car (CYC)	274	WP4
Operational (core)	TBC (proposed October-November 2014)	Online (Survey monkey) to smartCEM participants with extended circulation to	TBC	WP4

		members of Charge Your Car (CYC)		
Operational (supplementary)	TBC (proposed October-November 2014)	Online (Survey monkey) to smartCEM participants with extended circulation to members of Charge Your Car (CYC) - a selection of questions designed by CYC and Bosch specifically relating to CS management	TBC	WP4
Non EV users	September 2014	Online (Survey monkey) to students at University of Newcastle Civil Engineering department	725 (all sites)	WP4

Table 2-14. Summary of survey implementation (Newcastle)

Type of survey	Date administered	How administered	Number returned	Analysis
Baseline	June 2014	Hard copies of questionnaire to be filled in by users	TBC	WP4
Operational (core)	September 2014	Hard copies of questionnaire to be filled in by users	TBC	WP4
Operational (supplementary)	Late 2014 (TBC)	TBC	TBC	WP4
Non EV users	May-June 2014	Online to 60 people in two companies and up to 400 people via UNIMORE mailing lists	725 (all sites)	WP4

Table 2-15. Summary of survey implementation (Reggio Emilia)

Survey data will not be stored in the CDB. Online survey responses will be post-processed and analysed directly by the data analysis team in WP4, whilst emailed responses will be processed by local pilot sites for submission by WP4.

3. Extraction of Data from Central Database for Evaluation

3.1. Data Analysis and Processing for Electric Vehicles

The approach to data extraction and analysis presented here is a generic approach for analysing and processing electric vehicle data.

There are three main stages in converting raw vehicle data into usable metrics: data checking, data processing and data analysis.

3.1.1. Data Checking

Data checking is the process of initially examining the data for any fundamental flaws (e.g. no missing data points, the same number of columns, the same data types in each of the columns). In general this is not necessary for most forms of data collection but it is possible that errors further down the data flow can be traced back to fundamental errors in the data being collected. The types of errors that can be encountered at this stage are varied and not easy to predict. Hence there is no set technique for dealing with data errors at this stage. However, to facilitate the possible re-examination of data after an error has been encountered all data is permanently stored so that it can be retrieved and re-analysed if needed.

In previous projects data repair/removal has taken place at the point at which the data is being analysed i.e. the repair/removal was never written back to the original data set, thus preserving the integrity of the initial data set.

In general once an error has been identified in the data there are two possible options

1. Repair the data
2. Remove the data

The second option is by far the easiest option but is typically only done as a last resort. In general it is preferable to repair the data. fThere are numerous different techniques for repairing data and the exact method required will depend on the nature of the data to be repaired.

Repair from Internal Sources

In some instances the data being recorded may be unique to that particular vehicle at that particular moment (voltage, speed, etc.) and hence it will be impossible to derive the measurement from outside sources. In these circumstances the required data will generally be derived from either interpolating between good data sets, or deriving that data as a function of other metrics within the system.

A simple example of this in previous projects was the interpolation of missing GPS coordinates. It was assumed that for a single missing GPS coordinate, an approximate answer could be assumed to be the point in between the previous "good" coordinates. This type of simple interpolation can be vastly improved through the use of map matching algorithms and previous heading metrics.

Repair from External Sources

For some data types, the recorded data may be essentially a replication of an external data type. An example of this in a previous project was ambient temperatures. In one particular fleet of electric vehicles the ambient temperature sensors initially failed to work. Rather than replace the temperature with a null value, the ambient temperatures as derived from external weather stations was used as a proxy for the vehicle's ambient temperature. Whilst this was not an exact match, it was good enough to allow for an analysis that showed systematic variation in efficiency with temperature.

3.1.2. Data Processing

With the basic data sets derived from a GPS location and timestamp (date/time plus latitude, longitude and altitude) it is possible to create multiple different useful metrics for a vehicle including the average speed of a vehicle, the distance travelled per trip or journey duration.

Before deriving a metric, the data for each vehicle will generally be stripped of all data without a valid timestamp before being ordered by time stamp. Data from vehicle loggers, especially if local caching of the data is occurring, can frequently be sent out of order. If the metrics for the vehicle need to be created for each individual trip then the data will also need to be separated according to an algorithm that can split data up into individual trips. In previous work this was accomplished using the ignition, and flagging a timestamp as being either the start of a trip (if the ignition was turned on) or the end of a trip (if the ignition was turned off).

