

**EUROPEAN COMMISSION
DG CONNECT**

**SEVENTH FRAMEWORK PROGRAMME
INFORMATION AND COMMUNICATION TECHNOLOGIES
COORDINATION AND SUPPORT ACTION**

FOT-Net Data

FIELD OPERATIONAL TEST NETWORKING AND DATA SHARING SUPPORT



D5.3 FINAL REPORT ON WORKSHOPS

| | |
|---------------------------------|--|
| Deliverable no. | D5.3 |
| Dissemination level | Public |
| Work Package no. | WP5 |
| Main author(s) | Yvonne Barnard, Haibo Chen, Tony Jia (University of Leeds) |
| Co-author(s) | Helena Gellerman (SAFER), Satu Innamaa and Sami Koskinen (VTT) |
| Status (F: final, D: draft) | F |
| Version number | 1.0 |
| Date | 1 February 2017 |
| Project Start Date and Duration | January 2014, 36 months |



Document Control Sheet

Main author(s) or editor(s): Haibo Chen (University of Leeds), Yvonne Barnard (ERTICO)

Work area: WP5

Document title: D5.3 Final Report on Workshops

Version history:

| Version number | Date | Main author | Summary of changes |
|----------------|----------|---|--------------------|
| 1.0 | 1.2.2017 | Haibo Chen, Yvonne Barnard, Tony Jia (UNIVLEEDS) | 1.0 |
| | | | |

Review:

| Name | Date |
|--------------|-----------|
| Satu Innamaa | 31.1.2017 |

Circulation:

| Recipient | Date of submission |
|-------------------------|--------------------|
| EC | 1.2.2017 |
| FOT-Net Data consortium | 1.2.2017 |

Table of Contents

| | |
|---|----|
| Table of Contents | 3 |
| 1 Introduction | 4 |
| 2 The FOT-Net Workshops | 6 |
| 3 FOT-Net Data Workshop 1: FOT Data Re-use | 10 |
| 4 FOT-Net Data Workshop 2: FOT Data Anonymisation | 19 |
| 5 FOT-Net Data Workshop 3: A common methodology for road automation FOTs and pilots | 29 |
| 6 FOT-Net Data Workshop Extra: FOT Data Anonymization and feature extraction | 41 |
| 7 FOT-Net Data Workshop 4: Hands-on workshop on Data Re-use | 51 |
| 8 Hands-on Workshop on Data Re-Use, Guangzhou Session in China | 59 |

1 Introduction

1.1 FOT-Net Data project

FOT-Net Data is a Coordination and Support Action in the EU 7th Framework Programme for Research, submitted for the call FP7-ICT-2013-10. It stands for *Field Operational Test Networking and Data Sharing Support*.

The duration of the project is 36 months, effective from 1 January 2014 until 31 December 2016. The project is funded by the European Commission (EC) under Grant Agreement number 610453. The EC Project Officer is Ms. Myriam Coulon-Cantuer from Directorate General for Communications Networks, Content & Technology (DG CONNECT).

The project partners are VTT Technical Research Centre of Finland, as the coordinator, ERTICO – ITS Europe, SAFER Vehicle and Traffic Safety Centre at Chalmers University of Technology, Institut für Kraftfahrzeuge (ika) at RWTH Aachen University, Galician Automotive Technology Centre CTAG, University of Leeds, the European centre of studies on safety and risk analysis CEESAR and the automotive company Daimler. The project coordinator is Dr. Sami Koskinen.

FOT-Net is a networking platform established in 2008 to allow individual FOTs to benefit from each other's' experiences as well as giving a better overview of the scattered activities.

In all external communication to FOT-Net refers to a networking platform open to all stakeholders interested in FOTs, established in 2008. FOT-Net 1, FOT-Net 2 and FOT-Net Data are support actions, which have been the funded instruments enabling to operate the network. In all external communication, the activities will continue to be referred as FOT-Net in order to show continuity.

The main objectives of FOT-Net Data are to:

- Support efficient sharing and re-use of FOT datasets
- Develop and promote a framework for sharing data
- Build a detailed catalogue of available data and tools
- Operate an international networking platform for FOT activities.

1.2 FOT-Net Data WP5 – FOT Methodology

The objective of FOT-Net Data WP5 is to provide the FOT community with knowledge on data sharing:

- To include methods of data-sharing and data re-use in the FOT methodology based on experience and knowledge gained during the project and to keep the FESTA handbook up-to-date;
- To transfer knowledge about the FOT methodology, by organising webinars and making the presentations from these webinars available in a format that allows independent e-learning for newcomers; and
- To share experiences and gain knowledge about data-sharing by organising workshops with a hands-on character.

WP5 is composed of three tasks:

- Task 5.1 updates the FOT methodology with data-sharing processes FOT data sharing methodology and maintains the FESTA handbook. The results from this task are reported in Deliverable 5.1 “Identification of places in the FOT methodology where data-sharing should be integrated and templates for gathering information”.
- Task 5.2 organises webinars on the FOT methodology. The results from this task are documented in D5.2 “Six FOT methodology e-learning modules on-line”.
- Task 5.3 organises hands-on workshops on data-sharing, as described in this deliverable.

Task 5.3 generates knowledge and experiences on data sharing in the form of workshops in which people can actively work on problems and discussion items, share their expertise and experiences and learn from each other in order to come up with practical solutions. The workshops have a hands-on character and provided examples of datasets (or demonstration-sets) to be worked with in order to get a better understanding of the issues that play a role in re-using and sharing data and to come up with potential solutions. These workshops opened to the FOT-Net community and were publicised via the website, mailing lists and other publicity channels used in the project.

2 The FOT-Net Workshops

2.1 Summary of the workshops

The following workshops were organised:

Workshop 1: FOT Data Re-use In collaboration with EUDAT/BSC

This workshop was hosted in collaboration with EUDAT (<http://www.eudat.eu/>) and Barcelona Supercomputing Centre (BSC).

Barcelona, 16–17 December 2014

Topics:

- Datasets created in the project:
 - Data collection
 - Data pre-processing and quality assurance
 - How to make the data re-usable?
- Datasets collected in other projects and re-used in the project:
 - In which project (s) was the original data collected?
 - What kind of process did you adopt in order to re-use the collected data?
 - Overview of the results from re-using the collected data
 - What lessons and experience did you gain from this process?
- What is needed in order to make data easier to re-use?

Workshop 2: FOT Data Anonymisation

Gothenburg (Sweden), 1–2 September 2015

Topics:

- Vision-based systems for driver monitoring and de-identification
- Anonymising moving objects in the vehicle surroundings
- Using FACS-coder to anonymise drivers in eye-tracker videos
- DCode and DMask: Two approaches for video data anonymisation
- Open discussion on issues presented in the previous cases, Chaired by Marco Dozza
- Anonymisation: Balancing Re-use and Privacy – the US Perspective
- Trajectory Privacy: Measures and Preservation Methods
- Managing privacy risks in autonomous driving and cooperative intelligent transportation systems

Workshop 3: A common methodology for road automation FOTs and pilots

Leeds (UK), 3–4 February 2016

Topics:

- The FESTA methodology
- Automation FOTs and pilots in Europe
- USDOT Programs for Collecting Connected & Autonomous Vehicle Data
- Societal scenarios and research questions
- Study design

- Data collection and analysis
- Impact and socio-economic cost-benefit analysis
- Data sharing

Extra workshop: FOT Data Anonymization and feature extraction

Gothenburg (Sweden), 31 August – 1 September 2016

Topics:

- Importance of real world data for safety systems development
- Main issues with anonymization and feature extraction
- US perspective on anonymization
- Different privacy for data mining and querying
- Computer resources for large-scale anonymization / feature extraction
- Video anonymization of vehicle environment
- Feature extraction – how to assure you can delete data without information loss
- Using automated video annotation for feature extraction
- Automated labelling and recognition

Workshop 4: Hands-on workshop on data re-use

Several locations and online, 14 September 2016

Topics:

- US Research Data Exchange (RDE)
- US and European data sets and meta-data
- Research questions data sets could answer
- Data analysis methods

In addition to this workshop, a workshop along the same lines was held in Guangzhou (China) on 12 December 2016.

Below the number of participants is given. On average around a third of the participants were female.



2.2 Organisation of the workshops

The workshops were organised in order to transfer knowledge on data re-use and sharing in FOTs. The workshops were organised in a similar way as the seminars in the previous FOT-Net 2 project, because this had proved to be a successful formula.

Deliverable D4.1 of FOT-Net 2, entitled “FOT seminar planning”, defined detailed plans and checklists for such disseminations. This deliverable served as the guideline for organising the workshops.

All workshops were announced on the FOT-Net website (<http://fot-net.eu/>), in the FOT-Net Data newsletter and by direct mail to the FOT-Net mailing list. Several announcements and updates of agendas were made for each workshop. Participants could register electronically. After the workshops reports and all presentations were made available on the FOT-Net website, and short reports were published in the newsletters.

In the workshops there usually was a mix of presentations and interactive sessions. All workshops had activities in small groups in which participants were presented with datasets and problems they had to solve, for example defining research questions for re-using the collected datasets or discussing how data sharing could be supported. The results were reported and discussed in plenary sessions. Also interactive discussions, question sessions and panel sessions were part of the agendas. In this way we ensured that the workshops answered the goals of both knowledge transfer and exchange of knowledge and experiences. Most speakers and facilitators came from the European and national FOTs, with different backgrounds, coming from industry, public authorities, consultancy, and research institutes and universities. Also speakers and participants from the US joined all of the workshops. The attendees included both experts and novices in FOTs and related areas. Many attendees were involved in a FOT themselves.

Participants were asked to complete an evaluation form after each workshop. In general they gave very positive feedback; the appreciation of sessions was mostly “very good” and “excellent”. Most of the participants answered the question on whether the information was useful to them with “yes”. During the workshops most participants showed a very active attitude, engaging enthusiastically in the interactive sessions.

2.3 The “blended” workshop method

For the last workshop “Hands-on workshop on data re-use”, we adopted a new approach to the workshops. The workshop was organised at three different locations where small groups came together to work hands-on with data sets and discuss the re-use of data. In addition, participants could join remotely. During the day several plenary sessions with speakers were held in the form of webinars where all participants came together. For more details see the workshop report in Section 5. Originally the idea was to have some remote participants at all locations. However, as registrations showed a rather large number of people who wanted to join remotely, it was decided to have a separate group working completely online. In one of the locations two remote participants were expected, however, only one eventually joined. For this workshop it turned out that although originally 47 people registered, only 27 participated the whole day. Registering for an online event and not in effect joining may be relatively easy.

Although there were some concerns about logistics, the workshop went smoothly. The plenary webinars were moderated by ERTICO in Brussels, the chair of the day attended the

workshop in Aachen, the speakers were at the different locations, the US speaker delivered a pre-recorded presentation for the morning session (as the time-difference was too large) and joined live in the afternoon. The online group worked together but at times participants performed activities on their own. Before the workshop, but also during the day, remote participants were sent information and reminders on how to join the GoToWebinar and GooMeeting sessions. This software worked well and no technical problems were encountered. At the three locations video was also used.

The workshop was successful in reaching its goals, and reactions from participants were positive. Due to the plenary sessions, participants felt they were part of a common event. The one participant who joined the group in Aachen was also very positive and felt included in the process. In the local groups a new audience was reached, such as research students, who would not normally be able to travel to plenary workshops in other locations, and participating remotely was an important advantage for participants from the US. The nature of the workshop, working hands-on with data in small groups, provided insights in the opportunities (and difficulties) of re-using data and ensured the active engagement of participants. We concluded that this blended form of workshop is worthwhile to pursue in the future.

3 FOT-Net Data Workshop 1: FOT Data Re-use

On 16–17 December 2014, a workshop was held in Barcelona (Spain), in collaboration with EUDAT (European Data Infrastructure, <http://www.eudat.eu/>) and BSC (Barcelona Supercomputing Centre, <http://www.bsc.es/>), to address how FOT data collected in previous projects can be re-used. The agenda was as follows:

Agenda

Tuesday, 16th December 2014, 12:00 - 18:00

1. Overview of this workshop, Haibo Chen (University of Leeds)
2. FOT-Net Data introduction, Sami Koskinen (VTT)
3. Introduction to the current status of FESTA handbook, Yvonne Barnard (ERTICO)
4. Introduction on WP5 – FOT Methodology, Haibo Chen (University of Leeds)
5. What are the main issues with data re-use?, Adrian Zlocki (IKA)
6. DRIVE C2X: introduction and its approaches to data re-use:
 - a. The analyst's view, Satu Innamaa (VTT)
 - b. The test site's view, Marta Miranda (CTAG)
 - c. French re-use case, Laurette Guyonvarch (LAB)
7. EUDAT: project introduction and its approaches to data re-use, Damien Lecarpentier (CSC)
8. Compass4D / SISCOGA: introduction and its approaches to data re-use, Moises Martínez (CTAG)
9. Data re-use exercise, Moderator: Haibo Chen (University of Leeds):
 - a. Find errors
 - b. Examine the SPMD datasets (USA)
 - c. Suggest potential use of the SPMD data
 - d. Compare DRIVE C2X (EU) with SPMD
10. Reporting on exercise results and open discussion on how easy data can be re-used, moderated by Clement Val (Ceesar)
 - a. What are the methods and tools to be used for the quality assurance of a given dataset?
 - b. What are the best practices for labelling attributes?

- c. What applications would you like to use the SPMD and DRIVE C2X data and how easy could they be re-used? All

Wednesday, 17th December 2014, 09:00 – 14:00

11. Recap of Day-1 outcomes, Adrian Zlocki (IKA)
12. Data Re-use: US Perspective, Ram Kandarpa (Booz Allen Hamilton)
13. Introduction to BSC, David Vicente (BSC)
14. Guided visit to BSC
15. TRAFISAFE: introduction and its approaches to data re-use, Sami Koskinen (VTT)
16. smartCEM: introduction and its approaches to data re-use, Josep Laborda (RACC)
17. PhD projects, Joao Martins (Empa)
18. Round-table discussion, Moderator: Yvonne Barnard (ERTICO)
 - a. Whether there are data re-use challenges that apply in common across these different FOTs
 - b. Whether there are data re-use solutions that apply in common across these projects
19. Conclusions and Wrap-up, Haibo Chen (University of Leeds)

There were 18 workshop participants (the list of attendees can be found at the end of this document).

1. Overview of this workshop, Haibo Chen (University of Leeds)

Haibo welcomed delegates and outlined the programme of the seminar with a brief introduction of the objectives and the topics to be covered by the invited speakers.

2. FOT-Net Data introduction, Sami Koskinen (VTT)

With reference to FOTs, Sami informed that over the past six years, the EU has supported a number of large-scale field tests of latest vehicle information technology in which thousands of drivers have been able to test the most promising prototypes and products just entering the markets. The results from these projects have been used to evaluate the societal impacts of these technologies and contributed to their introduction.

