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

Application of Data Sharing Framework in Selected Cases

Deliverable no.	D4.3
Dissemination level	Public
Work Package no.	WP4
Main author(s)	Helena Gellerman (SAFER)
Co-author(s)	
Status (F: final, D: draft)	F
Version number	1.0
Date	2 February 2017
Project Start Date and Duration	January 2014, 36 months

www.fot-net.eu

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 610453

Document Control Sheet

Main author: Helena Gellerman (SAFER)

Work area: WP4

Document title: Application of Data Sharing Framework in Selected Cases

Version history:

Version number	Date	Main author	Summary of changes
1.0	2.2.2017	Helena Gellerman	Final

Review:

Name	Date
Satu Innamaa	1.2.2017

Circulation:

Recipient	Date of submission
EC	2.2.2017

Table of Contents

Executive Summary

A large number of Field Operational Tests (FOT) and Naturalistic Driving Studies (NDS) have been performed worldwide, resulting in a variety of datasets. The interest in re-using the data is rising, as knowledge of the content of the datasets is being spread. Safety is still the predominant research area that drives data analysis, but wider usage covers a vast number of research topics. This includes topics such as further development of safety performance measures, driver distraction, infrastructure analysis and fuel economy.

Re-use is taking place on a global scale, still the U.S. is predominant. Also in Europe, the data re-use is substantial, with a focus on the larger datasets collected in the Europe and in the U.S. The organisations re-using data are mainly universities and vehicle manufacturers, but also other private firms, public health organisations and laboratories work with the data.

This all implies the importance of making naturalistic datasets widely known, for each researcher to figure out how the datasets collected globally could be used in their respective research area. Another important area is to investigate the factors why not more datasets are provided for re-use and to provide assistance to overcome the identified obstacles.

The overall aim of the task WP4.3 was to facilitate the provision of collected datasets and to make a few interesting and high-priority datasets consistent with FOT-Net´s Data Sharing Framework (DSF) and ready for re-use. The findings from the investigations have fed into the final version of the DSF and also provided datasets for re-use.

Three datasets, the Australian Naturalistic Driving Study (ANDS), the European naturalistic driving study UDRIVE and the Danish ITS Platform were chosen to receive funding to facilitate future re-use of the data. The ANDS dataset contains the same data as the SHRP2 project and is therefore highly interesting as similar tools can be applied. The ANDS project was found to have implemented many of the DSF recommendations, still the DSF provided guidance when preparing for third party data sharing.

UDRIVE has implemented same or similar procedures as is recommended in the DSF throughout the project. UDRIVE has already implemented the full set of data sharing procedures as twelve analysis sites are remotely performing analysis on a the same dataset. The hands-on experience from the project was fed back to the final version of the DSF.

The last interesting case incorporating GPS, ITS Platform, developed the necessary procedures, including anonymization method, to be able to provide the data on-line for free.

This report includes the full reports regarding each of these datasets. It also contains a smaller pilot on the Trafisafe dataset that was done to investigate what documents were needed to provide such a small dataset.

Finally, U.S. DOT provided a report comparing the FOT-Net documents to similar U.S. frameworks. The conclusion was “this review make clear that FOT-Net (as it is embraced by CARTRE) and the ITS JPO can benefit from continuing to learn about each other’s policies, practices, and procedures regarding data sharing, protection and related topics. It is recommended that this comparison of practices periodically be updated and shared with both parties and the broader international community.”

1 Introduction

1.1 Enhancing the availability of collected datasets

A large number of Field Operational Tests (FOT) and Naturalistic Driving Studies (NDS) have been performed worldwide, resulting in a variety of datasets. These datasets are collected to answer research questions formulated within the projects, but the potential for re-use to further serve the research community is large. Several datasets are re-used already today, where large-scale datasets are pre-dominant, such as SHRP2, euroFOT and datasets collected with event data recorders.

The interest in re-using the data is rising, as the knowledge of the content of the datasets is being spread. The initial main interest in the data and the driving force in collecting these datasets came from the safety research, understanding crash causation and the impact of safety functions. This is still the predominant research area that drives the data usage, but the complete usage covers though a vast number of research topics. This includes topics such as further development of safety performance measures, driver distraction, infrastructure analysis, in-depth look at younger and older drivers, driver fatigue and impairment, roadway lighting, fuel economy and pedestrian/vehicle conflicts. As the research area is still fairly young, the re-use of the data for development of data analysis tools and methods to make the data easier to use is large.

The re-use is taking place on a global scale. Still the U.S. is predominant in re-using the data on a larger scale, as they have been in the forefront of collecting large datasets that are also provided for re-use after the projects. Also in Europe, the data re-use is substantial, with a focus on the larger datasets collected in the U.S. and in Europe. The re-using organisations are mainly universities and OEMs, but also other private firms, public health organisations and laboratories work with the data.

This all implies the importance of making naturalistic datasets widely known, for each researcher to figure out how the datasets collected globally could be used in their respective research area. Another important area is to investigate the factors why not more datasets are provided for re-use and to provide assistance to overcome the identified obstacles.

The purpose of this report is to describe the result of the work of facilitating the provision of collected datasets. The overall aim was to make a few interesting and high-priority datasets consistent with the data sharing framework and ready for re-use. The report documents the selection criteria for the prioritization of which datasets to fund. It also addresses what could be funded and the meaning of being consistent with the DSF. The focus of the report is though the three datasets that were chosen for funding, the Australian NDS, the European project UDRIVE and the Danish ITS platform. The selection process is described followed by a description of the data, the work done, conclusions and a common lessons learnt. Even though not funded, a report from the U.S. DOT is also included. The full reports of the performed tasks per dataset are included as annexes.

1.2 FOT-Net Data Project

FOT-Net is a networking platform open to all stakeholders interested in FOTs. It was established in 2008 as a European support action to let FOT experts benefit from each other's experiences as well as to give an international dimension to local activities. It organizes international workshops, publishes a series of newsletters and promotes FESTA – a European handbook on FOT methodology.

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. FOT-Net Data is a continuation of FOT-Net's activities. In external communication the activities will be referred to 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.

The duration of the FOT-Net Data 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 Ltd., 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, VTT.

2 Conditions for funding enhancement of datasets

The overall aim of the WP4.3 is to make a few interesting and high-priority datasets consistent with the DSF and ready for re-use. A budget of €106k was reserved for this purpose. In order to be able to differentiate the datasets from each other and make consistent choices regarding which datasets to fund, selection criteria for prioritisation and a list of items that could be funded were decided. Similarly, a list identifying what could be considered as consistent with the DSF and the topics it covers was developed.

The information collection regarding the status of different datasets started as identified data providers were approached in task WP4.1.

2.1 Selection criteria for prioritization of datasets

The selection criteria was based on knowledge gathered during previous work within the FOT-Net community relating to the FESTA methodology (FESTA, 2014) and data usage, together with the decade long experience regarding FOT/NDS and using the related data among the current partners of FOT-Net Data.

The number one question is if the data could be shared at all, based on the project agreements. The interest from the research community is also a mandatory criterion, otherwise it is no use funding the work. The data should be of high value and sufficient support should be available. Finally, the cost for making the data ready for re-use must be taken into account, also relating to the time needed. In cases where the work or legal barriers is seen to exceed project possibilities, the required steps and their estimated price could still be documented for future use.

The list of criteria is provided below together with some explanations for the interpretation of the criteria. Each criterion is evaluated on a 1–5 scale, 5 being the highest value.