Distance Example

To derive the total distance driven in each trip, the Haversine formula is used to calculate the distance between each successive GPS point. The Haversine formula is a formula that calculates the distance between successive paired values of longitude and latitude. The initial output from this is a vector of numbers with each element of the vector containing the distance between the two positions of the vehicle. Calculating the total distance travelled is then a simple matter of summing the distance.

$$Dist(vehicle) = \sum_{t=1}^{t=n-1} Haversine(pos_t, pos_{t+1})$$

Average Speed per Trip

To calculate the average speed for each trip it is first necessary to calculate two numbers: the total distance of the trip (as calculated previously) and the total duration of the trip. The duration of the trip is simply derived from the first and last time stamps in the full data set for each individual trip. Care must be taken that no extreme time stamp values are used in this calculation.

Instantaneous Speed

As well as the average speed it is also useful to calculate the instantaneous speed at every point in the trip. This data can then be used in further analysis (for example to check the speeds at certain points in a journey or check the speeds at specific points on a map) The instantaneous speed is calculated here in the same way as the average speed except that instead of a single number for the whole data set, a speed is produced for each entry in the data set.

$$Speed_t = \frac{Haversine(pos_t, pos_{t+1})}{Duration(t, t+1)} \; for\ t\ in\ 1, t-1$$

The instantaneous speed is one data point shorter than the total size of the data in the processed data set.

Instantaneous Acceleration

In an analogous technique to calculating instantaneous speed it is also possible to calculate the instantaneous acceleration. Extreme care should be used when analysing this data as it is a second order derivative and as such is derived from data taken from three data points, which leads to a reduced temporal resolution. Wherever possible an accelerometer should be used to gain a more accurate result for acceleration. However, this result can be useful in certain circumstances.

$$Acceleration_t = \frac{Speed(t, t+1)}{Duration(t, t+1)} \; for\ t\ in\ 1, t-2$$

Power Metrics

With the addition of an instantaneous fuel usage record in the data set it is possible to derive multiple useful additional metrics. The following examples are presented mathematically.

Total fuel consumption

$$Fuel\ Consumption(vehicle) = \sum_{t=1}^{t=n} Fuel(t)$$

Total Fuel Efficiency

$$Fuel\ Efficiency = \frac{Fuel\ Consumption}{Total\ Distance}$$

Instantaneous fuel Efficiency

$$Instantaneous\ Fuel\ Efficiency = \frac{Fuel(t)}{Haversine(pos_t, pos_{t+1})}\ for\ t\ in\ 1, t-1$$

3.1.3. Data Analysis

Data analysis is a much more involved process and will generally depend on the type of research questions being asked of the data.

Sanity Checks

One common component of all analysis should be a "sanity check" of the results provided by the data processing. Generally speaking the results from a large group of vehicles will form a statistical distribution and any extreme deviation from this distribution will require a manual inspection to either (i) accept the result as a statistical outlier or (ii) to identify the problem in either of the two previous steps.

As the volume of data increases it may become unfeasible to manually investigate every possible flawed data set. At this point it will likely be valid to simple drop any suspect data point, as the overall data set will be so large that losing any single data point will have no overall effect on the statistics.

Data Analysis Possibilities

Data analysis, as used in previous projects, has been broken into two major components.

1. Descriptive statistics. This is the basic form of the data collected so far and may include such data as total distance driven, average distance driven per vehicle, total time spent charging, etc. It is typically a simple process to extract the statistics, assuming that the data processing has been completed.
2. Analytical statistics. To fully understand the underlying driving behaviour behind electric vehicle uptake and usage, it is necessary to understand those factors that are driving this behaviour. Simple descriptive statistics may not be enough and at this point it is necessary to delve more deeply into the data.

Data Analysis Example

An example of basic data analysis can be shown below. In this graph the average efficiency of an electric vehicle, in terms of gCO_2/km can be seen to systematically vary between summer and winter. Winter, although peaking at approximately the same point as summer, exhibits a greatly increased low efficiency tail.