Sami reiterated that FOT-Net is a networking platform open to all stakeholders interested in FOTs to let them benefit from each other's experiences as well as to give an international dimension to local activities.

Sami then introduced the FOT-Net Data project which appreciates the Commission's recognition of the importance of making the collected data widely available to the research community and explores the potential for reusing the data in new studies. Sami highlighted the project objectives, data sharing benefits, the project structure, the partner network and the FOT-Net Wiki which is a free living encyclopaedia of FOTs that anyone in the FOT community can edit (wiki.fot-net.eu).

3. Introduction to the current status of FESTA handbook, Yvonne Barnard (ERTICO)

Yvonne gave an overview of the current status of the FESTA handbook by explaining why a methodology is needed for field operational tests of vehicle information technologies, how it works and what steps need to be followed. Yvonne highlighted that the FOT-Net data project will make specific contributions to data sharing and re-use which will be used to update the FOT methodology later in the project.

4. Introduction on WP5 – FOT Methodology, Haibo Chen (University of Leeds)

Following the previous presentation, Haibo explained that the objective of WP5 ('FOT methodology') in the FOT-Net Data project is to provide the FOT community with knowledge on data usage via enhancement and update of the FOT methodology; knowledge transfer of the FOT methodology (e.g. Webinars); and sharing experiences of data usage (e.g. Workshops).

5. What are the main issues with data re-use? Adrian Zlocki (IKA)

Adrian explained that FOT-Net Data is developing a common description format for the data. He informed that a FOT-Net data catalogue is planned to be issued before the summer of 2015. Datasets and their metadata are described in this catalogue. Contact information will be provided, so interested people can use this to negotiate access to the datasets.

In the US there is the RDE website (www.rde-its.net) where people who are registered can directly access data. The way data access will be handled in the EU will evolve over time, similar to accident data.

6. DRIVE C2X (www.drive-c2x.eu): introduction and its approaches to data re-use:

a. The analyst's view, Satu Innamaa (VTT)

Satu explained that a schema for good metadata is needed. One aim is to provide a minimal set of content. There are international standards, but they have to be mapped and the current practise provides room for improvement.

b. The test site's view, Marta Miranda (CTAG)

Marta introduced the DRIVE C2X test sites, in particular the Galician Intelligent Corridor which is 90km long consisting of A55, A52 and AP9 motorways, and densely instrumented with RSUs, cameras, VMSs, and high precision methodological stations. Marta showed how the events data are transmitted, processed, analysed and archived. She also explained how data re-use was addressed in OpEneR (Optimal Energy consumption and Recovery based on a system network).

c. French case, Laurette Guyonvarch (LAB)

Laurette gave a presentation of a French data re-use case by starting with "Find the data". She highlighted a number of practical issues when dealing with other people's data including missing records, low quality, synchronisation, baseline settings, and so on.

Laurette provided recommendations for data re-use during the experiment, at the end of the recordings, and when re-using data, and above all she recommended to be patient!

7. EUDAT (www.eudat.eu): project introduction and its approaches to data re-use, Damien Lecarpentier (CSC)

Damien explained that EUDAT is a pan-EU e-infrastructure solution for data challenges. It offers common data services, supporting 30+ research communities as well as individuals, through a geographically distributed, resilient network connecting general purpose data centres and community-specific data repositories. The infrastructure provides users with a number of powerful CDI services/tools including:

- B2DROP: secure and trusted data exchange service for researchers and scientists
- B2SHARE: for storing and sharing research data
- B2SAFE: for replicating large data safely
- B2STAGE: to staging large data sets
- B2FIND: to find data

Damien told that there are several areas where EUDAT can support FOT-Net including:

- Storage capacities located at selected centres in Europe to keep pace with the accelerated generation of data (B2AFE - based on clear service offer and SLAs; calls for collaboration projects in 2015);
- Open Data Sharing platform tailored for FOT-Net stakeholders (researchers, citizen scientists, etc.) and FOT data sets via B2SHARE;
- Dissemination and better discoverability and reusability of FOT data sets through specific solutions to access the data and metadata catalogue (e.g. B2FIND for FOT-Net data hosted inside and outside EUDAT)

8. Compass4D / SISCOGA: introduction and its approaches to data re-use, Moises Martínez (CTAG)

Moises provided an overview of the VIGO text site in Spain, and how the SISCOGA data was re-used to test Compass4D apps. He explained that the projects collected CAN, GPS, traffic light, road events, and TMC data which were reused for multiple apps including:

- Traffic management, traffic signs position, signal retiming, find congestion, traffic sensor positioning, and weather impact
- Transit routes improvement, hybrid bus benefits, and transit impacts in traffic management

9. Data re-use exercise, Moderator: Haibo Chen (University of Leeds):

a. Find errors

In this exercise, the participants were given an on-page dataset which consists of 1 Hz exhaust CO2 emissions and GPS data from an instrumented car. There were 11 errors in the data records and Haibo confirmed that they were related to real errors:

- 1) Incorrect date format: 02/15/2005
- 2) Invalid time format: 07:00:60
- 3) Invalid time format: 07:01:61
- 4) Invalid CO2 value: -12.924
- 5) Invalid velocity value: -1.3
- 6) Invalid latitude: 0.0000
- 7) Invalid longitude: 1.70743
- 8) Invalid GPS-velocity value: 999
- 9) Invalid number of satellites: N

- 10) Missing 1 second: 07:00:08
- 11) Missing 10 seconds: 07:00:31 - 07:00:40

The results showed that only one participant found all the errors.

b. Examine the SPMD datasets (USA)

In this exercise, the Safety Pilot Model Deployment One Day Sample Metadata descriptions were provided to participants with no prior knowledge to examine. It was found that several names are not comprehensible for people with a European background.

c. Suggest potential use of the SPMD data

It was suggested to show speed-limit/-change information on a map. It depends on creativity. What is the audience? The audience consists mainly of analysts, but datasets may be useful for Masters and PhD students. They can use modified (faked/anonymised) datasets to learn statistical tests, for example.

TRAFISAFE includes trip and event data and it needs to be checked whether travel diaries look strange, e.g. when the start and the end in controlled trials are the same.

d. Compare DRIVE C2X (EU) with SPMD

Both projects include more or less the same variables. It takes time to comprehend data.

The DRIVE C2X documentation is good, but can it also include references to other deliverables? Not every deliverable has the same number of cameras, driver characteristics, etc. It is important to keep track of these changes. Stand-alone documentation could be the solution and FOT-Net should stimulate this solution. However, a common problem is the lack of time and resources; at the end of a project the budget runs out and people leave without having time left to provide information.

10. Reporting on exercise results and open discussion on how easy data can be re-used, moderated by Clement Val (Ceasar)

Clement led a discussion on the following issues:

- a. What are the methods and tools to be used for the quality assurance of a given dataset?
 - There is almost no way to only rely on the human eye for checking all data, for example to see if there is a count < 0
 - It is recommended to include in the metadata file descriptive statistics of the data such as the mean, media and mode. This can help find missing values, which should also be described by metadata. Furthermore, outliers as well as coherence could be visualised automatically.
 - Re-use can uncover errors and implausibility, so possibly open a Pandora's box. Hopefully any such errors have not been influencing the analysis in the original project.
 - There may be implausible differences between measures, such as speed and yaw rate.

- It was agreed that novelty detection comprises a lot of work. The analysis is the faster part.
 - In EuroFOT a monitoring tool was used that sends emails when data is no longer logged from a certain vehicle.
 - When one goes from the tool to the method level, how would these situations be documented?
- b. -What are the best practices for labelling attributes?
- Long, describing names that are still short enough to be used in queries. They should imply what is measured and which sensors are used. One dataset uses '_', others join words by capitalising the first letters of words. It is important to note that some tools are unable to recognise certain conventions, but they may still benefit the researcher, if such a way of writing is permitted.
- c. What applications would you like to use the SPMD and DRIVE C2X data and how easy could they be re-used? (Confirmed by Ram, potential applications of the SPMD data will be presented and discussed in his presentation on the second day. Haibo commented that both datasets could be used to validate and update the road design standards set in the Design Manual for Roads and Bridges (DMRB) in which many studies were carried out more than 30 years ago.)
- Examples of research questions are: What are the user responses to external road signs? Could we get rid of all signs? Can we improve the road infrastructure based on the data? Can better infrastructure improve behaviour?
 - The SPMD data are more interesting, because they are naturalistic, rich and therefore better for re-use. However, there are benefits to both sets. Controlled studies can provide rich data and then naturalistic studies can complement the results.
 - One can compare car populations between the EU and US.
 - DRIVE C2X data can be useful for individual variability in reactions, extremes and boundaries.
 - One may extrapolate results with other datasets and make predictions, e.g. "We have seen people do ... in this situation". One can look at an average or range.

Wednesday, 17th December 2014, 09:00 – 14:00

11. Recap of Day-1 outcomes, Adrian Zlocki (IKA)

Adrian started the second day session by summarising the key issues found in the first day session, highlighting the difficulties of re-using data from previous projects but reiterating that more and more people are re-using these datasets.

12. Data Re-use: US Perspective, Ram Kandarpa (Booz Allen Hamilton)

Ram provided an overview of the Research Data Exchange (RDE) system in the US which promotes sharing of both archived and real-time data from multiple sources and modes (www.rde-its.net). Ram introduced structured approaches to facilitating data re-use. The presentation also included a number of suggestions for motivating data re-users, e.g.:

- Can multi-modal data be fused and used for traveller information and systems management?
- How can “big data” analytic techniques be applied to transportation data to improve transportation planning, operations and management?
- How can connected vehicle data be used for better measurements of queue lengths and bottleneck detection?
- How can connected vehicle data improve commercial and transit operations?
- How can connected vehicle provide better measures of travel time reliability and highway and intersection performance?
- Can temporal analysis of vehicle operation measures inform variety of network management policies ranging from HOT lanes to dynamic speed limits?
- Can temporal analysis of acceleration metrics uncover insights into driver behaviour versus dynamic network / environmental features such as adverse weather?

Ram also described how the RDE system measures the extent of re-use, and solicits feedback from users to enable continuous improvement.

The presentation stimulated a lot of comments, questions and discussions, e.g.:

- Data can be used to see how drivers navigate junctions.
- Contextual data in the US is situational data in Europe.
- In the US there is a University Transportation Centre programme, working with Universities and doing outreach activities. Students, even undergraduates, should use such data.
- Can people re-use data for any purpose? It depends on the permission they are granted. Usually data is free for research purposes, but there are restrictions for commercial use.
- It is important to disconnect data from drivers, but it might still be possible to identify them, e.g. with origin and destination information.
- There may be policies in place that restricts data use after the download, e.g. in the UK data need to be kept on protected networks.
- Asking data users for their wishes can make projects very expensive.
- In the EU data collection is more scattered, whereas in the US it is more systematic.
- Outreach can be done at conferences, for example, which may trigger a snowball effect.
- Students can work with data during their modules or in summer schools, where they can work on small research questions and basic statistics. FOT-Net Data should investigate this.
- Guidelines for logging need to be found, handling questions such as how deeply we can look at data and data protection.

- Capture, outreach and sharing are all interrelated. Guidance needs to be developed for all three areas. FESTA has dealt with the capture part. Guidance needs to be put into practise, for example by making data sharing a requirement for all FOTs.
- Outreach needs to be targeted beyond the transport field, e.g. computing. Different people can use their knowledge to do different things with the data. Outreach is an important objective of FOT-Net Data.

13. Introduction to BSC, David Vicente (BSC)

David presented an overview of the BSC facility and its services, databases and users. David explained how the advantage of BSC is the cost efficiency when several elements such as staff and room are provided by one organisation. He also indicated that the cost of using the BSC is free for research purposes.

14. Guided visit to BSC (www.bsc.es)

15. TRAFISAFE: introduction and its approaches to data re-use, Sami Koskinen (VTT)

Sami introduced the TRAFISAFE project and focused on driving style feedback for novice drivers. Sami indicated that it can be possible to standardise feedback using the study's results, which can also be handed to national agencies. This way they can be used in driving schools and other communication aimed at young drivers, for example with statistics. Data from an earlier VTT project, Tele-ISA, was used as a reference. Data from TRAFISAFE are available for re-use under a non-disclosure agreement.

16. smartCEM: introduction and its approaches to data re-use, Josep Laborda (RACC)

Josep described the data analysis approaches used in smart Connected ElectroMobility (smartCEM). Josep explained that smartCEM has amended the FESTA Methodology as an evaluation framework as smartCEM not an FOT. Josep commented that it is necessary to add some points to FESTA, contextual data are very important, and data of this project are available.

17. PhD student presentation, Joao Martins (Empa)

Joao described his PhD study which is concerned with the derivation of current braking forces load model for the assessment of road bridges. He explained that sources of randomness in braking events are determined by traffic characteristics and driver behaviour, which is why he needed FOT data.

Joao described his experience in gaining access to FOT data and how difficult to understand the collected data without proper documentations.

18. Around-table discussion, Moderator: Yvonne Barnard (ERTICO)

19. Conclusions and Wrap-up, Haibo Chen (University of Leeds)

Due to time limit, Yvonne and Haibo joined the sessions and focused on capturing key pits from the two-day presentations and discussions.

- Whether there are data re-use challenges that apply in common across these different FOTs
- Whether there are data re-use solutions that apply in common across these projects

The discussion recognised that much project budget is spent on data collection, field tests and data selection but not always enough time and budget is left to address research questions set in the beginning of the project. Very little effort can be given to meta-data documentation and hence data re-use is not easy.

Finally, Haibo closed proceedings by informing that there will be 3 webinars and another workshop to be held in 2015 covering other 3 FESTA areas and a focus on data sharing.