Table 1: Selection criteria for prioritization

Criteria	**Dataset** (Level 1–5)	**Comments**
Possibility to share the data • OK for CA, consent forms, data owners • Special conditions for sharing data		
External demand for re-use • Potential for re-use from scientific and also maybe business perspectives • Number of projects using the data		
Valuable and of high quality • Sufficient amount of data • Rich dataset • Known/sufficient data quality • Collected according to FESTA		
Reasonable support available • Answer questions in supporting data re-use • Mandatory tools to access to the data		
Cost of facilitating the data set for re-use • Efforts needed to have the dataset available		

2.2 Consistent with the Data Sharing Framework

The DSF consists of seven high-level topics, which have been defined and further developed and detailed over the past four years within the FOT-Net projects. These are: project agreements including necessary pre-requisites for data sharing, data and metadata descriptions, data protection, training, support services, financial models and applications procedures. For each topic, recommendations have been provided that have to be considered in relation to the dataset at hand.

The objective was to make a few interesting and high-priority datasets consistent with the Data Sharing Framework, also being able to provide funds for the efforts. What do we mean by consistent? The following definition was decided by the project.

Table 2: Data Sharing Framework

Topic	Minimum level
Project agreements	Data sharing is possible and decided. Possibility to fund preparations and decision meeting with a foreseen positive outcome?
Data and metadata descriptions	Detailed descriptions of the data to be shared.
Data protection	Requirements fulfilled for their country. Fulfills the data provider site requirements how the data could be shared.
Training	Training procedure available and in use. Especially important for data privacy and IPR, must include these items.
Support and research services	Name of contact person and information on level of support available; if a highly interesting dataset, some level of support is mandatory.
Financial models	Available data re-use cost model
Application procedure	Data access procedures available and in use including application template stating mandatory information. Information of the availability of the data.

2.3 Possible items to be funded

During the course of almost a decade of FOT-Net activities, the data sharing issues have been discussed in international meetings and workshops. The questions regarding what are the main obstacles why datasets cannot be shared have always been present. Based on the documentation of the former work, the following list has been developed taking into account what should have been done in the project collecting the data, still keeping in mind that one of the pre-requisites for the selection is to enhance the number of interesting datasets. Another factor that has been a condition for becoming an item on the list is the minimum level to become consistent with the DSF, detailed in the previous chapter.

The following list of possible items to fund to enable datasets to be available for re-use was decided:

Table 3: Dataset funding possibilities

Item to fund	"Dataset"
Detailed data and metadata documentation of a dataset according to the FOT-Net's Data Sharing Framework and the Data Catalogue	
Dataset revival & proper archival	
Review of a dataset regarding the possibilities to re-use it in future projects (a smaller funding would be expected for a review only)	
Conversion of confidential data into non-confidential formats.	
Legal agreement work to enable a re-use case. Lessons learned are to be published	
Limited support services for a re-use case. Lessons learned and the case itself are to be documented and presented in FOT-Net's events	

3 Prioritization of datasets

The selection of the datasets to fund involves two main activities. The first one is identifying potential candidates for funding. This work will rely on the common work done during the development of the Data Catalogue, identifying the datasets that could be interesting to share and therefore to make visible in the Data Catalogue. Further investigations were done to find out which of those datasets that have the highest potential, but cannot be provided due to various reasons. These obstacles will be investigated and a list of potential candidates with related items needing funding will be identified. The second step is the actual selection of the datasets to fund, taking the conditions set forth in chapter 2 into account. The result of this selection will be a description of the dataset and the items that will be funded.

3.1 Interesting high-value datasets

The specific datasets to contact was identified in the beginning of 2016, based on the decided conditions for funding. This work will utilize the work during 2015 on identifying potential dataset providers to approach for adding content to the Data Catalogue. A more detailed investigation of the re-usability of the datasets was performed and potential selected candidates was approached for more information and to investigate the interest in participating. The potential work to be done and the related funding needs were investigated. A brief assessment was carried out regarding the legal, privacy protection, documentation and data processing steps that would be required to share the data. Also, the possibilities of the host organisation e.g. to offer support for analysts and keep up the dataset for the years to come was discussed.

During the first project year, FOT-Net identified several datasets that are of high value for re-use purposes (listed below). They were known for their quality and were both large and rich enough for various research purposes.

- DRIVE C2X: Finnish Test Site, French Test Site, Swedish Test Site
- euroFOT: French Test Site, Swedish Test Site
- TeleFOT: Greek LFOT, Italian LFOT, OuluFOT (Finnish LFOT), UK LFOT, UK DFOT1, UK DFOT2, UK DFOT3
- UDRIVE: Car/Truck/PTW data gathered in a common Central Data Centre; Dutch Site (Trucks and cars), French Site (Cars), German Site (Cars), Polish Site (Cars), UK Sites (Cars), Spanish site (PTWs)
- Research Data Exchange: U.S.
- Australian NDS

Identified smaller datasets interesting for re-use included:

- SeMiFOT
- INTERACTION: questionnaire data from nine countries
- ITS Platform: Denmark
- Trafisafe: Finland

3.2 Selection of datasets

3.2.1 Process

The partners of the consortium have together a good knowledge of the kind of data incorporated in the different datasets in the chapter 3.1. The selection criteria were applied and quickly just a few datasets emerged that all partners agreed would be interesting to promote, the Australian NDS, UDRIVE and then finally ITS platform as a smaller dataset with potential to become openly shared. The Trafisafe was selected to start with a pilot to test the ideas. The selection ratings and the decided items to fund are seen in tables 4-7 and Table 8.

3.2.2 Selection including ratings

The four datasets were rated according to decided criteria and comments were provided. Some parts were to be investigated, but as the knowledge of the kind of data and the amount was known, the decision to select the four datasets were unanimous.

Table 4 shows the outcome of the rating for the ANDS data set. Even if all facts were not known at the decision, the mere fact that the project collected the same data as the SHRP2 made the dataset interesting, as the same analysis tools used for SHRP2 also applies to the ANDS data.

Table 4: The Australian NDS

Criteria	ANDS	Comments
Possibility to share the data • OK for CA, consent forms, data owners • Special conditions for sharing data	3	Conditions to be investigated
External demand for re-use • Potential for re-use from scientific and also maybe business perspectives • Number of projects using the data	5	Same data as SHRP2 combined with Mobileye and eyetrackers (Seeing Machines)
Valuable and of high quality • Sufficient amount of data • Rich dataset • Known/sufficient data quality • Collected according to FESTA	4	To be investigated
Reasonable support available • Answer questions in supporting data re-use • Mandatory tools to access to the data	3	
Cost of facilitating the data set for re-use • Efforts needed to have the dataset available	4	Rich dataset

The UDRIVE data got high ratings in the selection process. Two main reasons is that the project already have a tested analysis process relying on remote access to a common

dataset utilised by 12 partners today, and secondly that the data is very rich. Table 5 gives the outcome of the rating for UDRIVE.

Table 5: UDRIVE

Criteria	UDRIVE	Comments
Possibility to share the data • OK for CA, consent forms, data owners • Special conditions for sharing data	4	Video and GPS at UDRIVE partner
External demand for re-use • Potential for re-use from scientific and also maybe business perspectives • Number of projects using the data	5	Similar data as SHRP2
Valuable and of high quality • Sufficient amount of data • Rich dataset • Known/sufficient data quality • Collected according to FESTA	4	To be investigated
Reasonable support available • Answer questions in supporting data re-use • Mandatory tools to access to the data	5	
Cost of facilitating the data set for re-use • Efforts needed to have the dataset available	4	Rich dataset

The ITS Platform was an interesting case, as it got such high ratings. There are many research areas that would benefit from that kind of dataset. The table 6 provides the complete result for ITS Platform.

Table 6: ITS Platform

Criteria	ITS Platform	Comments
Possibility to share the data • OK for CA, consent forms, data owners • Special conditions for sharing data	5	Anonymize GPS
External demand for re-use • Potential for re-use from scientific and also maybe business perspectives • Number of projects using the data	3-4	GPS based data can be used for many research questions
Valuable and of high quality • Sufficient amount of data • Rich dataset • Known/sufficient data quality • Collected according to FESTA	4	10 million km driving data from 1.3 million trips The dataset is not rich, only GPS and accelerometer
Reasonable support available • Answer questions in supporting data re-use • Mandatory tools to access to the data	5	The project leader has contributed to many FOT-Net activities and are well aware of the services to be provided.
Cost of facilitating the data set for re-use • Efforts needed to have the dataset available	3	Anonymisation is costly Storage and provision of data is moderate.

