

Recommendations and future action items on Data Analysis (RT1)

Busan, 25 October 2010



Panellists

- Chairperson: Yvonne Barnard, University of Leeds,
Y.Barnard@its.leeds.ac.uk
- Ilja Radusch, Fraunhofer Institute, FOKUS
ilja.radusch@fokus.fraunhofer.de
- Shane McLaughlin, VTTI
SMcLaughlin@vti.vt.edu
- Farida Saad, INRETS
farida.saad@inrets.fr

Busan, 25 October 2010



Questions

1. Dealing with large data-sets
2. How to analysis video-data efficiently?
3. How to detect driver behaviour issues such as distraction in the FOT data?
4. How to connect data on driving behaviour from the car sensors with attitudinal data from questionnaires, log-books and interviews?
5. What new techniques and tools are available for data analysis?
6. What are the differences between data analyses in naturalistic driving studies and FOT's?

Busan, 25 October 2010



3

How to select what data to analyse if there is a lack of resources to analyse everything?

- Space mission: collect everything you can (difference of opinion) www.fot-net.eu
 - If you collect everything, make sure you have a plan for analyses later on (making data open, data dictionaries etc)
 - Depends on research question
 - Do not collect too much if it is not relevant later on
- Lack of resources for analysing everything, esp. human resources, computational resources no problem
- Delays: data analysis is squeezed
 - Factor in your data-analysis in your project
- Re-using data from other studies may be a good option but difficult:
 - Stakeholders want fresh data, like demonstrations not research
 - How long are data valid?
 - Fundamentals do not change, context does
- Compromise between strict experiment and explorative study
- Need for meta-data base to be able to identify relevant data bases
- Analyse data in layers

Busan, 25 October 2010



4

How to analysis video-data efficiently?

- Video is useful
- Time-consuming
- Knowledge needed to analyse
- Define the interesting events to guide the fragments to look at, start very large, add additional data, focus, and enrich data
- Harder to find patterns than isolated events
- Easy to collect, hard to synchronise and to analyse
- Automated analysis, quick and dirty:
 - Automated computer vision: international project needed
 - Learning mechanisms
 - Hard to identify misses
 - Reverse from manual: determine from video context (e.g. Rural road) and identify events
 - There will be an error rate, statistical analysis
 - Compare with human analysts
- Transfer function: start with sample set of data

Busan, 25 October 2010



5

Use of video recordings : How to analyse video-data efficiently ?

Analyses of video data are :

- Time consuming
- Knowledge and method challenging

Efficient data analyses will depend on :

- The objectives set (identifying the contexts of use of a driver support system; assessing the frequency of “critical interactions” with the support system, with other road users, ...)
- The moment when the data analyses will take place (pilot study, experimental study, ...)
- The data associated with the video-recordings (synchronisation with automated data collection, Drivers' verbal reports , Drivers' diary, ...)
- The means (Human and Technical) available

How to detect driver behaviour issues such as distraction in the FOT data?



www.fot-net.eu

- You need more data than what you measure
- It's more than just numbers
- Subjective data: questionnaires, interviews, diaries etc.
- You need to know what "normal behaviour" is:
 - here is not a model of driver behaviour yet
 - Naturalistic Driving studies may provide information
- Hard to detect in data collected by car, interviews, video etc needed:
 - Drivers avoiding critical situations
 - Anticipation behaviour
- Combine different performance indicators
- Patterns identification, machine learning techniques

Busan, 25 October 2010



7

Detecting Behavior Issues/Distracted

- Identify behavior/distracted detection methods that are independent of study measures
- Methods:
 - Algorithms to evaluate vehicle measures (e.g., steering wheel)
 - Eye glance – real-time or post-processed
 - Human review of video
 - Data mine vehicle network messages for interaction with in-vehicle systems
 - GPS and road network maps for driving related distraction



How to connect data on driving behaviour from the car sensors with attitudinal data from questionnaires, log-books and interviews?



www.fot-net.eu

- Diaries can be connected with automatic data, because of time-stamps
- Design questionnaires to be integrated with behaviour data, specific to functions
- Ask for events in questionnaires, interviews etc.
- Use of questionnaires:
 - Selection of drivers based on questionnaires
 - Identify groups of drivers
 - Select interesting drivers for focus groups/interviews

Busan, 25 October 2010



9

What new techniques and tools are available for data analysis?



www.fot-net.eu

- Data-mining
- Automated video-analysis
- Data-stream management
- Machine-learning techniques
- Simulation for pre-analysis
- Using new techniques for (re-)using old data
- Need to look at data before and after events: continuous data recording but high resolution storage around events
- International project bringing together new techniques
- Standardisation:
 - Data-exchange standardisation
 - Standardisation of event definition
 - Standardisation of eye-movements measurements (ISO group on this)
 - Semantic structures, annotations, ontologies, data dictionaries
- Different techniques answer different questions (detailed studies needed for particular questions)

Busan, 25 October 2010



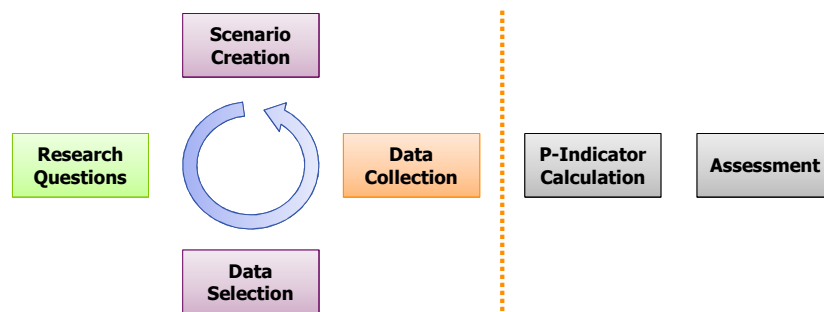
10

What are the differences between data analyses in naturalistic driving studies and FOT's?

- More noise in ND
- ND explorative, finding new hypotheses
- In FOT you can ask people to go to some place, like for cooperative systems, but FOTs may be more naturalistic, some FOTs are more controlled experiments, (different opinions)
- Data collection related to purpose
- Iteration needed in both types between data analysis and formulation of hypotheses/research questions

Outlook

- Why not use scenario specification in a posteriori data selection process?



Other issues



www.fot-net.eu

- Have specific questions to guide analysis
- Hard to answer impact questions because of noise
- Bias in data:
 - How to correct for bias? (drivers included in FOT are favourable to system) How representative is the sample?
 - Re-use of data may give information on bias
 - Using raw data without having a complete picture may miss bias
 - On the other hand high level data may have more data
 - Iterate between the two

Busan, 25 October 2010



13

Thank you for your attention!



www.fot-net.eu

More information or want to cooperate?

www.fot-net.eu

Busan, 25 October 2010



14