Participants FOT-Net workshop, 16–17 December 2014, Barcelona

| Last Name | First Name | Company |
|----------------------|------------|---------------------|
| Barnard | Yvonne | ERTICO |
| Pampel | Sanna | University of Leeds |
| Guyonvarch | Laurette | LAB |
| Chen | Haibo | University of Leeds |
| Gonzalez Villafranca | Sara | IREC |
| Martins | Joao | Empa |
| romero | Josep | appintelligent |
| Koskinen | Sami | VTT |
| Casas Vilaro | Jordi | TSS |
| Innamaa | Satu | VTT |
| Kandarpa | Ram | Booz Allen Hamilton |
| Krishnakumar | Reakka | CEESAR |
| Val | Clément | CEESAR |
| Zlocki | Adrian | ika |
| Rial Martínez | Moisés | CTAG |
| Miranda | Marta | CTAG |
| ALattas | ALaa | upf |
| Perarnau | Josep | Aimsun |

4 FOT-Net Data Workshop 2: FOT Data Anonymisation

Agenda

On 1–2 September, a workshop was held Gothenburg (Sweden) to address how FOT data can be anonymised. The agenda was as follows:

Tuesday, 1 September 2015, 13:00 – 18:00

1. Overview of the workshop, Haibo Chen (University of Leeds)
2. FOT-Net Data introduction, Sami Koskinen (VTT)
3. Current status of FESTA handbook, Yvonne Barnard (University of Leeds)
4. Knowledge exchange activities related to FOTs, Haibo Chen (University of Leeds)
5. “What are the main issues with data anonymisation?”, Helena Gellerman (SAFER)
6. Vision-based systems for driver monitoring and de-identification, Fernando de le Torre (Carnegie Mellon University, USA)
7. Anonymising moving objects in the vehicle surroundings, Clement Val (CEESAR)
8. Using FACS-coder to anonymise drivers in eye-tracker videos, Selpi (Chalmers/SAFER) and Torsten Wilhelm (Smart Eye)
9. DCode and DMask: Two approaches for video data anonymisation, Amir Tamrakar (SRI International, USA)
10. Open discussion on issues presented in the previous cases, Chaired by Marco Dozza (Chalmers/SAFER)

Wednesday, 2 September 2015, 09:00 – 12:30

11. Recap of Day-1 outcomes, Satu Innamaa (VTT)
12. Anonymisation: Balancing Re-use and Privacy – the US Perspective, Dwayne Henclewood (BAH, USA)
13. Trajectory Privacy: Measures and Preservation Methods, Gyözö Gidofalvi (KTH)
14. Managing privacy risks in autonomous driving and cooperative intelligent transportation systems, Henrik Broberg (VCC)
15. Open discussion on issues presented in the previous cases and issues in a broader context (e.g. accident data, surveys etc.), chaired by Niels Agerholm (Aalborg University)
16. Wrap-up, Yvonne Barnard (University of Leeds)

There were 50 workshop participants (the list of attendees can be found at the end of this document).

1. Overview of this workshop, Haibo Chen (University of Leeds)

Haibo Chen welcomed the participants and outlined the programme of the workshop.

2. FOT-Net Data introduction, Sami Koskinen (VTT)

Sami Koskinen introduced the FOT-Net Data project behind the workshop. He commented that anonymisation is a rising topic and is included in the EU's ITS directive. It forms part of the requirements for new products and projects. In H2020 there are plans to arrange FOTs on connected systems (C-ITS) and automated driving. Data sharing will be an important part of those projects.

FOT-Net has the first version of Data Sharing Framework available for comments, divided into three documents. The FOT Data Catalogue is also ready and we are inviting data owners to create entries in the catalogue: wiki.fot-net.eu.

3. Current status of FESTA handbook, Yvonne Barnard (University of Leeds)

Yvonne Barnard presented the FESTA methodology, introducing the need for a common methodology and going through the methodology steps in the so-called FESTA-V. She presented FESTA texts regarding anonymisation. FESTA recognises and emphasises the need for anonymisation of personal data and points out the need to describe the anonymisation process, but does not offer much technical advice. FESTA considers that complete anonymity is not necessarily feasible for all FOT data. It also raises the topic of the difficulty of anonymisation of video data - being a central topic of the day. The latest version of the handbook may be found at: <http://fot-net.eu/Documents/festa-handbook-version-5-2014/>. The handbook will be updated at the end of 2016, and the results of the workshop and on-going work on anonymisation will be included.

4. Knowledge exchange activities related to FOTs, Haibo Chen (University of Leeds)

Haibo Chen presented briefly the knowledge exchange activities of FOT-Net: webinars and workshops. The webinars address the FESTA handbook content and recordings and materials may be found on-line. Presentations and the report from the 2014 workshop on data re-use are also available at fot-net.eu/library.

5. "What are the main issues with data anonymisation?" Helena Gellerman (SAFER)

Helena Gellerman gave a presentation on main issues with data anonymisation. Privacy is nothing new. Previously data anonymisation has been much about removing names and addresses. The last years with the internet search and map capabilities there is a growing need for stronger anonymisation. Otherwise combination of data can give out personal details.

She gave examples that in accident data company names, road signs, faces of emergency personnel etc. are anonymised. However, where accident data used still images, naturalistic driving study data has continuous video. There is also a general trend towards open data and re-use of data. Anonymisation is becoming more and more important because of legal requirements on privacy, participants that have only given consent to share anonymous data and requirements in funding and consortium agreements.

One of the main questions is how to keep the original richness of the data; anonymising data while keeping essential information. Helena discussed especially key information recorded in video, both internal (of the driver) and external (of other road-users and environment). The other topic to be discussed at this workshop is GPS data. Topics that need to be addressed in the future are e.g. low-cost, real-time feature extraction from video and position privacy in C-ITS and automation.

6. Vision-based systems for driver monitoring and de-identification, Fernando de le Torre (Carnegie Mellon University, USA)

Fernando de le Torre from the Human Sensing Lab at Carnegie Mellon Robotics Institute presented a gaze tracking system for detecting drivers' distraction. IntraFace software will be released in October. It can detect facial features emotions and show them on an avatar face.

He demonstrated the effects that common anonymisation options such as blurring, pixelation, bar mask and negative colouring have to images. He presented the project on de-identification in the automobile environment, which uses a new approach: remove person-specific facial features, while preserving head pose, gaze and facial expression, and transfer it to the face of someone else. He gave a demonstration using president Obama as the masked face of a driver on a video. He showed results on SHRP2 video data, anonymised using an image of Helena Gellerman and discussed difficulties due to low resolution. Future development addresses e.g. hair and clothes. Other challenging situations can be caused by noisy camera images during the night or by large changes in the driver's pose.

Fernando presented Facial Action Unit Coding for representing expressions, e.g. upper lip raised being one "action unit". They have created a FAT (Facial Action Transfer) technology for processing videos. He showed competitive results against other known techniques. (humansensing.cs.cmu.edu)

In the discussion the problem was raised that to prove in court that a face has been anonymised, 100% anonymisation is needed. That's why Volvo is looking for really effective algorithms. Fernando commented that can it be easier to de-identify a video for humans than for computers. Computers are able to recognize faces by using various measurement and search techniques. Background search may also form an issue, e.g. Google can search for photos that have the same background. The driver may have posted a photo of him- or herself online, with the same background.

There was a question on why not to share only action units. Fernando replied that high-level actions can be shared like head pose and action units. Low-level features depend on algorithms. The high-level actions are totally anonymous and can truly be a way to share information safely.

He explained that their face tracker in three-year tests used in day-time performs well over 90% of the time. Night-time performs currently less. A likely performance goal is 95% - but will that be enough for products that anonymise data in real-time?

7. Anonymising moving objects in the vehicle surroundings, Clement Val (CEESAR)

Clement Val gave a presentation on video anonymisation in UDRIVE, the European Naturalistic Driving Study. Legal constraints for this study in France were that pedestrians should not be a part of the dataset. It was requested to systematically blur pedestrians' faces and vehicle licence plates. Moving objects outside a moving vehicle can be difficult to blur: there are occlusions, inconsistent illumination, shadows. A large quantity of videos needs to be processed in a short time in the project. They used the WASSA company to develop robust and low-cost anonymisation solutions, and the achieved performance satisfied the French authority. He showed an example of the processing working in an intersection scenario.

8. Using FACS-coder to anonymise drivers in eye-tracker videos, Selpi (Chalmers/SAFER) and Torsten Wilhelm (Smart Eye)

Selpi & Torsten Wilhelm gave a joint presentation about using the FACS (Facial Action Coding System) coder to anonymise drivers in eye-tracker videos. Selpi explained that a driver's face is a rich source of information and continued to discuss how to access this data without breaching data privacy. Ekman (1970) used facial expressions (with motions of facial muscles) as a way to express six basic emotions and she continued to introduce the FACS, the system: objective measures to describe facial expressions in terms of Action Units.

Torsten explained that Smart Eye has built an automatic FACS coder, extracting features and classifying Action Units (e.g. brow raisers, lid raiser, cheek raiser). These are fed into a face reconstruction algorithm; the output being anonymised video. They use 16 detectors, covering 96% of all variations in training data. They concluded that the system is able to anonymise the driver's face, but does not have enough accuracy in preserving the facial expression. Potential improvements would be to improve feature extraction stage, using both shape and grey value to classify Action Units and improving the translation of wrinkles and precision around mouth, eyes, and eyebrow.

9. DCode and DMask: Two approaches for video data anonymisation, Amir Tamrakar (SRI International, USA)

Amir Tamrakar presented two projects that could enable sharing of the personal data in the large dataset from the SHRP2 Naturalistic Driving Study: DCode, the Comprehensive Automatic Coding System for Driver Behavior Analysis to assist in the automatic coding of features relevant to safety researchers and DMask, the Reliable Identity Masking System for Driver Safety Video Data, to generate identity masked video that can be disseminated to a wider audience. These two projects provide different approaches, anonymisation by coding driver activity and driving context (DCode), and by masking the driver's face and body (DMask). Masking the location is still only a proposal. The approach to masking consists of three tiers: core feature extraction by tracking relevant features such as faces and upper

body position, intermediate feature extraction by monitoring and analysing gaze, expression, gestures etc., and feature integration, leading to the final coded features of the driver actions, state and driver environment. The masking approach masks out the driver's head with an overlaid synthetic avatar. Both approaches make use of learning mechanisms to improve performance. Although results are good so far, the success of facial motion transfer and identity masking is totally dependent on the accuracy of the facial feature tracking. And to have 100% accuracy, still a human analyst has to be involved in the process.

10. Open discussion on issues presented in the previous cases, Chaired by Marco Dozza (Chalmers/SAFER)

On the question what the next challenges are in order to be able to 100% anonymise, the speakers agreed that the input data quality, such as the head tracking quality, should improve. To really reach 100% is very hard, although progress is being made and commercial systems are available. Ideal would be to anonymise in real-time, at the source, instead of afterwards on a dataset. That is also important when these techniques are used in real-time, such as for automation of vehicles instead of for research purposes. It is important to understand the reasons for systems' failures. The models that are used should be adaptable. New sensors are going to help to improve data quality, such as 3D visual data sensing and infrared sensors for bad lightning conditions.

Recommendations for FESTA should include the emphasis on improvement of data quality. Good sensors are becoming cheaper. As the presented mechanisms include learning, good meta-data are needed and training data. This training data may come from other datasets. Location remains a difficult problem, it is easy to identify famous locations, or search for them on Google.

The issue was raised that there is not yet enough experience with how well researchers can do their analysis with anonymised data. Combination of techniques may also make analysis easier.

The question was raised of how much anonymisation is really needed; research, ethical and legal requirements are not always the same.

There is a lot of emphasis on extracting facial features, but other features, such as body language are also important to understand what is going on. In data-collection such as SHRP2, cameras do not always capture other body parts very well.

A question from the audience asked how far we are from being able to query a database for specific episodes, for example "give me all frames where the driver is using a cell-phone". Most anonymisation using marking techniques are running nearly real-time. This kind of queries is already (or will soon be) possible in batch processing, given sufficient computing power. Storing outputs from these anonymisation processes do not require a huge additional space.

We need to know more about what researchers need, how accurate the marked data should be, and what error rate and false alarms rates are acceptable? How much information is needed to understand the driver and the situation? For example when using avatars, is gender and ethnicity important? The current algorithms are rather decent, but not perfect, some data are lost. More evaluation is needed.

Marco gave as his wish list to be able to identify (1) critical situations (the “oops” reaction by the driver) (2) engagement in secondary tasks and (3) driver impairment (e.g. yawning because of fatigue). He wanted to know whether these are easy to identify reliably. The panel answered that giving the right sensors all three are doable (but there may be a cost issue), it is just a matter of tracking. However, the rule of thumb is that everything you want to find that is subtle, it is much harder. And if you want to predict frames, such as predicting when the driver will fall asleep, it is much more difficult.

The learning mechanisms for the anonymisation processes, such as the use of neural networks, are usually unsupervised in the lower stages but supervised in the later stages, such as for detecting drowsiness.

Day break

11. Recap of Day-1 outcomes, Satu Innamaa (VTT)

Satu Innamaa kicked off the second day with a recap of the first day's discussions. She highlighted that users of the anonymised data, the analysts, must be informed about the quality of processing, e.g. what expressions carry over accurately and which ones could become mixed up. Such quality topic could also be included in FOT metadata documentation regarding quality of processed information.

12. Anonymisation: Balancing Re-use and Privacy – the US Perspective, Dwayne Henclewood (BAH, USA)

Dwayne Henclewood (Booz Allen Hamilton under contract to the US Department of Transport) presented anonymisation of data in the RDE (Research Data Exchange), the repository and portal to archive and re-use real-time connected vehicle, traveller and roadway transportation data. He stressed that the same questions of protecting participant privacy have been discussed in both Europe and the US. He gave an overview of RDE and the Safety Pilot Model Deployment (SPMD). He also told personal experiences about using the NGSim dataset (next generation simulation, roadway in Atlanta). Three categories of sensitive data are distinguished: personally, commercially, and research sensitive data that may compromise the goals and objectives of a research endeavour. RDE is using a range of anonymisation methods. Traditional methods to anonymise trajectory data include masking true trajectories, suppressing trips' origin and destination, trajectory reconstruction via a range of mathematical techniques, and redefining the geospatial reference layer by offsetting latitude/longitude pairs. These techniques are effective in masking PII but limit data utilization. Dwayne commented about metadata, that "if you don't describe data, they will never use it in the context it should be used."

RDE has used an advanced method of truncating start and end part of trips, being further developed for being able to share 60-day sample of SPMD. In an iterative process intervals of concern are identified (e.g. stops), extended to privacy intervals, and intervals finally eliminated. Detection is done by using different features and methods, for example the use of reverse and park often can indicate the stop/end location as well. GPS quality had been a

issue, as poor positioning can cause problems for determining the value of distance and time parameters.

13. Trajectory Privacy: Measures and Preservation Methods, Gyözö Gidofalvi (KTH)

Gyözö Gidofalvi explained that a large portion of all data generated today are flowing through mobile networks. They capture the location, but users are often unaware of collection and use of this information. Location traces could be a goldmine - but privacy is a huge concern. Simple anonymisation seems not enough either. Research has shown that four observations from cell phone data are sufficient to identify 95% of individuals. He introduced k-anonymity, where it is required to have at least k records for every possible value of any subset of attributes (e.g. k persons for any gender and birthdate). He explained common options, generalising a value to make it less specific, e.g. age 34 becomes 30-40, suppression by simply deleting the value, and perturbation, replacing the actual value with a random value out of the standard distribution of values for that attribute. Also l-diversity is required: diversity of the sensitive attribute within the k-anonymity set. Gyözö has developed a "grid-based framework" for trajectory anonymisation, with several benefits. Evaluation of this method shows good results.