Trafisafe was a test of how much paperwork is needed to share a small dataset. The result of the evaluation of the dataset is given in table 7.

Table 7: Trafisafe

Criteria	Trafisafe	Comments
Possibility to share the data • OK for CA, consent forms, data owners • Special conditions for sharing data	5	
External demand for re-use • Potential for re-use from scientific and also maybe business perspectives • Number of projects using the data	3	
Valuable and of high quality • Sufficient amount of data • Rich dataset • Known/sufficient data quality • Collected according to FESTA	3	
Reasonable support available • Answer questions in supporting data re-use • Mandatory tools to access to the data	3	
Cost of facilitating the data set for re-use • Efforts needed to have the dataset available	2	Small dataset

3.2.3 Items to be developed or documented

During discussions with each of the selected datasets, a work plan was agreed. In all cases, the first item was somewhat changed. Instead of focusing only the data and metadata description, all aspects of the DSF should be investigated and documented. This was a way to rise the awareness of the content of the complete DSF and also get feedback on what is seen as the major advantages of the DSF and what needs additional developments.

Table 8: Items to be funded

Item to fund	ANDS	UDRIVE	ITS Platform	Trafisafe
Detailed data and metadata documentation of a dataset according to the FOT-Net's Data Sharing Framework and the Data Catalogue		X		
Dataset revival & proper archival			X	
Review of a dataset regarding the possibilities to re-use it in future projects (a smaller funding would be expected for a review only)	X			
Conversion of confidential data into non-confidential formats.			X	
Legal agreement work to enable a re-use case. Lessons learned are to be published				X
Limited support services for a re-use case. Lessons learned and the case itself are to be documented and presented in FOT-Net's events				X

4 Enhancing datasets to facilitate future re-use

This chapter documents the actual work performed in enhancing the selected datasets towards a level where they can be made available for further research. Each of the organisation(s) evaluating and enhancing the datasets have documented the work in a report. In this chapter selected sections of the reports are provided, to give an overview of what has been done. Each report is an annex to the D4.3.

The datasets are described together with the tasks that were performed. Finally, for each dataset, conclusions from the report is included. An overview is provided showing the work done and the results from the different datasets divided into the main topics of the DSF. Finally, a lessons learned chapter, is gathering the findings and thoughts for future work on providing interesting datasets.

4.1 Funded datasets and their enhancements

Four different datasets, the Australian NDS, UDRIVE, ITS Platform and the Trafisafe are presented together with the work and conclusions provided by the organisations related to the datasets. The full reports are available as annexes to the D4.3.

4.1.1 The Australian NDS

Dataset description

The aim of the Australian Naturalistic Driving Study (ANDS, www.ands.unsw.edu.au) is to understand what people do when they drive their cars in normal and safety-critical situations. The study will eventually have collected data from 360 vehicles, using the same logger as was used in SHRP2 study in the US. This opens up the possibility to re-use the SHRP2 research algorithms on the ANDS dataset. December 2016, around 70 percent of the planned vehicles had been instrumented, and the study is expected to be complete in 2018. The collected data includes video, GPS, radar, accelerometer, cell phone, Mobileye data and data from Seeing Machine systems.

Performed work and results

University of New South Wales (UNSW), Sydney, investigated the applicability of the DSF guidelines to ANDS. The work was a start for eventually making the valuable dataset available internationally. Due to the phase of the ANDS, the main comparisons between the DSF and ANDS data management were made regarding data protection and metadata.

The investigation showed good compatibility between FOT-Net Data recommendations and existing data access and the management structures of the ANDS project. A few topics were identified that need to be addressed by the ANDS consortium. They relate mostly to balancing freedom of access to data and privacy, and what to include in agreements to ensure access to data also for members outside the consortium. On these topics, the DSF documents were able to provide a useful starting point for development of extended and specific permissions for access. UNSW reported that the DSF guidelines have highlighted also several topics for future discussions in the ANDS consortium, e.g. regarding financial models and training material to be put together.

Viewed from the perspective of the ANDS project, joint projects with Europe will open the possibilities to expand the scope of our research in a number of areas and to validate others.

There is a wealth of common research questions to be investigated re-using already collected data such as the datasets from ANDS or the European NDS project UDRIVE. If planned ahead in project proposals, funding for accessing existing datasets could be included in proposals.

Conclusions

The FOT-Net's DSF has been a most valuable resource for establishing a data protection and sharing platform for the ANDS Project. The DSF provides a comprehensive and usable approach to establishing datasets that facilitate data sharing and maintain privacy and ethical standards for access. Overall, the recommendations of the DSF addressed decisions already made for the ANDS project and so were entirely consistent and compatible. On a few issues the DSF documents have highlighted areas where decisions are still to be made. In these cases, the DSF has provided guidance on a number of aspects of the ANDS database that need further development by showing the issues that the ANDS Governance Group will need to address.

The report is Annex #1 'Investigating the FOT-Net Data project's Data Sharing Framework for use in the Australian Naturalistic Driving Study'.

4.1.2 UDRIVE

Dataset description

The European naturalistic driving study UDRIVE (Eenink et. al., 2014) started in 2012, and has thus been running in parallel with both FOT-Net Data and its predecessor FOT-Net 2.

UDRIVE is the first large-scale European Naturalistic Driving Study on 120 cars, 50 trucks and 40 powered two-wheelers (PTW). The data is collected in six sites located in: France, The Netherlands, Germany, The United Kingdom, Poland, and Spain. The acquired data includes: CAN, Mobileye, video (five, seven or eight views depending on vehicle type: driver face, pedals, cockpit, steering wheel, front middle, left front, right front), GPS, and questionnaires.

Performed work and results

UDRIVE has used the checklists and recommendations of FOT-Net when taking data sharing into account in each step of the development of the project. At the same time, UDRIVE needed to detail the data protection requirements in its Data Protection Concept (Gellerman et al. 2016) covering the whole data handling chain, and the requirements for data centres and analysis sites were provided to FOT-Net Data who generalized the requirements and incorporated them into the DSF.

UDRIVE included in the report how the different DSF topics were implemented in the project. In short, the UDRIVE description of work incorporated a text where data would be provided to third parties, if financial means were provided. The consortium agreement followed the DSF checklist and ownership, storage and access after the project was addressed. The participant agreement included written consent to data sharing within specified areas. The same template was used for all countries. The external agreements were given special attention, following the DSF.

The data description recommendations are followed, but not yet fully documented, and a Data Protection Concept was developed covering the data handling from collection to data re-use after the project. All partners need to document how they handle the data.

The data can be remotely accessed after the project by third parties. To protect the personal data, video and GPS, these data can only be accessed via a secure enclave at one of the project partners having remote access to the CDC.

The analysis in UDRIVE is performed by eleven partners remotely connected from thirteen analysis sites to one Central Data Centre (CDC), hosting all data collected in six countries including video, GPS and confidential data. The same data sharing set-up is planned to be used after the project and training and support services are already in place. Application forms and procedures will follow the DSF. The project has not been provided funding for maintaining the dataset after the project and discussions are on-going to seek for possibilities. Data sharing is depending on that a solution can be found.

Conclusions

The Data Sharing Framework was developed in parallel with the UDRIVE project. For this reason, most of the suggested guidelines were applied in the project and some others were updated with the lessons learned from the project. The recommendations in the DSF definitely provided a reference point when shaping all of the data-related issues in UDRIVE. The DSF does not necessarily imply a single and strict structure on how to ensure proper data sharing for every possible project but it can be considered as a starting point and could evolve into different adaptations depending on specific needs. Based on lessons learnt in UDRIVE, we have also proposed improvements to the DSF.

Several challenges have emerged within UDRIVE when trying to ensure data sharing. A very time-consuming task has been getting the acceptance and clear requirements to share data from a specific country according to the specific legislation.