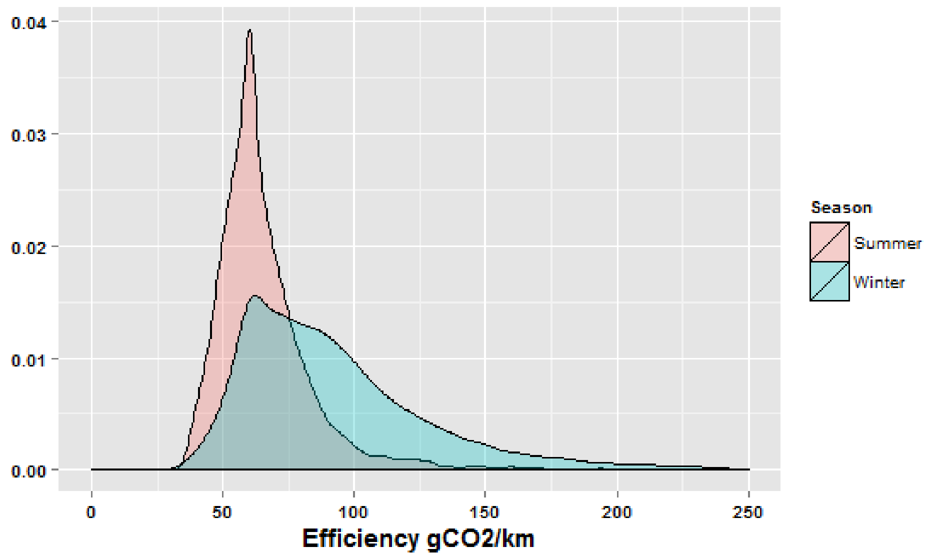

Fig. 3-1. Example of basic data analysis

This example shows how analysis of data can lead to a much greater understanding than simple descriptive statistics. If the data was presented with no further analysis, then it would appear that there was a substantial number of journeys that are less efficient than expected. However, by splitting the data it is possible to see that the low efficiency journeys are, in the main, being driven by those at a colder temperature.

3.2. Managing data quality

Whilst guidance is given in the preceding section for checking data, D3.1 has also presented a list of recommendations to ensure data quality. Within smartCEM it has been agreed that data quality is the responsibility of local pilot site managers. Guidance has been provided by the Evaluation partners in a spreadsheet "smartCEM Quality Data Assurance" Excel file. This file contains four worksheets to be uploaded to the project's shared space as follows:

- Implementation: to be filled only once (unless pending issues exist to be updated)
- Daily (acquisition): calendar table with data to be filled daily
- Weekly (acquisition): calendar table with data to be filled daily
- Period (acquisition): summary of numbers and lessons learned on acquisition

Pilot sites will complete the template as follows:

Pilot Site	smartCEM Data Quality Assurance (Excel template)
Barcelona	All sheets are updated manually once Creafutur has delivered a new archive of historical operational data and this has been processed by IDIADA
Gipuzkoa	Implementation sheet is completed manually; daily and weekly acquisition sheets are also completed manually; period acquisition is completed as required
Newcastle	Implementation sheet is completed manually; daily and weekly acquisition sheets are completed automatically; period acquisition is completed as required
Reggio Emilia	Implementation sheet is completed manually; daily and weekly acquisition sheets are completed automatically; period acquisition is completed as required

Table 3-1. smartCEM data quality assurance method of completion

3.3. *Access to data after smartCEM*

The CDB will continue to exist after the conclusion of the project as a series of SQLite files that will encompass the totality of the data stored by the CDB. Judging from past experiences on previous projects it is expected that the total size of the data set will be in the 1-2 GB range and as such will be easily transferrable.

Multiple options exist to anonymise the data before sharing, which, in addition to possible encryption or redaction should make the data suitable for sharing for a variety of different purposes.

4. Conclusion

This deliverable has presented the following:

- A description of the structure and characteristics of the CDB developed in smartCEM WP3;
- An explanation of the data exchange mechanisms used to transfer the data from the local databases to the CDB;
- An explanation of how data quality is managed by the pilot sites;
- An explanation of how the data analysis partners may extract, process and analyse the data to obtain performance indicators for the analysis and evaluation process;
- Availability of data after the termination of the smartCEM project.

The mechanics of the CDB in terms of data transfer and protocols can be summarised as follows:

- Each site uses its own system, with transferred data stored in a simple csv format;
- Data is transferred daily from each individual site to the CDB through an FTP system;
- The central database stores the data in a PostgreSQL server with the option of exporting data in any needed format;
- Final results are created using SQL queries on the raw transferred data.

D3.3 *Final Operation Report* will present a review of the smartCEM operations phase at each site, and will include final facts and figures relating to the surveys performed. A full description of the data analysis methodology as implemented, along with results of the smartCEM data analysis and evaluation will be presented in D4.5 *Results of the Evaluation*.

5. References

[1] smartCEM, D2.4, Logging tools DB definition, 2013

[2] smartCEM, D3.1, Operational plans for smartCEM platform, 2014

[3] smartCEM, D3.3, Final operation report, 2014 (forthcoming)

[4] smartCEM, D4.3 smartCEM experimental design, 2013

[5] smartCEM, D4.4, smartCEM assessment tools, 2013

[6] smartCEM, D4.5, Results of the evaluation, 2014 (forthcoming)