14. Managing privacy risks in autonomous driving and cooperative intelligent transportation systems, Henrik Broberg (VCC)

Henrik presented work being done in the Security workgroup from the Car 2 Car Communication Consortium and projects on connected ITS. The security problem consists of questions on the acceptance and trust of information exchanged between cars and infrastructure (C2I) given the threat of manipulation, and the disclosure of identity given the threat of eavesdropping. Cryptographic signatures are used to guarantee the integrity, authenticity as well as authorization of senders. He explained the PKI (Public Key Infrastructure) architecture and the way it is deployed in C2I communication, as well as the legal basis for dealing with personal data within the connected ITS context. He summarized what can be done to ensure security: avoid the risk by removing sensitive information, mitigate the risk by containing sensitive data, set a proper controls baseline, and ensure good governance (remember that a data protection officer is required by law).

15. Open discussion on issues presented in the previous cases and issues in a broader context (e.g. accident data, surveys etc.), chaired by Niels Agerholm (Aalborg University)

The participants discussed the balance between saving and deleting data for privacy reasons. Communication and transparency are crucial about what can be made public and what not, and the reasons for this.

It was remarked that mobile services operators already collect a lot of data about us and our locations, we also collect passively data ourselves when using apps. What will be the value for us of giving our (vehicle) data away? Trust in companies and data control are essential. We have to be aware that legislation on data protection is not aimed specifically at the automotive industry, but is more general. Also data issues are relatively new for the industry,

but they are addressing this issue, for example by appointing data protection officers. It is important to work together with lawyers and to inform lawyers so they can provide better support.

When opening the data, like in RDE, it is the objective to get researchers to do something with it. While asking to cite the source, the RDE team does not check publications coming from the (re-)use. If you start to share data you can no longer fully control it. That is why you have to anonymise and protect the sensitive data before sharing it.

Discussion also took place on the data quality; it's not just a matter of deleting bad data (like impossible locations) but also about the trust that the other data are accurate. There used to be a lot of problems with the quality of GPS data, but high fidelity positional data can now be acquired or will be more easily accessible in the near future.

It was concluded that it is very difficult to 100% anonymise all data; even if one type of data is complete anonymous, combinations of data (car type, colour, position etc.) can point to one individual. Even if legal requirements are met, what people consider as private and do not want to be disclosed is very personal. Would it be possible to make an anonymisation on an individual level, through asking the person what they do not want to have revealed, for instance, which GPS positions do they not want to reveal? Trying to make data as secure as possible and to communicate about how data are protected, and the benefits of studies, is probably the best way forward.

16. Wrap-up, Yvonne Barnard (University of Leeds)

Yvonne Barnard wrapped up the workshop. New and promising techniques were discussed but not all problems related to anonymisation are solved. Anonymisation of driver faces is rather advanced, but context is more difficult, and anonymising location and trajectories is still problematic. The latest anonymisation techniques still have to prove their value for the large FOT datasets. If we want to use them for automation and C2C systems instead of research, 100% perfection is needed. Data (and metadata) quality is key for being able to apply the techniques successfully.

Anonymisation is becoming more and more important, due to growing awareness of the need to protect participants' privacy, new laws and regulations, and requirements from project funding bodies. This interesting workshop concluded that (international) collaboration and discussion between researchers, computer scientists and legal experts is necessary to advance the development towards more and more successful ways to keep personal data anonymous. If the data could be anonymised while still keeping the information that is essential to research, the access and the re-use of the valuable data would be greatly facilitated.

Participants FOT-Net Data workshop, 1–2 September 2015, Gothenburg

| Last Name | First Name | Company |
|------------------|--------------|--|
| Agerholm | Niels | Aalborg University |
| Barnard | Yvonne | University of Leeds |
| Bode | Fredrik | Volvo Group |
| Brizzolara | Davide | ERTICO |
| Broberg | Henrik | Volvo Cars |
| Bärgman | Jonas | Chalmers University of Technology |
| Carlsson | Anders | Volvo Group |
| Chen | Haibo | University of Leeds |
| Christoph | Michael | SWOV |
| de la Torre | Fernando | Carnegie Mellon School of Computer Science |
| Ding | Chengkai | Autoliv |
| Dobberstein | Jan | Daimler |
| Dombrovskis | Sergejs | Volvo Cars |
| Dozza | Marco | Chalmers University of Technology |
| Eliasson | Mats | SRI International |
| Fredriksson | Rikard | Autoliv |
| Gellerman | Helena | SAFER |
| Gidofalvi | Gyöző | KTH Royal Institute of Technology |
| Gjertz | Mats | Volvo Cars |
| Grönvall | John-Fredrik | Volvo Cars |
| Gustafsson | Dan | Volvo Cars |
| Heinig | Ines | SAFER |
| Henclewood | Dwayne | Booz Allen Hamilton |
| Hermanson | Thomas | Volvo Cars |
| Hogema | Jeroen | TNO |
| Inaama | Satu | VTT |
| Isaksson-Hellman | Irene | If Skadeförsäkring AB |
| Ito | Hiroshi | Japan Automobile Research Institute |
| Johannesson | Pär | SP Technical Research Institute of Sweden |
| Johansson | Mikael | Volvo Group |
| Karlsson | Johan | Autoliv |
| Karlsson | Tobias | Volvo Cars |
| Koskinen | Sami | VTT |
| Kovaceva | Jordanka | Chalmers University of Technology |
| Linse | Leif | Trivektor |
| Maghsood | Roza | Chalmers University of Technology |
| Perez | Miguel | Virginia Tech |
| Pirnia | Elham | Volvo Group |

| | | |
|-----------|-----------|-----------------------------------|
| Roos | Nathalie | Volvo Group |
| Rydberg | Gunilla | Sjöland & Thyselius |
| Selpi | Selpi | Chalmers University of Technology |
| Sui | Bo | Autoliv |
| Svanberg | Erik | SAFER |
| Tamrakar | Amir | SRI-International Sarnoff |
| Val | Clement | CEESAR |
| Wallgreen | Pontus | Chalmers University of Technology |
| van Nes | Nicole | SWOV |
| Vedad | Sajic | Volvo Cars |
| Wiberg | Per-Johan | Ericsson |
| Wilhelm | Torsten | SmartEye |

5 FOT-Net Data Workshop 3: A common methodology for road automation FOTs and pilots

In Europe, Field Operational Tests (FOTs) and pilots have been designed and conducted using the FESTA methodology, developed by the FESTA consortium and maintained and updated by the FOT-Net support actions. The focus of these projects has been on changes in driver behaviour, and resultant societal impacts due to use of in-vehicle systems and cooperation with the infrastructure and other vehicles. In the coming years large-scale projects will be started for testing automated or even autonomous vehicles.

In this seminar, we went step-by-step through the FESTA methodology, focussing on those issues that need to be addressed in automation FOTs and pilots. Five major areas were addressed:

- Societal scenarios and research questions
- Study design
- Data collection and analysis
- Impact and socio-economic cost-benefit analysis
- Data sharing

The objectives of this seminar were to explain the current methodology, to extend the FESTA activities to automation pilots, and to work together taking the first steps towards a common methodology. The seminar provided both plenary presentations and small group discussions and exercises. We worked with example FOTs, and covered all the methodological steps.

Agenda

On 3–4 February 2016, a workshop was held in Leeds, UK, entitled ‘A common methodology for road automation FOTs and pilots’. The agenda was as follows:

Wednesday, 3 February, 12:00 – 17:30

1. Overview of this workshop: Haibo Chen (University of Leeds)
2. FOT-Net Data introduction: Sami Koskinen (VTT)
3. Knowledge exchange activities related to FOTs: Haibo Chen (University of Leeds)
4. Automation FOTs and pilots in Europe: Davide Brizzolara (ERTICO – ITS Europe)
5. USDOT Programs for Collecting Connected & Autonomous Vehicle Data: Cory Krause (Noblis / U.S. Department of Transportation)
6. Current status of the FESTA methodology: Yvonne Barnard (University of Leeds)
7. Societal scenarios and research questions: Oliver Carsten (University of Leeds)
 - a. Case introduction: PEGASUS project: Jan Dobberstein (Daimler)
 - b. Small group work
8. Study design: Yvonne Barnard (University of Leeds)

- a. Case introduction: Nordic Way Lone-Eirin Lervåg (SINTEF)
- b. Small group work

Thursday, 4 February, 9:15 – 13:15

9. Data collection and analysis: Sami Koskinen (VTT)

- a. Case introduction: Barbara Metz (Würzburg Institute for Traffic Sciences)
- b. Small group work
- c. Impact and socio-economic cost-benefit analysis: Satu Innamaa (VTT) Case introduction: Benefits estimation framework for automated vehicle operations: Scott Smith (Volpe National Transportation Systems Center / U.S. Department of Transportation)
- d. Small group work

10. Data sharing: Helena Gellerman (SAFER)

- a. Case introduction: RDE research data exchange in the US: Corry Krause (Noblis / U.S. Department of Transportation)
- b. Small group work

11. Questions and discussion

12. Wrap-up: Yvonne Barnard (University of Leeds)

There were 31 workshop participants (the list of attendees can be found at the end of this document).

1. Overview of this workshop: Haibo Chen (University of Leeds)

Haibo Chen welcomed the participants and outlined the programme of the workshop. He introduced the Institute for Transport Studies of the University of Leeds.

2. FOT-Net Data introduction: Sami Koskinen (VTT)

Sami Koskinen introduced the FOT-Net Data project that organised the workshop. For the upcoming automation FOTs and pilots data sharing will be an important issue. FOT-Net Data has developed a Data Sharing Framework, which covers e.g. data and metadata description and data protection recommendations (<http://fot-net.eu/library/?filter=data-sharing-framework>). The FOT Data Catalogue is ready and we are inviting data owners to create entries to the catalogue: wiki.fot-net.eu.

3. Knowledge exchange activities related to FOTs: Haibo Chen (University of Leeds)

Haibo Chen presented briefly the knowledge exchange activities of FOT-Net: webinars and workshops. The webinars address the FESTA handbook content, and recordings and materials may be found online at FOT-Net Library (fot-net.eu/library). Presentations and reports from the previous workshops are also available there.

4. Automation FOTs and pilots in Europe: Davide Brizzolara (ERTICO – ITS Europe)

Davide Brizzolara gave an overview on European and national projects and pilots dealing with road automation. Ongoing FP7 projects include:

- AdaptiVe: Automated Driving Applications and Technologies for Intelligent Vehicles
- CityMobil2: pilot platform for automated road transport systems, implemented in several urban environments, focus on public transport
- Companion: developing a framework of co-operative mobility technologies to enable supervised vehicle platooning of heavy-duty vehicles
- iGame: Interoperable Grand Cooperative Driving Challenge AutoMation Experience, with a focus on merging on highway and mixed traffic intersection
- AutoNet 2030: developing and testing a co-operative automated driving technology based on a decentralised decision-making strategy which is enabled by mutual information sharing among nearby vehicles
- Cargo-Ants: Cargo handling by Automated Next generation Transportation Systems for ports and terminals.

National projects include:

- UK: 3 projects from the INNOVATE UK competition: Venturer, GATEway and UK Autodrive
- Sweden: Drive Me
- Germany: PEGASUS
- Belgium: BE Zaventem
- The Netherlands: WEPOD, DAVI (Dutch Automated Vehicle Initiative)

More information about these and other projects may be found at the wiki of the VRA (Networking in Automation) coordination and support action: <http://vra-net.eu/wiki>.

Davide concluded by discussing the upcoming calls in the Horizon2020 framework that are relevant for road automation.

5. US DOT Programs for Collecting Connected & Autonomous Vehicle Data: Cory Krause (Noblis / U.S. Department of Transportation)

Cory Krause explained the US Department of Transportation (US DOT) programs for Collecting Connected & Autonomous Vehicle Data (CDS) (www.its.dot.gov).

The US DOT's vision is to show how emerging data can transform surface transportation systems management. The CDS Program seeks to develop, test and make available methods to capture, manage and exploit the potential of high-volume multi-source data to enhance current operational practices, and transform future surface transportation systems management.

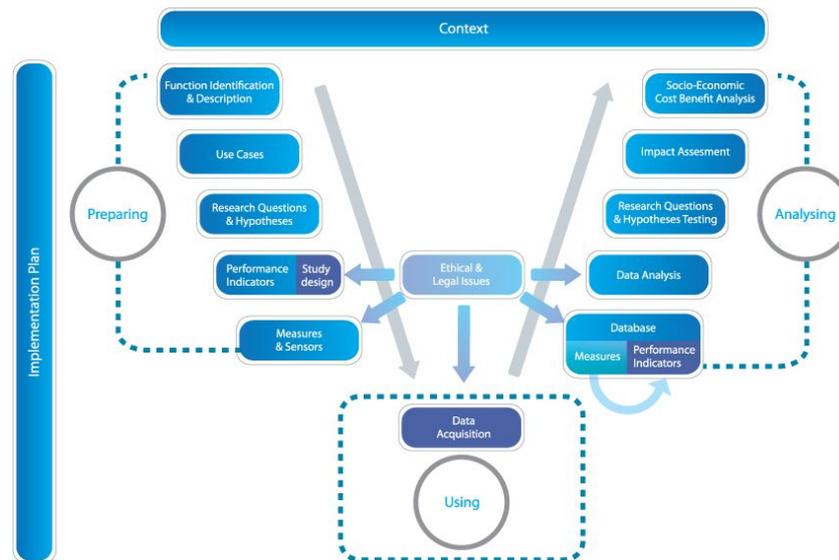
In 2015, three sites were awarded:

- Wyoming, focused on the needs of the commercial vehicle operator, aiming to reduce the number and severity of adverse weather-related incidents
- New York City: to improve safety and mobility of travellers through connected vehicle technologies
- Tampa: to alleviate congestion and improve safety during morning commuting hours

Big data collected in the Safety Pilot project, held at Ann Arbor, MI, is one of the datasets that is being made available for other researchers at the Research Data Exchange (www.its-rde.net). This pilot, with 2836 vehicles driving on 73 miles of roads, examines the effectiveness of connected vehicle safety applications.

6. Current status of FESTA handbook, Yvonne Barnard (University of Leeds)

Yvonne Barnard presented the FESTA methodology, introducing the need for a common methodology and going through the methodology steps in the so-called FESTA-V (see below). The question to be addressed in this workshop is what the next step will be: FESTA for automation studies. Many practical issues remain the same, designing and performing an FOT, and analysing the results, and the systematic and scientifically based approach to gather sound evidence on the effects of automation. A common methodology is important for being able to study impact of automation, combining all knowledge gathered from different kinds of FOTs. New focusses are needed on users, vehicles and contexts. Sharing experiences, knowledge and data will remain important. The handbook can be found at: <http://fot-net.eu/Documents/festa-handbook-version-5-2014/>. The handbook will be updated at the end of 2016.