Another great challenge which continues to be time-consuming and that will have an important financial impact is ensuring the post-UDRIVE data & tool maintenance. It has been difficult to find a business model involving a monetary risk partners are willing to take based on the overall uncertainty of the availability of project that can eventually fund the costs. Finding a sustainable solution is crucial to ensure the longevity of the data.

Overall, the UDRIVE project can recommend the DSF as providing excellent guidance in what pre-requisites and procedures to implement in the project.

The report is Annex #2 ‘Application of the FOT-Net Data Sharing Frameworkon the UDRIVE dataset’.

4.1.3 ITS Platform

Dataset description

Data were collected in 2012–2014 in the Danish research and innovation project ‘ITS Platform North Denmark’, also known as the ITS Platform. The ITS Platform data consists of GPS-based floating car data (FCD) from 425 privately owned cars for about two years. Each On-board unit (OBU) collected FCD with 1 Hz ID, position, map-matched position, direction, speed and a number of other attributes, which are mostly related to position reliability. In addition, acceleration data was collected with 10 Hz in three dimensions. Data consists of about $1.4*10^9$ positions. The recorded distance driven is about 15 million km in total. The number of accelerations recorded approximates $42*10^9$.

Partners were Aalborg University, Gatehouse and the at the time start-up company Inntrasys.

Performed work and results

Aalborg University (AAU) applied the DSF to a finalized Danish FOT: ITS Platform. The work included anonymisation of the dataset, compiling metadata and creating a training manual for re-users. The AAU provided detailed feedback on the DSF based on the case.

The DSF was used as a guideline for developing a more complete metadata description and a training manual. Though the development of the anonymization algorithm was the main effort, considerable time was spent beforehand on discussing the financial model and the related data protection issues. The decision was to anonymize the data and provide it openly without cost. The interesting idea of getting indirect funding through citations was identified.

In order to make the dataset publicly available and to avoid any privacy issues, data anonymisation procedures were developed and carried out: Time and data and car identifiers were removed. In addition, all data close to (using varying distances between 200 and 500 meters) start and ending points of each individual trip were removed.

The first parts of this dataset are already available and the remaining part is planned to be available for the public in the beginning of 2017. This public dataset is expected to consist of more than 10 million km driving data from 1.3 million trips. The dataset is made available at http://fcd-share.civil.aau.dk/.

Conclusions

AAU has opened up its FCD for the public and ensured that the data is sufficiently anonymised. The main source of inspiration and guidance, especially with regard to data protection is from the Data Sharing Framework developed by FOT-Net Data. Also, the nature of the data, the local conditions and resources, but in particularly the valuable contribution from the FOT-Net Data society with recommendations and especially many good scientific discussions, have shaped the quality of the publicly available FCD.

The report is Annex #3 ‘Aalborg application of the FOT-Net Data Sharing Framework on the ITS platform dataset’.

4.1.4 Trafisafe

Dataset description

The dataset Trafisafe is from a Finnish research project Trafisafe (2012–2014), which focused on driving style feedback for young drivers and their parents. The dataset consists of GPS, tri-axial acceleration data, OBD-based fuel consumption and engine RPM values, and questionnaire responses. Trafisafe partners were the Finnish Transport Safety Agency, EC-Tools and VTT.

Performed work and results

VTT and the University of Leeds carried out a test case for sharing a dataset. The main purpose of the test was to clarify required agreement details for accessing a dataset with the help of a real re-use case and two lawyers. The case also served as an example for assessing the level of additional efforts required in documenting a dataset, when re-users were not familiar with the study that collected the data.

Agreement negotiations raised up interesting discussions in three main areas:

(1) Liability clauses, i.e. what could be the level of damages (financial and reputation) if parts of the dataset would leak out as the dataset contained full GPS data.

(2) Common confidential data protection clauses were included in the contract: e.g. to ensure that those having access to the confidential information should receive advice, and for reasonable measures to be taken in the handling of the dataset. Further, according to good scientific practises, evaluation reports should not show full GPS tracks of single persons.

(3) Mainly the existing data description documents had to be translated into English and further comments to be added to avoid misunderstandings.

The complete DSF was compared with the Trafisafe implementation and used to enhance for instance training material.

Conclusions

The efforts to share the dataset were mostly related to discussions and numerous e-mails from both sides, involving two lawyers. In comparison, translation of key documentation took only a day to complete and required support for re-users maybe two days.

University of Leeds found the dataset to be easy to use and has planned further use for it in student work and research projects.

The report is Annex #4 ‘Test case Trafisafe’.

4.2 Overview of applying DSF on the selected datasets

All organisations receiving funding investigated how the DSF applied to their dataset. Table 9 provides an overview of the outcome. The use of the DSF is though different as two datasets are a few years old and two are still collecting data.

Table 9: Overview of applying DSF to selected datasets

DSF topic	ANDS	UDRIVE	ITS Platform	Trafisafe
Agreements	Data sharing prerequisites in all agreements	All four types of agreements include the data sharing pre-requisites.	Pay especially attention to the participant agreement, as it difficult to go back for second OK.	Re-use possibilities inserted in original agreements by Trafisafe partners
Data description	Descriptive and Study Design and execution is done. The other to will be done when storing our own data.	Similar structure as proposed by the DSF is implemented in the project and is supported by the tools.	All five descriptions are made for the ITS Platform and downloadable as a scientific report.	Test design description and detailed data description. Translated to English
Data protection	In place for most components.	The Data Protection Concept was developed covering the complete data handling in the project.	DSF gave good guidance in showing possibilities. Final decision, anonymising data. Original data protected.	Encrypted transfer, NDA, reasonably anonymous; normal precautions.
Training	Under development using DSF guidance.	Training is mandatory in the DPC. All partners develops their own.	Training descriptions was updated as part of FOT-Net work.	Powerpoint pres: install DB, insert data, some scripts were provided.
Services	Not applicable (yet)	Data sharing is done in the project with related support services.	Keep it simple as key person usually busy with other tasks.	Support services: information and tools, info on earlier research.
Financial models	Models are under discussion using DSF.	No sustainable model found yet. If no funding the data will not be available after 1 July 2017.	Setup with anonymized data, self-explanatory services and no charges. Payback via citations	Small dataset, no costs. Costs for lawyers when sharing data.
Applications	In place for partners. Other users, under discussion.	Procedures developed within DPC. Post-project still to be developed	Keep it simple: Enter webpage, fill in form, read and accept, download data.	No formal application procedure, handled case by case.

4.3 Lessons learned from applying the DSF to datasets

Providing funding for hands-on application of the DSF on an organisations own datasets, raises the awareness of the procedures involved in data sharing. Maybe could be used to spread the DSF.

Mandatory documentation of data protection implementation as in UDRIVE revealed many misinterpretations, which could become costly for the project.

An awareness of the personnel efforts needed for data sharing, which led to not charging costs, but forced to find alternative payment methods as citations. It also led to self-explanatory set-ups.

Data sharing pre-requisites were part of the project agreements in all four cases, a pleasant surprise.

Data protection guidance was appreciated, also in the U.S. DOT report.

The amount of time involved in anonymizing a dataset needs to be taken into account. Same experience were made in the U.S., when the Safety Pilot data was to be anonymized before upload to RDE. Automated tools are needed.

Financial support for projects collecting data is important, as the expensive dataset might not be further re-used otherwise. The cost of maintenance funding is far less than collecting another set of data.

Collaborative research projects reaching out cross the border of EU, should be investigated. Many driver behaviour could be investigate simultaneously re-using datasets collected in a similar manner, such as the SHRP2, UDRIVE and ANDS datasets.

Finally, it has taken some effort to make people use their time to provide comments on the DSF. Those that did, have all conveyed how they appreciated the framework and that it really is something that could be used as a starting point in various discussions. Or as one participant in the Leeds workshop pleasantly surprised said, “it contains important information on all subjects we are discussing right now.”