7. Societal scenarios and research questions: Oliver Carsten (University of Leeds)

Oliver Carsten opened this session by addressing the upper left part of the V (for more information see the FESTA handbook). He stated that vehicle automation could become a societal game-changer, the issues being “what are the wider lifestyle, transport, economic and social impacts?” FESTA, addressing its systems, identified as impact areas: safety, efficiency, environment, mobility, acceptance and trust, usage, adoption (making inherent in behaviour). With automation, new impact areas will also become important: transport demand, value of time, travel costs, mode choice, network operation, infrastructure costs, land use, urban planning, environment, and equity.

7a. Case introduction: PEGASUS project: Jan Dobberstein (Daimler)

Jan Dobberstein presented the German PEGASUS project. The core research questions are: What criteria do systems for highly automated driving have to answer to? What is necessary in order to ensure that systems actually fulfil those criteria? The project will perform virtual tests, proving ground tests and field tests. They will investigate socially accepted risk criteria. More information on: www.pegasus-projekt.info.

7b. Group work

In small groups societal scenarios and research questions were discussed, starting from one of the example cases:

- Driverless bus in restricted area
- Vehicle with driver on the highway, lane keeping and forward control fully automated
- Platooning of trucks
- Fully autonomous car in city environment

The following table was used to capture the results.

| Questions | Case study | Recommendations for FESTA | Methodological issues |
|--|------------|---------------------------|-----------------------|
| Societal scenarios and research questions | | | |
| 1. Describe a scenario on which you will focus | | | |
| 2. What would be the main impact areas? | | | |
| 3. Define three main research questions | | | |

Groups looked at different scenarios and use cases such as fully automated cars, shared urban transport, driverless public transport, and remote control of vehicles. Impact areas are very broad as we are dealing with very different eco-systems and business models. Next to safety, environment, efficiency, and mobility, also for groups who have now restricted mobility, security is an important area. But also questions about comfort, reliability, trust, vehicle ownership, roads and infrastructure, traffic flow, actual use, vulnerable road users, public health, mode change, willingness to pay and (many) other questions were discussed. In addition, impact on life-styles was mentioned, as well as perception of time and use of time freed up by automation. It was recommended for FESTA to address new impact areas. As methodological issues, it was mentioned that we should go from observation to prediction, and that we need to develop economical models for the automation eco-system. Another issue is the scaling up, what is the testing that is needed for deployment purposes? Different groups, (young, old, disabled...) need to be taken into consideration.

8. Study design: Yvonne Barnard (University of Leeds)

Yvonne Barnard explained the lower part of the left branch of FESTA: study design, specifically the definition of hypotheses, measures and performance indicators. She asked whether we know what the performance indicators will be for automation FOTs as new impact areas and research questions will lead to new types of indicators. In the study design, the experimental set-up of the study, the question will be who the participants are: drivers, monitors in the vehicle, passengers, other road users and general public, traffic managers,? An interesting question is that of the baseline. In FESTA the baseline is the situation without an ITS system being active, so comparisons can be made with the situation in which the tested system is active. In automation FOTs this is not so clear: what do we compare automation with? No automation, "normal driving" (findings from naturalistic driving studies), lower levels of automation, individual driver baseline, different penetration levels of automation? Do we need a baseline?

8a. Case introduction: Nordic Way Lone-Eirin Lervåg (SINTEF)

Lone-Eirin Lervåg presented the study design in the NordicWay project, the pre-deployment piloting of C-ITS services in the Nordic countries. Their goal is to enable vehicles to communicate safety hazard and road status information using cellular networks (3G and LTE/4G) offering interoperable services to the users on a cooperative road corridor. Technical performance, impacts and user acceptance are evaluated, resulting in a roadmap for large-scale deployment. The ultimate goal is to prepare for road automation in the Nordic countries. NordicWay uses the program theory as the basis for their methodology: how the intervention is expected lead to effects, and in which conditions.

8b. Small group work

Discussions were held on the use cases from the first round of group work, with the following questions:

- Formulate 3 hypotheses and define performance indicators for each hypothesis
- Sketch a study design
- What would be the baseline for your study?

Participants came up with a wide range of options, but it became clear that it was not easy to formulate hypotheses, there are too many options. For the baseline, a study in which commuters could be compared with train passengers (or other public transport) was mentioned. Another baseline could be the experience of (customers of) car sharing companies. Another idea is to study a currently available transport mode and restart a service that is now no longer available (but now in automated form), and see what happens and compare them.

Methodological issues identified were that the impact of automation on mode choice is very hard to study. Sensitivity analysis to identify the most essential parts in the study may be needed. Pre-studies will become more important (for example on current car sharing). Non-users should also be included in the studies.

9. Data collection and analysis: Sami Koskinen (VTT)

Sami Koskinen presented the lower part of the FESTA V: data collection, data processing and the data analysis steps. In automation FOTs, new emphases may be chosen: (1) High emphasis on recording video, with the need to annotate events to learn contributing factors and reasons for incidents, and maintaining privacy. (2) Advanced sensing, the driver's pose and various environmental features need to be recorded. (3) Addressing the question how human driving differs from automated driving (like differences in lane keeping, lane change, speed at intersections, time to collision in some situations). And (4) Opinions, reactions and behaviour of other road users are of interest.

9.a. Case introduction: Barbara Metz (Würzburg Institute for Traffic Sciences)

Barbara Metz presented recommendations for data collection derived by re-analysing euroFOT-data. Examples of research questions are: usage and acceptance of different navigation systems, and impact of navigation systems on driving behaviour, safety and efficiency. For an efficient and valuable re-use of data, more signals should be recorded than needed for primary analysis. Signals should also cover aspects of driving that are not in the scope of the original project (e.g. signals for all available ADAS, air condition, etc.).Video

data is essential to develop and validate algorithms for new research questions. They made three recommendations for new approaches in the analysis of FOT data: (1) situational variance should not be controlled through experimental design but through the selection of relevant data sections, (2) algorithms need to be developed that divide the data into meaningful sections (e.g. driving manoeuvres), and (3) algorithms need to be adapted and validated for every vehicle type.

9b. Small group work

In small groups further discussions were held on data collection and analysis, with the following questions:

1. What kind of data would you collect?
2. What sensors would you use?
3. What would be the main challenges for data collection and analysis?

Next to the usual sensors measuring vehicle behaviour, video was mentioned as a key means to collect data, collection of images in every possible direction. The other major type of data are the experiences and opinions of users. The response of users should be measured before and after experiencing automated transport. Also measuring participants feelings in real-time was mentioned. We also have to be aware that new behaviours may emerge, which are not anticipated, but may be extremely interesting. Data on societal change would need to be collected. In addition, data about the infrastructure and the environment is needed. As one group stated: we need “everything”.

Methodological issues identified are: Huge amounts of data will be collected and the filtering of relevant data, e.g. for identifying incidents will become very important. Challenges are in the area of video analysis, automated/intelligent video annotation and interpretation of sensor data. New data analysis expertise will be needed. With (high resolution) video the issue of privacy comes up. Another issue is that long test periods may be needed. Sensor data from OEMs and TIER1s may become an issue, as they may not be prepared to share it. An interesting question is if you do AD simulator or control studies, how will you find participants prepared to do nothing for a long time? The biggest challenge may be dealing with the wide variety of data, maybe it is good to start with small case scenarios.

10. Impact and socio-economic cost-benefit analysis: Satu Innamaa (VTT)

Satu Innamaa introduced the upper right part of the FESTA V: Impact and socio-economic cost-benefit analysis. This analysis is needed to increase knowledge about impacts of ITS and to provide the impact assessments in numerical estimates for the cost-benefit analysis. It provides information to stakeholders and decision makers about potential business cases and input for future work. Automation FOTs will present new challenges, and many impacts take time to form, like mobility patterns, new mobility services, land use, and public health (accessibility to health services, walking & cycling). How can we measure or assess these and how do we estimate the benefits in financial terms? She discussed the benefits of a common impact assessment framework allowing the exchange of best practices, the designing of tests and studies to maximize insight, and enabling meta-analyses. She presented the Trilateral Impact Assessment Subgroup for Automation in Road Transportation (between USA, Japan and EU), of which she is the EU co-chair.

10a. Case introduction: Benefits estimation framework for automated vehicle operations: Scott Smith (Volpe National Transportation Systems Center / U.S. Department of Transportation)

Scott Smith (the US co-chair of the trilateral working group) presented the development of the Automated Vehicle Multimodal Benefits Framework, the US framework to estimate the potential safety, mobility, energy and environmental benefits (including dis-benefits) of technologies contributing to the automation of the US surface transportation system. The objectives are to identify metrics, develop a framework for quantifying impacts, provide a high-order assessment of the state of knowledge, and incorporate current research by other parties. The final phase 1 report, Benefits Estimation Framework for Automated Vehicle Operations (FHWA-JPO-16-229), is now available in the National Transportation Library (ntl.bts.gov). This research is sponsored by the US DOT Intelligent Transportation Systems Joint Program Office.

There is substantial interest in the framework and a need is identified for a clearing-house on research, to facilitate sharing. We need to understand the big picture to ensure the right data are collected. Data sources and automation applications for initial modelling need to be identified, and linkages between micro and regional mobility models examined. It is necessary to develop AV impact models, starting with safety, mobility and environment, but continuing to other areas. International coordination of evaluation activities is very important.

10b. Small group work

In small groups discussions were held on impact assessment and the methods and information needed, with the following questions:

1. Select an impact area and discuss where your focus would be
2. What additional information would you need?
3. How would you proceed to determine the impact?

Again impact areas were discussed that were identified in the first group work session. From the group discussion, it became clear that it is not possible to address everything at the same time, some groups focused on limited scenarios, like parking and found it easier to come up with a plan to determine the impact. Also the group that focused on accessibility for different groups came up with a plan. These groups combined sensor data with questionnaires and statistics available in the public domain.

Impact assessment is a very difficult, but essential topic, and will require more thought in order to come up with recommendations.

11. Data sharing: Helena Gellerman (SAFER)

Helena Gellerman explained the role of data sharing in the FESTA methodology, and the data sharing framework that is further developed in FOT-Net Data. Data sharing needs to be considered from day 1 in an FOT: in agreements, funding, determining what data to collect, documentation, analytical tools, and data protection. She presented areas where support may be needed for data sharing, like training, financial models, support services, and application procedures. More information can be found in the (draft) Data Sharing Framework available on the FOT-Net website (www.fot-net.eu). Specific challenges for data sharing in automation FOTs are related to the data gathering from competitive systems (sensors, algorithms) and data ownership. Video will be used and will gather sensitive information, also

about other road users, so we will need to deal with legal restrictions and privacy issues. If we want data to be more open, data protection is of the utmost importance.

11a. Case introduction: RDE research data exchange in the US: Corry Krause (Noblis / U.S. Department of Transportation)

Cory Krause gave a further presentation on RDE. The purpose of RDE is to provide a variety of data-related services that support the development, testing, and demonstration of multi-modal transportation, weather, and environmental applications. RDE enables the sharing of data collected from connected vehicles, mobile devices, and infrastructure. It performs data quality checks, provides clean, well-documented data sets, and integrates data from multiple sources into data environments. RDE provides real-time and archived data, probe data from field tests, and data from research projects including simulations. On the website, it is easy to search for datasets, and they can be downloaded by registered users, the registration procedure is very simple, and a try-out is strongly recommended: www.its-rde.net.

ITS researchers are invited to submit data sets to RDE. US DOT assesses potential data on value and quality, and when a data environment passes this assessment RDE management adds the data to RDE.

11.b. Small group work

In small groups questions about data sharing were discussed:

1. What data would you be able to share?
2. What would be obstacles to sharing?
3. What would be the enablers for sharing?

Discussion centred on data and privacy protection, anonymization of data and permission from participants. The other discussion issue was the sharing between organisations, such as the problems manufacturers have in sharing data sensitive to their products and business. Enablers would be techniques for anonymization, which are currently not yet good enough, legislation/regulations and data sharing agreements.

12. Questions and discussion

In the discussion it became clear that thought about the impact of automation and the way in which to investigate it is moving in the same direction in the EU and the US, although sometimes different wording is used. It also became clear during the workshop that automation can mean very different things and multiple future scenarios are possible. The challenges are not only technological, but travel in the future may be quite different from what it is today. Nevertheless, we should focus on the next level of automation, becoming available in the near future, and for which FOT data may provide insights that are useful for the further development of automated vehicles. A concern for automation FOTs is the authorisation needed to conduct FOTs in Europe on the public road. The nature of FOTs may change. FOTs are not just about research per se, but also about gathering evidence that these systems/vehicles are safe. We may conclude that while recommendations from FESTA are still very useful, “FESTA for automation” is not “business as usual”.

13. Wrap-up: Yvonne Barnard (University of Leeds)

Yvonne Barnard concluded the workshop remarking that there is still a lot we don't know about how automation FOTs will be or needs to be designed and conducted. But in terms of FESTA we know that the context will be on a societal level (not just looking at specific systems or traffic environments) and will be constantly changing and at different speeds. The impact will be also in new areas, including areas outside the transport domain. The research questions may be hard to define at the beginning of a FOT and may be changing during the studies. Only a limited set of questions can be addressed, which will not cover all impact issues. We discussed whether we are ready to formulate hypotheses and determine a baseline (baselines may be divers). Many performance indicators will be related to user behaviour and attitudes.

Data collection was discussed. Maybe we should try to collect as many data as we can, we may need them to analyse the impacts. Also a large number of data will be needed from the environment, infrastructure etc. This brings us to the question of whether manufacturers will be willing and able to provide data. New expertise for analysts will be needed to deal with data and to draw conclusions about the impacts. We may need more visioning methods, to be able to predict the future consequences of automation. To complicate impact analysis even further due to automation values may change, such as the value of time, land use, ownership, etc. International sharing is the only way forward to tackle all these difficult questions, not only of data but also of data strategies and insights.