4.4 Data Sharing Framework vs. U.S. frameworks

FOT-Net have a long history of cooperation with the U.S based on a common view that we can learn from each other´s work and thereby enhance the knowledge in the FOT/NDS domain. It was therefor really interesting to read the report provided by the ITS Programs Connected Data Systems Program, where U.S. frameworks had been compared to the two main FOT-Net frameworks, FESTA and the DSF. The full report is included in as an annex, and some highlights are provided in this chapter. SHRP2, as the major data provider for re-use as of 2017, is presented as well, ending with a short reflection on the connection between the DSF topics and what has been implemented for the SHRP2 dataset.

4.4.1 RDE and JPO program

The applicability of the DSF on American conditions was examined in a report provided on behalf of the ITS Programs Connected Data Systems Program, U.S. DOT. It also contains reflections of the FESTA methodology. The following, and many more, thoughtful comments can be found in the report.

FOT-Net documents can be compared with similar resources and forms of support in the United States to conduct research involving FOTs and early connected vehicle (CV) deployments as well as NDS. Guidance for the connected vehicle pilot deployments consists of numerous documents similar to FESTA and there are many connections between documents provided for the users of the Research Data Exchange (RDE), a large data repository, and the content provided in the DSF.

The RDE also reveals implications for standards, intellectual property rights (IPR) data ownership, and privacy. The RDE datasets do not include any private or sensitive data, and the data is in the public domain because the U.S. DOT owns the distribution rights and data providers have signed agreements. Large efforts are though being made to develop methods for and anonymizing the data focusing on the GPS position, especially parts of the Safety Pilot dataset.

The posture of the DSF on the topic of financial models is not unlike the posture of the U.S. DOT in regards to insisting that the Connected Vehicle Pilot Deployments be financially sustainable.

The U.S. DOT connected vehicle deployment program and related activities can directly or indirectly use the reports, catalogues, data, tools and lessons learned from FOT-Net. Of considerable interest is protecting PII and IPR and making good use of material that could enhance the RDE as it evolves, for example by providing improved guidance. We share a common concern regarding PII and intellectual property rights.

One of the main strength in the DSF is the procedures for protecting data at Data Centres and Analysis Sites in the data protection chapter, according to the report; “As mentioned earlier these requirements may be offered for consideration as national regulations”. This shows one example of the potential benefits that could come out from a further common review of the documents.

The report ends with the following conclusion and recommendation.

FOT-Net and sister programs of the ITS JPO have fundamentally similar goals. Their approaches to data sharing, data protection, experimental design and implementation of vehicle and infrastructure applications overlap to a great extent but differ in many ways. The practices identified in this review make clear that FOT-Net (as it is embraced by CARTRE) and the ITS JPO can benefit from continuing to learn about each other’s policies, practices, and procedures regarding data sharing, protection and related topics. It is recommended that this comparison of practices periodically be updated and shared with both parties and the broader international community.

The report is Annex #5 ‘FOT-Net and the RDE – learning from each other’.

4.4.2 SHRP2

Since SHRP2 was finalized in the spring of 2015 and the data made available, many different stakeholders have begun to mine the data to explore questions in a variety of research areas. The SHRP2 database contains NDS data from over 3,500 drivers recruited from six locations in the United States, in total more than 5 million trips. Data include video, sensor, vehicle network, and participant assessment data, as well as summary data related to events and trips. Roadway elements can be obtained from the Roadway Information Database (RID).

The main SHRP2 usage covers understanding and further developing safety performance measures and developing data analysis tools and methods, whereas the full scope includes driver distraction, infrastructure analysis, driver age-related issues, driver fatigue and impairment, roadway lighting, fuel economy, and pedestrian/vehicle conflicts. The breadth of use underscores the importance of making naturalistic data sets widely known so that researchers can figure out how datasets collected globally can be used in their respective research areas.

The users of the SHRP2 data are from different parts of the world, the majority being from the United States. The data can be accessed either via a website or through research-specific requests for data. Ten percent of users are original equipment manufacturers (OEMs), 10% are private firms excluding OEMs, and 10% are public health organizations, federal laboratories, and overseas universities, as of spring 2016. The requestors have levels of expertise ranging from undergraduate students to noted researchers.

Data access is based on the level of detail requested and the need for personally identifying information (PII) either through the InSight website (https://insight.shrp2nds.us) or via a data use license (DUL). Video and GPS can only be accessed within a secure data enclave. There were 174 active DULs for SHRP2 data, and between 20 and 30 requests per month as of two years after the dataset was opened up for re-use.

The SHRP2 NDS data and analyses are already providing new insights into driver behaviour, both during safety-critical events such as crashes and during normal driving. The variety of researchers reusing the SHRP2 data points out the potential value still to be explored in naturalistic datasets worldwide.

The SHRP2 project had a requirement from the beginning that the data should be shared after the project. SHRP2 has therefore incorporated data sharing pre-requisites into the participant agreements and documented the data. The data protection procedures are developed to protect the collected personal data, such as video and GPS, and can only be accessed on the premises of Virginia Tech. Other datasets are made available via a web interface. The SHRP2 organisation was awarded 25 million dollars in 2014 to continue to keep the data available and fund research during four years. This funding is partly used to keep the data maintained and accessible, and provide support services for researchers wanting to re-use the data. Full set of application procedures and application forms have been developed.

SHRP2 is a success story showing that with substantial funding, the data can be made available for the research community in a professional manner.

5 Conclusions

This document has described the work done on applying the DSF to selected, high-value datasets. Funding has been provided to evaluate and enhance four datasets, to facilitate the provision of data. Through applying anonymization technics to GPS, the ITS platform data has been put online with an automated data download procedure.

The projects UDRIVE and ANDS shares many of the issues in sharing the dataset, as they contain video as personal data. Both projects have implemented most of the recommendations provided by the DSF on the seven topics that should be considered if sharing data. Still, these projects are still looking for solutions to be able to maintain and provide this data. Just relying on separate projects funding data access is not enough, as the uncertainty of future cost coverage inserts to much risk to keep it available.

The process of providing minor funding for investigating the applicability of the DSF on different datasets, has been proven a successful way of initiating work to provide more datasets. A proposal for the future would be to reach out to new projects and support that they utilise the Data Sharing Framework to provide guidance in what need to be addressed to provide data.

.

List of abbreviations

ANDS	Australian Naturalistic Driving Study
CA	Consortium agreement
CAN	Controller area network
CDC	Central data centre
CV	Connected vehicle
DSF	Data Sharing Framework
DUL	Data use license
FCD	Floating car data
FOT	Field operational test
GPS	Global position system
IPR	Intellectual property rights
NDS	Naturalistic driving study
OBD	On-board diagnostics
OBU	On-board unit
OEM	Original equipment manufacturer
PII	Person-identifiable information
PTW	Powered two-wheelers
RDE	Research Data Exchange
RID	Roadway information database
RPM	Rotation per minute

List of tables

List of annexes

Annex 1: Investigating the FOT-Net Data project's Data Sharing Framework for use in the Australian Naturalistic Driving Study (ANDS)

Annex 2: Application of the FOT-Net's Data Sharing Framework on the UDRIVE dataset

Annex 3: Aalborg application of the FOT-Net's Data Sharing Framework on the ITS Platform dataset

Annex 4: Test Case – Trafisafe

Annex 5: FOT-Net and the RDE – Learning From Each Other

References

Eenink, R., Barnard, Y., Baumann, M., Augros, X., Utesch, F., 2014. UDRIVE: the European naturalistic driving study. Proceedings of the Transport Research Arena, Paris.

FESTA, 2014. FESTA handbook v5. Available at: http://fot-net.eu/Documents/festa-handbook-version-5-2014/

Gellerman, H., Svanberg, E., Kotiranta, R., 2016. UDRIVE Data Protection Concept. In: Proceedings of the ITS World Congress 2016, Melbourne, Australia.