Participants FOT-Net Data workshop

| Last name | First name | Organisation |
|---------------|------------|--|
| Barnard | Yvonne | University of Leeds |
| Brizzolara | Davide | ERTICO - ITS Europe |
| Carsten | Oliver | University of Leeds |
| Chen | Haibo | University of Leeds |
| Dobberstein | Jan | DAIMLER AG |
| Dodsworth | Joel | ITS Leeds |
| Franzen | Stig | Chalmers |
| Gellerman | Helena | SAFER |
| Hibberd | Daryl | University of Leeds |
| Hill | Allan | Transport Scotland |
| Innamaa | Satu | VTT |
| Jackman | Gavin | telent |
| Koskinen | Sami | VTT |
| Krause | Cory | Noblis / USDOT |
| Landau | Andreas | WIVW GmbH |
| Lervåg | Lone-Eirin | SINTEF Technology and Society |
| Levin | Tomas | Norwegian Public Roads Administration |
| Madigan | Ruth | ITS |
| Merat | Natasha | ITS, LEEDS |
| Metz | Barbara | WIVW GmbH |
| Page | Yves | RENAULT SAS |
| Pampel | Sanna | University of Nottingham |
| QUINTERO | Karla | CEESAR |
| Rial Martinez | Moisés | CTAG |
| Romano | Richard | University of Leeds |
| Smith | Scott | United States Department of Transportation |
| Stephenson | Scott | AECOM |
| Thomasson | Erik | University of Leeds |
| Welsh | Ruth | Loughborough University |
| Willey | David | Jaguar Land Rover |

6 FOT-Net Data Workshop Extra: FOT Data Anonymization and feature extraction

Contents

1 Introduction

2 Day 1, Wednesday, 31 August 2016, 14.00-18.00

2.1 Workshop introduction, Anna Nilsson-Ehle

2.2 FOT-Net Data, Sami Koskinen (VTT)

2.3 Importance of real world data for safety systems development, John-Fredrik Grönvall (Volvo Cars)

2.4 What are the main issues with anonymization and feature extraction?, Helena Gellerman (SAFER)

2.5 US perspective on anonymization, Hannah Rakoff (VOLPE, US DOT)

2.6 Different privacy for data mining and querying, Devdatt Dubhashi (Chalmers)

2.7 Computer resources for large-scale anonymization / feature extraction, Ann-Charlotte Sonnhammer (SNIC, Uppsala University)

2.8 Video anonymization of vehicle environment, Jan-Erik Solem (Mapillary)

2.9 Open discussion, John-Fredrik Grönvall (Volvo Cars) and Helena Gellerman (SAFER)

3 Day 2, Thursday, 1 September 2016, 09.00-12.30

3.1 Wrap-up of Day 1, Satu Innamaa (VTT)

3.2 Feature extraction – how to assure you can delete data without information loss, John Lee (University of Wisconsin-Madison)

3.3 Using automated video annotation for feature extraction, Flora Dellinger (WASSA)

3.4 Automated labelling and recognition, Jan Erik Solem (Mapillary)

3.5 Open Discussion, Marco Dozza (SAFER)

3.6 Wrap-up, Helena Gellerman (SAFER)

4 Participants

6.1 Introduction to the report

Personal data such as video and GPS is collected in many FOTs and NDSs. This data contribute with essential information that e.g. enables an objective view on the driver behaviour during different driving scenarios and facilitate map-data enrichment to the collected naturalistic data. The personal data demands though a high level of data protection not to reveal the personal identity outside the group of researcher that has approved access to the data. If the data could be anonymized while still keeping the information that is essential to research, the access and re-use of the valuable data would be greatly facilitated.

This workshop provided insights into the current state-of-the-art of different ways of anonymizing and extracting features from video data. It also included discussions regarding the trends, the needs and the possibilities for the near future, where stakeholders working with data from self-driving vehicles have an interest in the result of the progress.

There were 51 workshop participants (the list of attendees can be found at the end of this document) with a variety of backgrounds, such as road user behaviour research, computer science, vehicle system development, law, and data protection and privacy. The selection for the topic of the workshop 'FOT Data Anonymization and feature extraction' was successful, as there were numerous discussions based on the presentations among the speakers and participants.

All presentations are available on the FOT-Net website; fot-net.eu/library.

6.2 Day 1, Wednesday, 31 August 2016, 14.00–18.00

2.1 Workshop introduction, Anna Nilsson-Ehle (SAFER) and Helena Gellerman (SAFER)

Anna Nilsson-Ehle opened the workshop by telling the history of field operational tests (FOT) and naturalistic driving studies (NDS) in Gothenburg, starting 2006. The privacy and data protection have always been at the core of the research, since in the beginning, four vehicle manufacturers were uploading their data into the same database at SAFER. Helena Gellerman also welcomed everyone and told about the background of this workshop. She presented the agenda of the two-day workshop.

2.2 FOT-Net Data, Sami Koskinen (VTT)

Sami Koskinen presented FOT-Net Data project which is organising this workshop. Since 2008 there have been a lot of FOTs in Europe. FOT-Net was established at the same time to support FOTs. FOT-Net Data project (2014-2016) is focusing on sharing of FOT data. FOT-Net Data's results include the Data Sharing Framework, the FOT Data Catalogue and the FOT network operation. FOT-Net's website includes links to past webinars on different aspects related to performing an FOT, FESTA, (e-learning material), and all presentations and documentation from past workshops, and more. FOT-Net Wiki is a free encyclopaedia on FOTs.

2.3 Importance of real world data for safety systems development, John-Fredrik Grönvall (Volvo Cars)

John-Fredrik Grönvall gave a presentation on the importance of real world data for safety systems development from an OEM point of view. Extensive knowledge is needed on how

vehicles and people interact to be able to develop automated vehicles. For this, good data is needed.

John-Fredrik gave definitions for 'anonymised data' and 'un-identified data'. The difference is whether you can go back or not (un-identified is something for which you cannot go back). It is easy to share un-identified data.

OEMs have lots of data on our mobility. This data is sensitive. Thus, access to it must be kept restricted although there might be demand for it, i.e. there is a conflict of interest (wish list of data vs. what you can provide). John-Fredrik gave an overview on different monitoring systems included in the new cars.

The sensor possibilities and the need to collect more data about the driver and especially the vehicle environment is developing fast. So, we have to develop the anonymization methods, too. Methods are needed that protect the privacy but still do not lose the essential information, vital to the study. FACS classification was presented as one method to keep the essential information without having to keep the face. To achieve reliable feature extraction will take time but development is going in that direction.

It depends on the level of automation whether information on the driver is needed; for automation levels lower than 5 it still is.

2.4 What are the main issues with anonymization and feature extraction?, Helena Gellerman (SAFER)

Helena Gellerman discussed on what the main issues are with anonymization and feature extraction. Data collection started to include continuous video in large scale some 10 years back. Also continuous GPS positions of the start and end of the trips can identify a person.

A new privacy law is enacted and will apply to the European Union in May 2018. Currently, all member states have their own legislation, though all implementing a common Directive. From 2018 onwards, all countries will have the same law. There will be an EC Data protection group to help with the interpretation of the law. Persons will have the right to get data deleted. Pseudo-nomisation is introduced, where someone in the company has the key between the original data and the anonymized data, whereas everyone else only can access the anonymised data. There are legal and cultural differences between member states today, after 2018 it would be easier to file a privacy case in another country as the law is the same. There will be a fine up to 4% of gross profit in case of violation of privacy.

Protecting data privacy includes data protection, anonymization and feature extraction. Anonymization can be made back-office or by design. It depends on possibilities for user consent if anonymization by design might be the only alternative. Manual anonymization includes extensive efforts. Therefore, automated methods are needed and are being developed.

There is a need to identify the essential data that should be kept. An ISO group is focusing on this topic. E.g. emotions of the driver, head and eye movements, body movements and tasks are of interest. For external videos, traffic scenarios and interactions are most interesting.

There are different methods for making video annotations from blurring to avatars and masked faces. In the validation of the processed data, it is important to verify that the quality of the annotation is high enough and for this, a large amount of manually annotated data is

necessary. Thus, validation is a time-consuming effort. Possible final gains still motivate the initial efforts.

People act differently and can therefore be identified by other means than faces. The big question is how far do we have to go with anonymization? If people are sharing their lives on social media, why do we need to anonymise the research data? It is different to have one snapshot than the whole mobility pattern of people. We have a responsibility by law to protect the privacy in the data. Therefore, we need to find ways to anonymise or un-identify the data, while still keeping the necessary data for research or development of e.g. vehicles.

2.5 US perspective on anonymization, Hannah Rakoff (VOLPE, US DOT)

Hannah Rakoff gave the presentation on the US perspective on anonymization; operationalizing privacy. The ITS Joint Program Office has a special mission to work on this subject. Advancing automation, connected vehicles and smart cities are their strategic priorities. The data is the foundation of the ITS strategic plan. Enterprise Data Program is funding this research.

Data Program approach will enable data use by early deployers (CV-Pilots, Smart City Challenge), develop tools (Operational Data Environment, ITS Privacy Module algorithm, ITS Privacy Sensitivity Model, ITS Data Policy Playbook), and share data (RDE, National Strategy).

It took more than 2 years to anonymise one month of Safety Pilot data, thus, there is a need to make anonymization easier and cheaper. Tools will be provided. A focus of the connected vehicle anonymization has been on archiving, particularly Basic Safety Messages (BSMs). “Privacy by design” must be applied to the single file level to make it anonymous. The aim is to make it more operational and efficient to also be applicable to real-time data.

The new challenge is how to share operational (real-time) data that originates at state and local level. The data program philosophy is to address technical and policy/institutional issues, to set priorities and make product development driven by user needs, and to have many program products ultimately handed-off to user communities.

The objective of the ITS Privacy Module is to remove enough data to prevent derivation of a specific traveller’s origins and destinations while maintaining the value of the data (trade-off). The approach focuses on truncating data at appropriate junctions. Initial focus is on BSMs.

The ITS Privacy Sensitivity Model is a tool for assessing the trade-off between data utility vs. privacy protection. Users can adjust privacy algorithm parameters and see the implications for the resulting (anonymised) data.

ITS Data Policy Playbook is an online collection of policies, principles and real-world examples of successful policies and practices, such as sample data retention policies, open source procurement language, and guides to conducting privacy impact assessments and staying up to speed on current standards. The idea is to hand-off the long-term maintenance of the system to the deployment community.

Hannah also showed an example of video anonymization funded by the DOT Exploratory Advanced Research (EAR). The algorithm performs a one-way merge of the pixels of the original face with another face. It is partly done manually.

2.6 Differential privacy for data mining and querying, Devdatt Dubhashi (Chalmers)

Devdatt Dubhashi gave a presentation on differential privacy for data mining and querying. The topic of privacy has become an important issue in computer science, with the rapid increase of the use of personal information for different purposes in a variety of organisations, ranging from business organisations to authorities.

Netflix announced a 1 M\$ prize for improvement of their recommendation system. Data regarding their customers were released and researchers were able to de-anonymize it completely by connecting it to another dataset. It was powerful example of the issues associated with privacy.

Privacy enhancing technologies include differential privacy, data minimisation and differential privacy in programming languages. Apple has announced that they do research on differential privacy. In differential privacy methods noise is added to data. You release probabilistic data. There is a trade-off between utility and privacy. Machine learning can be applied also in a way that privacy is not compromised.

Data minimisation is based on the fact that you only collect information that is relevant, not more or with more precise accuracy. This kind of approach is challenging if you do not know how you will use the data or what features in the data you will need. Several examples on privacy respecting programming systems were given.

There are different levels of randomness in adding random noise. To keep the original data is a separate question to making anonymised data available to others. All data does not need to be anonymised (internal data), only when it is shared or made public.

2.7 Computer resources for large-scale anonymization/feature extraction, Ann-Charlotte Sonnhammer (SNIC, Uppsala University)

Ann-Charlotte Sonnhammer gave a presentation on computer resources for large scale anonymization/feature extraction. SNIC is a Swedish research infrastructure with national funding that provides resources for research. The amount of collected data is increasing fast. Privacy and security principles must be followed and specific e-infrastructure was established to store sensitive data.

The data is classified based on confidentiality, data integrity, and availability. Risk and security assessment is made including confidentiality, data integrity, traceability, and availability. Information security courses are given for system administrators, staff and researchers. The technical solution for the research includes strong authentication, logging of activities, encrypted data transfers and avoiding data leakage from and within resource. Support with legal matters and contracts are also given, such as to write and finalise contracts.

Regarding the protection of the data, it is easier to deal with technical issues compared to finding solid solutions to make people understand the necessity to secure the privacy of persons. Sometimes the data users do not fully comply with the rules for data use.

2.8 Video anonymization of vehicle environment, Jan-Erik Solem (Mapillary)

Jan-Erik Solem gave a presentation on video anonymization of vehicle environment. He has previously developed face recognition for mobile, web and cloud areas at Polar Rose, a company later sold to Apple. Mapillary is a service for crowd-sourcing street views for maps, to allow for continuous updates. People upload images and videos, which are automatically

anonymized before published. Different rules apply for data collection inside the vehicle (consent needed) vs. outside the vehicle (face blurring, license plate blurring).

There is a quick development in face and feature recognition. You should keep the original data, as in a couple of years you can do so much more with the data compared to what you can do today. This also implies that persons, sufficiently anonymized today, can be re-identified in a couple of years.

Face recognition based on facial features is possible when the group of people is small but the performance gets worse as the pool of people gets larger. One-to-one verification works well. Typically a sequence of images would provide better recognition but several frames are not used unless liveliness test is needed.

People can recognise people without seeing the face, on movement or body pose. For a machine it is trickier. Machines (artificial intelligence) nowadays are much better than five years ago. Video-blurring in real-time is possible. From video the un-blurring is easier than from still-photos.

The participants discussed identification of people based on different features and patches of images or additional data that you could link to. Adding more data sources makes it easier to identify the anonymized people.

2.9 Open discussion, John-Fredrik Grönvall (Volvo Cars) and Helena Gellerman (SAFER)

What is good enough anonymization?

In the US, large efforts are used to estimate how good the current methods are, to decide what should be the target level. It won't be possible to guarantee 100% privacy.

Legal approach is one thing, respecting the customers may require a different level of anonymization. This also applies to test participants.

'Good enough' could be when you cannot identify someone easily. If you use a lot of resources, you may still be able to de-anonymise a picture/video. Sensitive parts of the data need to be anonymised as early as possible as a leak might appear somewhere. There is not an absolutely secure place for data storage.

It is important to work together in this field to progress. The US is especially interested in methods to anonymise real-time trajectory (GPS) data. E.g. FOT-Net's recommendations on participant agreements are of common interest, what is good enough anonymization and operationalisation of anonymization as well.

Anonymization as close as possible to the data source is a good take-back for us all.

6.3 Day 2, Thursday, 1 September 2016, 09.00–12.30

3.1 Re-cap of Day 1, Satu Innamaa (VTT)

Satu Innamaa did a re-cap of Day 1.

3.2 Feature extraction – how to assure you can delete data without information loss, John Lee (University of Wisconsin-Madison)

John Lee gave a presentation on feature extraction and how to assure you can delete data without information loss. He presented a list of projects that have developed tools for video

reduction, funded by FHWA and targeted for SHRP2 NDS data. This presentation shows the results from the tool project awarded to Wisconsin.

The project has developed a tool that provides fast, but imperfect (good enough?), facial landmark info from difficult video. It shows that it is feasible to code all SHRP2 face video. The following aspects need to be considered: features should capture the essence of video, but the essence of the video may depend on the analysis purpose; video quality governs algorithm selection; multi-level visualization is used to identify pathologies; and performance metrics include accuracy and confidence calibration.