Gellerman, H., Svanberg, E., Kotiranta, R., Heinig, I., Val, C., Koskinen, S.,Innamaa, S., Zlocki, A., Bakker, J., 2017: Report from the FOT-Net Data Sharing working group, In preparation.

ANNEX 1

Investigating the FOT-Net Data project's Data Sharing Framework for use in the Australian Naturalistic Driving Study (ANDS)

Prof Ann Williamson

Transport and Road Safety (TARS) Research Centre,

University of New South Wales, Sydney, Australia

Background to Project

The FOT-Net Data project is established under the EU 7th Framework Programme for Research. The objectives of this project include the efficient sharing and reuse of data generated by Field Operational Tests (FOT) and Naturalistic Driving studies (NDS). Aligned with this objective is the development and promotion of a framework for sharing data. Over the past three years, the FOT-Net Data project has produced a Data Sharing Framework (DSF) including Data protection recommendations and a framework for Data and Metadata description. These documents have been disseminated widely both within the FOT-Net group and beyond.

The aim of this report is to investigate the applicability of the DSF procedures for the Australian Naturalistic Driving Study (ANDS), which is currently being conducted. Specifically the required tasks are to undertake the following activities:

- Comment on the DSF
- Investigate DSF procedures for ANDS
- Document data and meta data according to DSF at a high level
- Investigate data management incorporating data protection
- Investigate long-term financial models for data access
- Investigate joint data analysis possibilities with Europe
- Write a report on the findings

The current stage of development of the ANDS project provides an opportunity to trial and evaluate the DSF. This report describes the results of the investigation and each of the aspects of the DSF. First, the report describes the background to the ANDS including the rationale, characteristics and current status of ANDS. The report then describes the analysis of the Data Sharing Framework and its applicability to ANDS.

Background and characteristics of the Australian Naturalistic Driving Study

The aim of the ANDS is to understand what people actually do when they drive their cars in normal and safety-critical situations. There have been a number of Naturalistic Driving studies around the world, especially in the USA. How well their results apply to Australian conditions, however is unknown. Factors like driving conditions and environment, vehicle fleet composition, and driving regulations can be different in other parts of the world. Country and regional differences such as population density and structure and culture and socioeconomic factors may play a role in road safety outcomes. These differences mean

that the applicability of the results of previous NDSs to the Australian context is unknown. Furthermore, few of the previous NDS have explored many of the high priority road safety problems identified in the *Australian National Road Safety Strategy (2011–2020)* such as: speed choice, vulnerable road user interactions in different situations and urban versus regional areas and none have addressed the Australian context.

The ANDS also presents new opportunities for new insights into road safety that have not been possible using previous methods. These include:

- Normative data: Analysis of fundamental data on such aspects as how people drive, how they *avoid* crashes, navigate, maintain speed, adhere to traffic laws; stay within their lane; control the vehicle, etc.
- Human Factors Data: Analysis of perception reaction times and hazard recognition times for different drivers and ages and compare this data with current published literature where closed circuit tracks and simulators have been used.
- Exposure: Provides new/more detailed data on when we drive (driver, vehicle, road, traffic and environmental factors) and so are more or less likely to be on the road: for design and policy.
- Validation: Capitalise on the opportunity to validate findings from existing surveys, observational studies, simulator studies and data collected on Police-reported crashes.

The ANDS project involves a Consortium of road safety research centres from four universities (the Transport and Road Safety (TARS) Research Centre from the University of New South Wales, Monash University Accident Research Centre, the Centre for Automotive Safety Research at the University of Adelaide and the Centre for Accident and Road Safety – Queensland from Queensland University of Technology), five road safety authorities from NSW (Transport for NSW), Victoria (VicRoads and Transport Accident Commission), South Australia (Motor Accidents Commission) and Western Australia (Office of Road Safety), one road user association (NRMA) and two commercial partners (Seeing Machines and Hyundai). Approximately half of the funding for the project was supplied by the industry partners in the Consortium and half by the Australian Research Council in the form of a Linkage grant. The initial ANDS project is planned to take three years (2015-2017 to collect data and develop databases ready for further analysis by the Consortium users.

As for many other similar projects, the ANDS project involves three main governance groups, as follows:

1. Core Management Group which is responsible for the management of the project including progress, direction, results IP disclosure and protocols and implementing the study design, data analysis and dissemination of findings.
2. Stakeholder Advisory Group which is responsible for ensuring that all stakeholder requirements are met and liaison with the Core Management Group on all matters.

3. ANDS Governance Group which is responsible for deciding on further use of the DAS units and further use of the ANDS data in future studies.

Methodology

The ANDS was designed to recruit 360 drivers and vehicles into the study. Each vehicle is owned by the participating driver and is instrumented for four months. Drivers recruited must have a full drivers licence and be in the 21 to 70 years age range. Around half of the drivers/vehicles are recruited from the Sydney region in NSW and half from the Melbourne region in Victoria. The final study sample will include drivers residing in urban and rural areas in the proportion 70:30. The data collection is occurring in waves of around 45 vehicles in each location.

Three different types of data acquisition systems are being used to instrument vehicles (see Table 1). All vehicles are being instrumented using the NextGen Data Acquisition System (DAS) which were developed by Virginia Tech Transport Institute (VTTI) and originally used for the SHRP2 study. Most vehicles are also fitted with the Mobileye system and up to 26 vehicles will be fitted with Data Management Systems from Seeing Machines. Table 1 also provides an overview of the information collected by each type of instrumentation.

Progress so far (at 1 Dec, 2016)

Currently, around 70 percent of the planned 360 vehicles have been instrumented. The composition of the participating drivers is consistent with the original plan for the sample of drivers, including an equal representation of males and females, even distribution of ages across the planned range 21 to 70 years and approaching one-third residing in rural areas.

ANDS Data Analysis Plan

The project plan for ANDS laid out seven key research themes for analysis selected through consultation with all researchers and partner organisations. These included:

- Safety at intersections
- Speed choice
- Interactions with vulnerable road users
- Fatigue
- Distraction and inattention
- Crashes and near-crashes
- Interactions with intelligent transport systems (ITS)

Table 1: In-vehicle instrumentation in ANDS and summary of the information collected.

Data Acquisition System	Information collected
VTTI NextGen Data Acquisition System (110 units)	• Video (4 cameras) • Camera (snapshots) • Machine Vision: – Eyes Forward Monitor – Lane Tracker – Driver ID (post hoc) – Head pose • Accelerometers (3 axis) • Gyro Rate Sensors (3 axis) • GPS – Latitude, Longitude, Elevation, Time, Velocity • Forward Radar: – X and Y positions – X and Y Velocities • Cell Phone – GSM, ACN, health checks, location notification – Health checks, remote upgrades – Illuminance sensor – Infrared illumination – Incident push button (audio: 30 secs) – Passive alcohol sensor – Turn signals, Brake signals – Vehicle network (CAN) data – if available – Accelerator, brake pedal activation, ABS, Gear position, steering wheel angle, speed, horn, seat belt information, airbag deployment etc. …
Mobileye ADAS (90 units)	• Headway monitoring and warning (time headway threshold adjustable)** • Lane departure warning** – when wheels cross right or left line markings • Forward collision warning** – with cars, buses, trucks, motorcycles; moving or stationary • Pedestrian and cyclist collision warning** – in danger zone; imminent collision • speed; indicator use; brake use; indicator use; high beam on/off ** All warnings turned off
Seeing Machines – DMS (n =26)	• real-time analysis of head pose, gaze and pupil metrics and eyelid opening and classification of direction of gaze.

ANDS Data management and sharing framework

The aim of the Data management and Data sharing framework being developed for ANDS is to design a data management system to facilitate sharing of ANDS database with all project partners and ultimately other road safety groups. This aim is consistent with that of the FOT-Net DSF and, as the ANDS Dataset will be large, the DSF should be particularly suited to assisting in achieving the aim of establishing a database that can be readily shared with relevant and suitably qualified users.