Feature coding was used to quantify distraction and engagement by applying video analytics to NDS data. GPS data was utilised in the process. It is important that the things that you code are generic i.e. can be used for many purposes. Still, is it possible to extract purpose-specific features from general features, such as FACS codes, percentage of eye closure, facial tone or phone-to-ear from landmarks?

The SHRP2 video includes some challenging elements such as low video quality, motion blur, low light and over exposure, unpredictable and significant illumination changes, non-frontal head pose and occlusion. If video analytics could be applied to the complete video dataset of SHRP2, and non-privacy data extracted, the video data would become more useful.

The coding strategy is based on branching cascaded shape regression, where a simpler regression problem is solved in each step. The coded facial features include head bounding rectangle, 68 facial landmark coordinates and visibility values, confidence indicator, head orientation and openness of the eyes and mouth. Performance of the algorithms was tested and found to be better than state of the art. The method is 200 times faster than real-time. The algorithm is hosted at a private GitHub repository. Send email to John Lee if you want to get access.

John Lee presented the results of the assessment of the confidence and performance of the distraction prediction based on facial feature data. They have a macro-micro visualization tool for the validation. Going from overall picture, you need to also take a look at individual frames to see imperfections.

Future challenges and opportunities include shared database of standard annotated video data for video algorithm validation and benchmarking; a comprehensive catalogue of generic features to satisfy all potential purposes; error propagation to estimate what is good enough for a specific purpose; train algorithm for accuracy and confidence calibration; accuracy-confidence calibration tuning to define Pareto frontier of algorithm performance; and define and support computer-aided coding strategies with points on Pareto frontier of accuracy-confidence calibration.

To save data of the eyes would not risk privacy, thus it could be saved to get more information on the eyes. If the camera was located directly in front of the driver (above the wind screen or below) some problems would be solved directly.

Using automated video annotation for feature extraction, Flora Dellinger (WASSA)

Flora Dellinger gave a presentation on how to use automate video annotation for feature extraction. The goal was to produce a proof of concept of automated video annotation for secondary task studies on NDS data. The purpose was to extract automatically some specific

events linked to secondary tasks for more efficient manual annotation of start and end of task.

Events to be annotated automatically included position of feet on pedals, presence of hands on the steering wheel, texting on a phone, conversation with a hand-held phone, hand-free conversation and talking with a passenger. A software displaying the annotations on the videos should also be developed. Automatic video annotation algorithm included phases such as cropping of image region, computing of a feature image descriptor, application of a machine learning algorithm called SVM, and finally a decision. Deep learning algorithms with Tensorbox was used, where a framework trains a convolution neural network to detect the objects in images.

Performance of the Proof of Concept: Feet on pedals with precision 99.7%, recall 95.5%, passenger detection with precision 95.6% and recall 99.8%. Precision was lower for detection of hands on wheel (true 85%, false 3.1%), texting (true 50%, false 15%) and hand-held phone (true 72%, false 34%).

Work in progress includes speech detection, phone detection on right/left ear, and number of hands on wheel. Neural networks are also trained with more data and improved post-process filtering is applied to obtain consistent sequences.

A tool has been developed to visualise the result of annotation.

3.3 Automated labelling and recognition, Jan Erik Solem (Mapillary)

Jan Erik Solem gave a presentation on automated labelling and recognition, based on the work done at Mapillary regarding crowd-sourcing images and videos world-wide to provide open source map attributes. Currently street-view photographing is made in silos and the data is not shared nor can be used. Mapping is extremely labour intensive.

Mapillary is trying to solve scalability in the image collection and in the data extraction. They have one single platform where anyone can contribute by uploading videos or images of their surroundings. Map data is automatically generated within hours, including automatic blurring of personal data such as faces and license plates.

The business model is to charge for the data for business use. Otherwise everything is open and free. You can query on whatever you want. Also timewise differences can be obtained (changes made after certain day).

The basis is a 3D model where new images are combined with already existing images. No sensitive information is included. Semantic segmentation is also made to distinguish roadways, sideways, trees, buildings, pedestrians, bicycles, etc. Traffic signs are detected automatically and their positions can be downloaded.

Go to mapillary.com to see and play with the tool. You could also contribute to the maps by downloading the app and provide images.

3.4 Open Discussion, Marco Dozza (SAFER)

Marco Dozza hosted an open discussion with the presenters of day 2 and the audience.

Where is the automatic annotation in a couple of years from now? It is developing fast and many organisations are developing algorithms. The quality of the data is though a challenge.

Can we use data for which a consent form approves only one topic, for something else in anonymous form? If anonymization is made in-vehicle, we may be able to use the data for whatever. If raw data is transferred, maybe it cannot be used unless it is anonymised. But as anonymization may be reversible, that creates a problem. For similar purpose you may always use the data, but how to identify “similarity”.

Core question is: “What is anonymous enough?” We must take into account the ethics of research in addition to the legal aspects.

Keeping the raw data creates problems although it is needed for certain research and development. Customer should always be informed what is logged.

3.5 Wrap-up, Helena Gellerman (SAFER)

Helena Gellerman concluded the workshop. We had listened to some very promising progress on feature extraction – the core of anonymization of data. If we are to be able to protect personal data, still extracting the information needed for research and developments, feature extraction is the key.

The question still remain though, what is anonymized enough, taking into account the progress of technology where a level that is enough today can reveal personal data with more powerful tools tomorrow. Should data be continuously re-anonymized?

Helena thanked all the speakers and also the audience for a very interactive workshop. She invited the participants to send inputs for a possible new workshop, as the attention on anonymization will continue to grow the coming years.

6.4 Participants FOT-Net Data workshop, 31 August – 1 September 2016, Gothenburg

| Last name | First name | Company |
|-----------|--------------|-----------------------------------|
| Abraham | Michelle | Chalmers tekniska högskola AB |
| Agerholm | Niels | Aalborg University |
| Antignac | Thibaud | Chalmers University of Technology |
| Apoy | Camilla | Autoliv |
| Buendia | Ruben | SAFER |
| Carlsson | Anders | Volvo AB |
| Chen | Fang | Chalmers, Apply IT |
| Dellinger | Flora | Wassa |
| Dozza | Marco | Chalmers |
| Dubhashi | Devdatt | Chalmers |
| Ebadi | Hamid | Chalmers |
| Fahlvik | Beata | Chalmers University of Technology |
| Flament | Maxime | Ertico |
| Forsberg | Christian | Autoliv |
| Gellerman | Helena | SAFER |
| Gjertz | Mats | Volvo Cars |
| Gryte | Anders | SmartEye |
| Grönvall | John-Fredrik | Volvo Cars |

| Innamaa | Satu | VTT |
|--------------|---------------|---|
| Johansson | Emma | Volvo Group Trucks Technology ATR |
| Karlsson | Johan | Autoliv |
| Koskinen | Sami | VTT |
| Kovaceva | Jordanka | Chalmers |
| Krishnan | Amrit | Autoliv |
| Larsson | Jonathan | Autoliv |
| Lee | John | University of Wisconsin |
| Lindman | Magda | VCC |
| Lindström | Konstantin | Volvo Cars |
| Liu | Feng | Volvo Cars |
| Nabeel | Mohanad | Chalmers |
| Nilsson | Emma | Volvo Cars |
| Nilsson | Johan | Denso |
| Nilsson-Ehle | Anna | SAFER |
| Olsson | Jan | Eget bolag |
| Othmezouri | Gabriel | Toyota |
| Pirnia | Elham | Volvo AB - ATR |
| Quintero | Karla | CEESAR |
| Raisch | Florian | BMW AG |
| Rakoff | Hannah | U.S. Department of Transportation (USDOT) |
| Schneider | Gerardo | University of Gothenburg |
| Selpi | | Chalmers |
| Solem | Jan Erik | Mapillary |
| Sonnhammer | Ann-Charlotte | SNIC, Uppsala University |
| Svanberg | Erik | SAFER |
| Terada | Ryota | Denso Sales Sweden AB |
| Tivesten | Emma | Volvo Cars |
| Ustrabowski | David | Chalmers |
| Weiss | Jens Peter | BMW AG |
| Wells | Peter | Volvo Technology AB |
| Wiklander | Marianne | Volvo Cars |
| Zahirovic | Fikret | Chalmers tekniska högskola AB |

7 FOT-Net Data Workshop 4: Hands-on workshop on Data Re-use

On 14 September 2016, a workshop was held on different locations, using online tools. The workshop aimed to provide hands-on experience with the re-use of FOT data. Both US and European datasets were discussed, research questions were developed which may be answered using available data, and exchanging experiences with the re-use of data.

The workshop consisted of an alternation of plenary webinars with presentations from speakers, and hands-on group work at different locations. Next to physical locations, one group conducted their working group online. In one of the small groups, a remote participant joined in.

The workshop took place at several locations:

- IKA, Aachen, Germany
- VTT, Helsinki, Finland
- SAFER Vehicle and Traffic Safety Centre, Gothenburg, Sweden
- University of Leeds, Leeds, UK, using GotoMeeting

The central webinars were moderated by ERTICO, Brussels, Belgium, using GotoWebinar.

7.1 Agenda

The agenda was as follows:

1. Local introductions: Aachen, Helsinki and Gothenburg

Plenary webinar:

2. Introduction to the workshop, Yvonne Barnard (University of Leeds)
3. Introduction to FOT-Net Data and the FOT-Net wiki, Satu Innamaa (VTT)
4. Introduction of the US Research Data Exchange (RDE), Richard Glassco (Noblis, USA)
5. Data sharing, Helena Gellerman (SAFER)
6. FESTA methodology, Yvonne Barnard (University of Leeds)
7. Meet the four groups, introduction by the location leaders:
 - Aachen, Adrian Zlocki (IKA)
 - Helsinki, Satu Innamaa (VTT)
 - Gothenburg, Ines Heinig (SAFER)
 - Gotomeeting group: Haibo Chen and Tony Jia (University of Leeds)

Group work at the different locations:

8. Activities and discussions

- o General introduction
- o Demonstration of RDE, explore available datasets
- o Inspect some US and European datasets and meta-data
- o Discussion on types of research questions the datasets could answer

Lunch Break

Plenary webinar:

9. Short presentations of morning discussions by four location leaders
10. Presentation of data re-use of RDE data, Richard Glassco (Noblis, USA)
11. Questions and answers

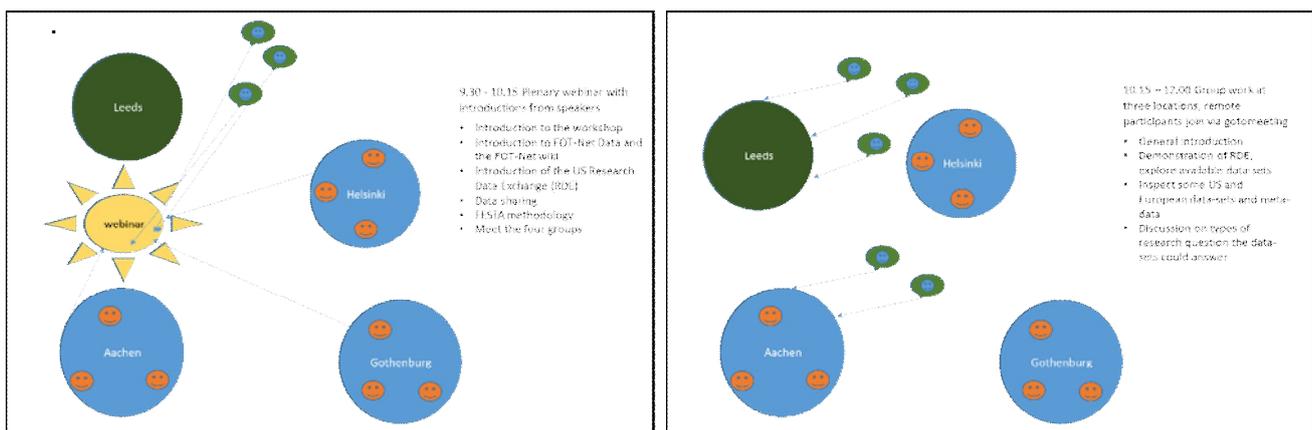
Group work at the different locations:

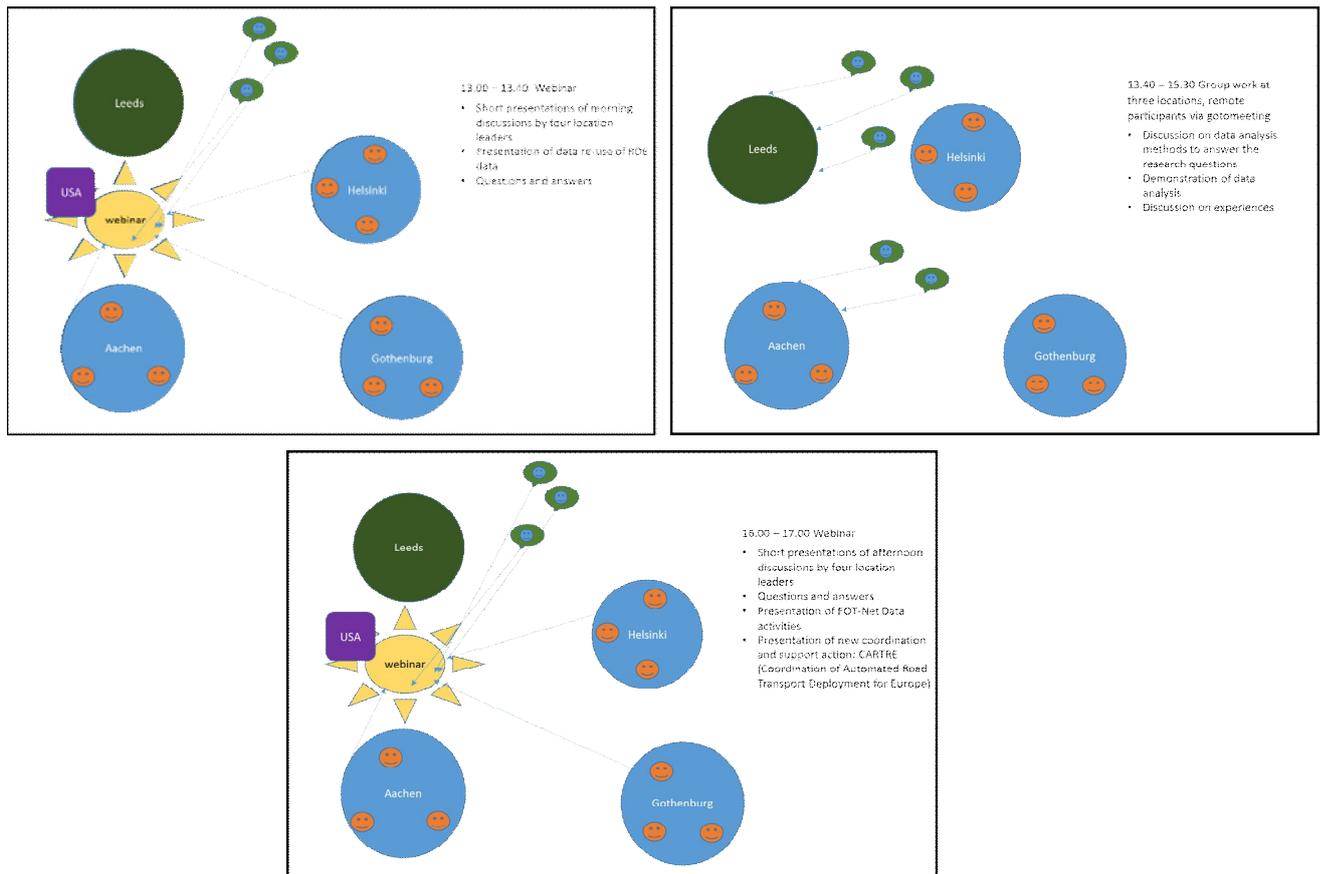
12. Activities and discussions
 - o Discussion on data analysis methods to answer the research questions
 - o Demonstration of data analysis
 - o Discussion on experiences

Plenary webinar:

13. Short presentations of afternoon discussions by the four location leaders
14. Questions and answers
15. Presentation of FOT-Net Data activities, Sami Koskinen (VTT)
16. Presentation of new coordination and support action: CARTRE (Coordination of Automated Road Transport Deployment for Europe), Maxime Flament (ERTICO – ITS Europe)

Below it is illustrated how the day was organised:





7.2 The plenary program

1. Local introductions: Aachen, Helsinki and Gothenburg

The people participating physically were welcomed at the three locations and information was given about the institutes in Aachen, Helsinki and Gothenburg.