Currently the ANDS Database management involves the structure shown in Figure 1. Data collected from the instrumented vehicles using the VTTI DAS units is uploaded to a Data repository held at VTTI. Prior to entering into the database, the uploaded data is cleaned and any incorrect or duplicate variables are removed and drivers who have consented to participate are identified and included in the database and data involving drivers who had not consented is removed. Data can be accessed by Analysis Sites which are ANDS Partners in the Project Consortium. Data Access Practices have been developed to ensure the data is managed ethically and takes into account privacy considerations. Access to data held in the VTTI Repository currently requires a two steps verification: an IP address and an Application Programming Interface (API) Key.

Currently, access to the VTTI database involves two components. OzSight is a web interface that allows access to non-sensitive data and is intended for general enquiries and the Endpoint database contains all data and involves an API interface for three languages: Matlab, Python or R. It is planned in the near future to create a copy of the ANDS dataset at UNSW in Sydney and to create a Data Centre at UNSW which can be accessed by the Analysis Sites.

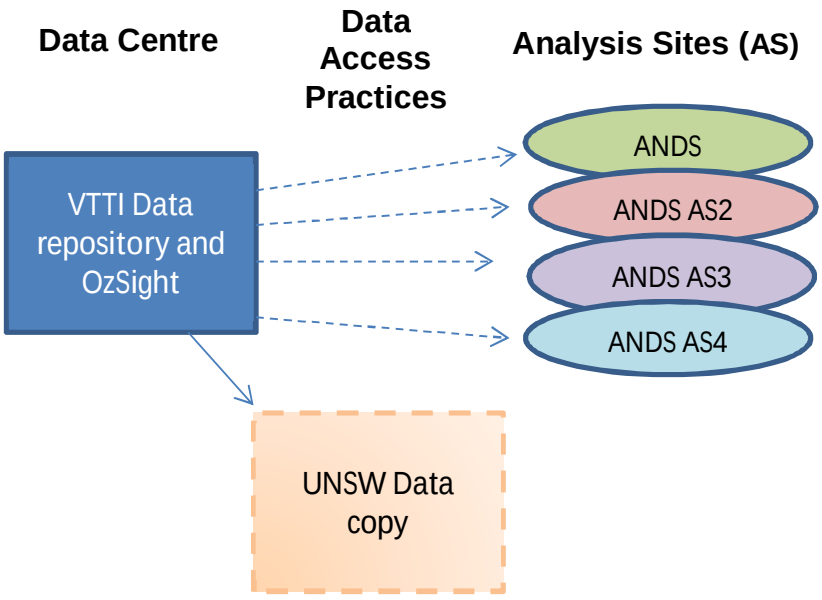

Figure 1: Current ANDS Data Management Process

A set of Data Dictionaries have been developed for all data contained in the VTTI databases. These include dictionaries for Time Series data, Trip Summary data, Critical Event data, Vehicle detail data and Driver characteristics data including questionnaire and driver personal data.

Data from the Mobileye devices is stored in a database being generated only at UNSW but will be available to Analysis Sites separately to the VTTI data as part of the Data Centre dataset. Data from the limited number of Seeing Machines devices being used in this project is commercial-in-confidence. It will not be part of the main data collection but the data will be used for certain research questions to aid in validating data collected using the VTTI system especially the video recordings of drivers.

Comparison of data management and data sharing for ANDS with FOT-Net Data recommendations

The major components of the DSF cover seven main areas which need to be addressed. Table 2 summarises a high-level analysis of each of the seven components that compares the recommendations by FOT-Net Data and the current ANDS project characteristics. This shows that the ANDS project has addressed and complied with the first three recommendations of FOT-Net Data, specifically the need for incorporating data sharing pre-requisites in the Agreements, Data and metadata descriptions and Data protection. The analysis is elaborated component by component in the sub-chapters below.

Table 2: Comparison of the major components of the FOT-Net Data Sharing Framework and characteristics of the ANDS Project

FOT-Net DATA		ANDS Project status
Component	Recommendation	
Agreements	Data sharing pre-requisites to be included in four types of project agreements	Done with all existing partners
Data and Metadata descriptions	Availability of data and metadata descriptions for all types of metadata.	Currently Study design and Execution metadata is available and Descriptive data as supplied by VTTI, Mobileye, Seeing Machines (see Table 3 below).
Data protection	Data protection requirements for Data Centre and Analysis Sites	In place for most components (see below)
Security and training of users	Security and personal integrity training for all	Under development using FOT-Net DATA guides
Support and research services	Support new users of the dataset	Not applicable (yet)

Financial models	Financial model for data management costs to allow for continued access to the database	Models are under discussion. See section above.
Application procedures for use	Application procedures and template for new data users	In place for use of data by Consortium members only. Procedures for further users are currently under discussion

Agreements

Agreements, all established before the commencement of the project, comply with the *National Statement on Ethical Conduct in Human Research* put forward by the Australian Research Council and National Health and Medical Research Council, the Human Ethics Committees of each University partner and the requirements of industry partners. This includes participant consent forms, funding agreements with all parties and the agreement between the consortium members and agreements with external providers of data including Mobileye and Seeing Machines. The overall objective of all of these agreements is to enable sharing of data within ethical guidelines on usage of data and privacy.

The participant agreements include requesting permission from participants to contact them again after the study to request further information or to invite them to participate in further research. This has already proved to be beneficial as we sought, and were granted, ethical approval to contact participants again after the naturalistic driving data was collected to undertake further cognitive testing in order to look at the relationship between participant impulse control and driving behaviour. This additional project is currently underway.

It should be noted that the two commercial partners, Mobileye and Seeing Machines have different agreements as the focus of their involvement is different. Mobileye is providing specific data to the database that is collected in every participating vehicle. Seeing Machines is participating in the Consortium with a more research-based focus in which data is being collected from a small percentage of vehicles and the ANDS Team will work jointly on variables that are generated by this technology including the development of new variables.

Data and metadata descriptions and data security

The next two components of the Framework: Data protection and Data and Metadata requirements are quite complex and justify a more detailed consideration. Four different types of metadata are defined in the DSF and documents: Study design and execution documentation and descriptive, structural and administrative metadata. ANDS has developed and documented the first two types and is currently working on developing Structural and Administrative metadata for the Australian Data Centre and the ANDS Team is working through obtaining a full copy of the dataset with VTTI and integrating this data with the data from Mobileye. These two components are discussed in more detail in a separate section below.

Security and training of users

ANDS is also currently developing Security and Training requirements and support tools for all users, based on the FOT-Net Data recommendations and on the range of tools available for preparing, using and analysing FOT and NDS data. Security of data and ethical use of the database are primary considerations in the development of the Data Centre and the use of data through Analysis Sites.

Financial models

The DSF identifies the need for sustaining funding to maintain databases such as ANDS. The ANDS Consortium has established a formal committee to consider use of the ANDS Database once the initial project is complete; however financial models for funding sustained use of the Database are currently under discussion. All of the financial models were discussed at a recent meeting of the Core Management group as potentially viable options for sustaining funding for ANDS. It is not clear at this stage which model might be most successful for ANDS. The original proposal for the ANDS project was not able to include funding for maintaining an on-going database. In the short term, the funding will be covered as part of the UNSW and TARS Research Centre core activities although a longer term solution needs to be found. The ANDS Team is currently seeking further supporting funding for this purpose from funding partners, but the solution will almost certainly also include a User-pays arrangement for data analysis services on a commercial basis which is Option G in FOT-Net DSF. The ANDS Consortium strongly agrees with the FOT-Net Data recommendation that it is essential to look for funding for the future, and it is the intention to have a funding model in place by the time the initial project is complete. The development of support services and materials for external users will also be developed as part of the model for sustained use and funding of the ANDS database. The issues of funding dataset maintenance and sustainability in the long term are discussed later in this report.