Plenary webinar

2. Introduction to the workshop, Yvonne Barnard (University of Leeds)

The plenary webinar started with an overview of the day's agenda.

3. Introduction to FOT-Net Data and the FOT-Net wiki, Satu Innamaa (VTT)

Satu Innamaa briefly described the FOT-Net Data project and introduced the FOT-Net wiki (<http://wiki.fot-net.eu>).

4. Introduction of the US Research Data Exchange (RDE), Richard Glassco (Noblis, USA)

A pre-recorded presentation (due to the time-difference) was given by Richard Glassco on the Research Data Exchange (RDE, <https://www.its-rde.net>). The purpose of RDE is to provide a variety of data-related services that support the development, testing, and demonstration of multi-modal transportation mobility, weather, and environmental applications. Objectives are: Enabling systematic data capture from connected vehicles, mobile devices and infrastructure, performing data quality checks, and providing clean, well-

documented datasets, and integrating data from multiple sources into data environments. RDE features are:

- Advanced Search Capabilities
- Multiple File Download Capability
- Playback Data Delivery
- FAQs
- Contact Information
- Standard Metadata Documentation
- Map of Data Location
- Registered Research
- Projects
- Sample Data Files

New datasets are added to RDE as they become available: from other USDOT projects or from outside submissions. ITS researchers are invited to submit datasets to RDE. Richard showed several example pages of RDE as a starting point for the work in smaller groups.

5. Data sharing, Helena Gellerman (SAFER)

Helena Gellerman briefly described the need for data sharing and the data sharing framework that is being developed in FOT-Net Data.

6. FESTA methodology, Yvonne Barnard (University of Leeds)

Yvonne Barnard gave an overview of the FESTA methodology, specifically for those who are not very familiar with Field Operational Tests and the methodology.

7. Meet the four groups, introduction by the location leaders

The four location leaders introduced themselves and their groups. Adrian Zlocki (IKA) in Aachen, Satu Innamaa (VTT) in Helsinki, Ines Heinig (SAFER) in Gothenburg, and Haibo Chen and Tony Jia (University of Leeds) with the GotoMeeting group.

8. Group work at the different locations

The four groups all had their own program, see the reports below. In general the following activities were performed: demonstration of RDE, exploration of available datasets, inspection of some US and European datasets and metadata, and discussion on types of research question the datasets could answer.

Plenary webinar:

9. Short presentations of morning discussions by four location leaders

In the afternoon plenary session the leaders reported back on their morning's activities. See their reports below, for more details see their presentations at the FOT-Net website.

10. Presentation of data re-use of RDE data, Richard Glassco (Noblis, USA)

Richard Glassco was able to join us remotely now, and he gave a further and more detailed presentation of the re-use of RDE data. Richard first explained datasets are included in RDE. The process for evaluating data and preparing it for posting is:

- Assess sample data for quality, completeness, value for research, and presence of sensitive data
- Assess full dataset for same issues
- Design logical structure for data file organization, and comprehensive documentation
- Add descriptions and search tags and post on the RDE
- Publicize to registered and potential users

The dataset is checked for personally, commercially and research sensitive data. Before data can be distributed to the public, personally sensitive data must be removed while maintaining usefulness of the data. He explained how trajectories may be anonymised. Other sensitive data such as commercial operations may not be distributed to the public. RDE is building visualization and playback tools to facilitate data discovery and use while preventing release of sensitive information. Richard provided several examples of the use of RDE data. RDE is used more and more, in the second quarter of 2016 1725 users were registered. Data from the RDE has been downloaded hundreds of times by hundreds of researchers from all sectors; some of them have shared what they have done with the data, including: congestion analysis, safety analysis, and driver behaviour analysis.

11. Group work at the different locations

After the plenary webinar participants again worked in smaller groups (see reports below). Different groups had different focus but in general they discussed data analysis methods to answer the research questions and short demonstrations of data analysis were given. Finally the participants' experiences from their workshops were discussed.

Plenary webinar:

12. Short presentations of afternoon discussions by the four location leaders

The location leaders reported back on their morning's activities. See their reports below, for more detail see their presentations at the FOT-Net website.

13. Presentation of FOT-Net Data activities, Sami Koskinen (VTT)

Sami Koskinen provided an overview of the past and upcoming events of FOT-Net Data.

14. Presentation of new coordination and support action: CARTRE (Coordination of Automated Road Transport Deployment for Europe), Maxime Flament (ERTICO – ITS Europe)

Finally, Maxime Flament and Davide Brizzolara presented the new coordination and support action CARTRE, starting in October 2016, funded 3 m€ by the EC, with a duration of 24 months. The main objectives are to:

- Establish a joint stakeholders forum in order to coordinate and harmonize ART approaches at European (e.g. strategic alignment of national action plans for automated driving) and international levels (in particular with the US and Japan)
- Create a solid knowledge base of all European activities, support current activities and structure research outcomes by enablers and thematic areas
- Setup a platform for sharing and re-using data and experiences from different ART
- Actively support Field Operational Tests (FOT) and pilots carried out at National and European levels
- Work on future visions, potential impacts and research gaps.

7.3 Report from Aachen

Group work morning

At this location there were six participants, one participant joined remotely. Adrian Zlocki provided a general introduction starting with a round table, FOTs, FOT-Net Data and different datasets (focusing on euroFOT data).

Yvonne Barnard showed how to use the RDE datasets. A dataset and metadata description was downloaded and demonstrated to all participants. A pre-downloaded dataset was inspected and all possibilities of data reuse were described to the participants.

Adrian Zlocki presented the structure and the content of the former euroFOT database. He described the data collection process and the data storage. The participants were given insight into which data was collected, and how the data was made available for data analysis.

Group work afternoon

Adrian Zlocki presented basic results of data evaluation from the euroFOT project. The analysis was based on the datasets and database which were presented in the morning. Therefore the participants could understand how FOT data was used and which kind of research questions could be applied to the available data. The analysis was the starting point for discussions on possible data analysis and data re-use.

The participants were interested in the evaluation of available data (re-use) for new ITS deployment. For new measures to be introduced, like e.g. usage of hard shoulder, a network management system could profit from data re-use if the “right” data is available. The FOT-Net wiki and Data Catalogue was introduced and presented as one possible source of available datasets.

Data re-use for automated driving was discussed. Especially issues with regard to the baseline for an evaluation of an automated driving function are of interest. Available data might be re-used as baseline data. The euroFOT dataset would be of interest concerning such baseline data. Furthermore, the following questions were addressed: Do we need a driver who is accurate in the same manner as the system? What about safety systems which have high penetration rates these days due to Euro NCAP? Who is the average driver? Can we find an average driver? Automation may be able to deal with incidents.

Datasets may become obsolete. The issue of dataset age was discussed as the real world situation may change.

For C-ITS evaluation, results for measures which are deployed are difficult to find. Not much data seems to be available.

Finally different driver profiles and the availability of driver profiles in existing data were discussed. The availability of different profiles in existing datasets needs to be investigated.

7.4 Report from Helsinki

The local session at Helsinki was targeted to Aalto University PhD students and researchers, especially students of transport planning. In total six persons participated in the workshop. At the start of the workshop the FOT aspects they were primarily interested in were future automated driving FOTs, driver attention monitoring techniques (e.g. eye tracking), and

combining FOT data with their usual data sources. As material for workshop discussions the participants were specifically presented with the TeleFOT and Trafisafe FOT datasets.

Regarding new research questions for FOT data re-use the group became especially interested in using travel data to assess the potential for future MaaS (Mobility as a Service) services: FOT data could provide a starting point for assessing current monthly travel costs for different family sizes and different vehicle types. These current costs could be compared to MaaS pricing models that are suggested and where a monthly fee paid to a service operator could cover most of the family travel.

Also, for example, the costs of shopping trips came up. Using FOT data it might also be possible to provide input to a comparison of family and societal costs related to shopping in nearby small shops vs. shopping in far-off supermarkets. Privacy aspects of mobility research was also discussed.

A clear benefit from the session was that FOT-Net was able to make contact with students who would not have been likely to travel to another country for a workshop. Further local cooperation could also be planned, e.g. related to upcoming automated driving test campaigns and sharing test tools.

7.5 Report from Gothenburg

The local session at Gothenburg was open for researchers who were either new to the field of NDS and FOT data or wanted to learn more about available datasets and data re-use; there were six participants. Topics discussed here included a summary of previous and ongoing NDS and FOT activities at SAFER, possibilities and challenges when answering several research questions by re-using different types of datasets as well as hands-on data experience. The datasets used during the day were Safety Pilot data available at the RDE site and euroFOT data available at SAFER.

euroFOT data include videos for both the outside environment of the car, and the inside view including driver behavior and gaze. The workshop participants were given an introduction to SAFER's specific tool for video annotations, and hands-on experience when creating their own annotation schemes as well as annotating sample videos.

In the next session participants went through a complete chain of analysis steps from accessing a database over selecting, filtering, processing and merging data to calculating performance indicator and answering hypotheses. Special attention was given to the differences between datasets and their related usability and applicability to certain research questions and areas. The practical session was concluded by presenting some analysis results from previous projects.

Finally the IGLAD initiative for the harmonization of global in-depth traffic accident data was introduced in order to make people aware of closely related datasets, inspire cross-research area analysis, and provide a bigger picture. This was especially appreciated by the participants.

7.6 Report from online workshop Leeds

At the University of Leeds a GoToMeeting session was hosted where seven participants formed an on-line working group.

The morning session started with an introduction and the organisers showing how to open Research Data Exchange (RDE) datasets. They gave a brief overview of the types of data, and discussed relevant projects included in RDE. Participants were encouraged to access the RDE datasets themselves, and download and explore sample datasets which could be used for their own research interests. Participants were also asked to propose questions regarding the use of RDE and to be discussed later.

The afternoon session focused on illustration and discussion of specific research issues. The organisers first demonstrated an example of travel time estimation using Bluetooth data, to show some basic statistical methods of datasets analysis. Then the participants joined in the wider discussion on research questions, such as the distribution of driving cycle on approach and departure at a traffic signal, different distributions of incident durations in different cities, and the evasive reactions by drivers to neighbouring vehicles on the highway. They also shared opinions on the issues of the potential utilization of RDE data on traffic research development (e.g. Can RDE data be used to detect drivers' response to road marks and signs? Can it be used to update the standards of road design?) as well as the global sharing of datasets.

The workshop inspired cross-area experts becoming aware of transport datasets opening up, and also created potential opportunities for future research collaboration, e.g. on scientific papers.

7.7 Conclusions

This multi-site and multi-online-session workshop attracted 27 people from Europe, USA and Japan. It encouraged local participants who have no budget to travel (e.g. PhD students) to attend the local physical meeting and learn from other remote sites via online facilities.

Several popular FOT datasets were introduced, analysed and discussed in parallel and experience and lessons learned were shared among the sites and with remote participants. It was a hands-on experience, a cost-effective meeting and a dynamic event during which the back and forth activities between online and offline sessions were very engaging and interactive, with a lot of questions, wonders, laughs as well as jokes.

Thus, the new concept of multi-site and multi-online-session workshop was proven successful. Yet, a lesson learned was that a half-a-day workshop might have attracted more people. As we were operating on several time zones, either the beginning of the full-day workshop is early in the Western Europe (now: UK) or the end of the workshop is late in the Eastern Europe (now: Finland). With half-a-day workshop the timing can be made better for all.

Nevertheless, the concept worked and the participants enjoyed being able to participate in an international event. Without the multi-site and multi-online-session concept they would not have been able to participate the workshop at all. Thus, this is a way to disseminate knowledge and engage people that are hard to reach otherwise.

8 Hands-on Workshop on Data Re-Use, Guangzhou Session in China

This workshop was held in South China University of Technology, Guangzhou, China, 12/12/2016, and moderated by Tony Jia (University of Leeds) as the session extension of the virtual online workshop on 14 September 2016. This session was targeted to the PhD/Master students and researchers from local universities, with 8 persons participating in.

The workshop started with the introduction of FOT-Net Data project and the open online dataset of Research Data Exchange (RDE). Participants showed great interests in how to access the RDE data sets and to re-use it in academic researches. The issues regarding opening data sets to the public was also discussed in particular, as normally these data sets are regarded as the core value especially to private enterprises, not only in the transport area, but also in other industry fields, e.g. the manufacturing industry. The participants agreed that the government and authorities should play a core role in encouraging more data sets becoming accessible to the public.

An example of re-using the VTT Trafisafe data set was presented in order to demonstrate driving behaviour analysis using different variables with different analysis methods. Regarding new research questions, the group discussed topics ranging from data acquisition, data processing to decision making. Some interesting questions included: potential data collection methods being applied in transport, the challenge of transport performance evaluation with limited sample data (e.g. Bluetooth data) with data loss and noise, improving context awareness with the constraints of data privacy and security, etc.

The workshop built up the link to universities of China, and also provided the potential to further research cooperation, e.g. funding applications.