Application procedures for use

Currently, Application procedures for data use by ANDS Consortium members are in-place including access to the VTTI OzSight database, although they are continuing to be revised using the DSF recommendations. Procedures for users outside the Consortium have yet to be developed but again, the DSF recommendations and guidance material will be very helpful in ensuring that all potential issues are addressed.

Investigation of data management and data protection recommendations

Since Data Protection and Data and Metadata descriptions were identified as significantly large issues in developing data sharing practices, a more detailed comparison of the FOT-Net Data recommendations and ANDS Practices is shown in Table 3 and the results are described in the next sections.

The FOT-Net Data Protection Recommendations provide a sufficiently broad and deep framework such that the structure of the ANDS database overall fits well into the DSF. The recommendations relating to the protection of access, accidental deletion of data, documentation and confidentiality are all very similar between the FOT-Net DSF and the ANDS practices for Data Centres and Analysis Sites. There are two exceptions. The FOT-Net DSF sets up four different types of data that may be included in a database, whereas the ANDS is only using three types as we collapse Personal and Sensitive personal data types. In ANDS, all information relating to an identified person should be treated with the same caution. We are currently not distinguishing personal data as more or less sensitive. The ANDS Team are aware that EU legislation (Article 10) specifically '*prohibits the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life'.* Such a distinction is not made in Australian legislation on ethical practice and privacy which restricts analysis and reporting of all data that specifically identifies individuals, requiring procedures for de-identification.

Secondly, ANDS does not have different permission for access and use of data for data downloads and data extractions. The DSF Data protection recommendations distinguish the process of Data downloads where part or all of the project data might be downloaded by an analyst or group for a purpose but the data can be reused for further analysis, from the process of data extraction where the data is not for reuse but only used for extraction of analysis results, plots and graphs for papers and presentations, which are all anonymised. No data for analysis is extracted. Our current plan is to continue to require the same Data access and use permissions for all users whether for data download or extraction of the results of analysis and for all research questions. In the future, if this requirement is seen as too onerous for analysis of ANDS data by users outside the consortium and the analysis only involves non-sensitive data, the permission requirements may be reduced.

In addition, as discussed above, we have not yet really evaluated the applicability of the DSF Recommendations on post-project use of the data although since ANDS data is being shared from a central database, data sharing procedures are very likely to be the same during and after the project. This is the task of the ANDS Governance Group which will be complete before the end of the ANDS project.

The FOT-Net Framework for Data and Metadata description is also compatible with the data dictionaries and metadata descriptions already established for the ANDS Database, and so it will provide a formal framework for the development of further data and metadata descriptions as new variables are developed. This will be included in the training resources we will develop for ANDS Database users. Members from the ANDS Consortium are members of the ISO TC 22/SC 39/WG "Naturalistic Driving Studies - Defining and Annotating – Safety Critical Events" currently being chaired by VTTI which has been established to set common definitions of Safety Critical Events.

The ANDS data corresponds well to the categories of FOT/NDS data described in the draft framework. ANDS collects context data, acquired data from vehicle sensors and video in the form of time-history data especially from in-vehicle measures and from this ANDS holds derived data from transforming the acquired data into more directly usable form. All acquired or derived data has documentation of its metadata attributes in a very similar way to that suggested in the framework.

Table 3: FOT-Net Data – Data protection recommendations

FOT-Net Recommendation	ANDS Actions
1. Data classified into four types: Personal Sensitive Personal Confidential Non-sensitive data	Uses the same classification typology but collapses Personal and Sensitive Personal groups and uses very similar definitions, as follows with exceptions in italics: i. Personal Data means any information relating to an identified or identifiable person, e.g. a participant; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number (ID number) or to one or more factors specific to their physical, physiological, mental, economic, cultural, social identity *or their travel patterns.* *ii.* Confidential Commercial Data refers to data that might need protection, e.g. Mobile Eye or Seeing Machines data. *There are three levels of categorising such data: 'Open' where all approved analysts and ANDS project consortium partners can access the data freely; 'Closed' where confidential commercial data is open to all ANDS consortium partners during the project on a per project approval provided by the Commercial stakeholder and the ANDS consortium that includes the Core Research Group; 'Confidential' data owned by the stakeholder that is never shared, as the commercial value of the data is too high for data sharing* *iii.* Non-sensitive data refers to completely anonymised data and does not include any confidential elements. This means that no personal identifiable data is available in the dataset (i.e. video, images or GPS traces) *that can identify the participant*. If video is included in the dataset, any traffic participant must be anonymised (e.g. blurred) to ensure the personal integrity, but also other objects that can be used to identify a person (e.g. number plates). If GPS-traces are included in the data it is important methods are used to protect the participant to be identified (e.g. by removing the first and last parts of a trip that can identify the participants address). The data cannot include any confidential data unless an agreement with the participant or other individuals identifiable in data allow public usage.

FOT-Net Recommendation	ANDS Actions
2. Roles: two types of roles are defined: – Data Centre – Analysis Site	These two distinctions will be used
3. Data access methods: – Public download – Conditional download – Remote access – On-site access	Currently ANDS data is only available to restricted users through a Conditional download arrangement with VTTI. In the future, a data repository will be established at UNSW which will involve Conditional downloads by accredited study participants and On-site access.
4. Data protection: a) Data Centres	
- Protection from unauthorised access (DS-1)	ANDS data is protected from unauthorised access in two steps: physically via an IP address as all servers and other equipment is kept secure with access only by authorised people through an authorised computer, and logically through limitations on which groups and individuals have full or limited access to data using an API key. Data uploading is through specific workstations and transfer of data between firewalls is encrypted. Mobile devices such as Disks and USB sticks can be used to transport data but personal or confidential data being transported must be encrypted.
- Protection from accidental deletion or corruption (DS-2)	ANDS will maintain an archive of data which will only be used for that purpose. A range of measures will be taken to ensure that in the analysis, we maintain multiple redundant hard drives and similar devices to ensure back-ups if data is lost for whatever reason. Regular back-ups will be made.
- Document data protection implementation (DS-3)	The data protection implementation will be documented and maintained by the Consortium and will be regularly reviewed. The FOT-Net Data implementation guidelines provide the outline of the ANDS documentation on data protection.
- Confidentiality agreements in place (DS-4)	ANDS has Confidentiality agreements that must be signed by all people using the database.
- Ensure data protection after end of project (DS-5)	This component is still under discussion by the Consortium. The FOT-Net Data Protection document has provided the framework for our discussions so far.
- Data sent between Data Centre and Analysis Site must be encrypted (DS-6)	ANDS data that is transferred between the VTTI repository and an Analysis Site is encrypted.

FOT-Net Recommendation	ANDS Actions
- Data downloads regulated by Project Agreements and informed consent from drivers (DS-7)	Data sharing of the ANDS data between VTTI and the Analysis Sites often involves downloading of at least part of the project data but only by Consortium members. This is regulated through the Access Permissions referred to in the next section on Analysis Sites and through Confidentiality agreements.
- Data extractions for specific purposes in accordance with consent forms and project agreement and documented (DS-8)	Currently, with the VTTI repository the main Data Centre, Analysis Sites will both download and extract the results of analysed data and access will be governed as described above. When UNSW becomes a Data Centre, the same will occur although it is also likely that many Analysis Sites will share extracted results of data analysis. The same access permissions will be required except where no personal information is involved including any video data or GPS data. It will be essential to manage access to this data in the same way as downloaded data and to document all activities.
b) Analysis Sites	
- Documentation of data protection implementation (AS-1)	Each Analysis Site will document the data protection procedures and practices implemented. As for the data protection practices for Data Centres, the Consortium will review the practices and documentation on a regular basis.
- Analysis work stations physically and logically protected (AS-2)	As for the Data Centres, work stations used in the analysis of personal or confidential data will be both physically and logically protected.
- Analysts trained in data protection and integrity issues (AS-3)	A training program is being developed based on the data protection and ethical issues relevant to the ANDS project.
- Confidentiality agreement for all analyst staff (AS-4)	All analysts and any individual working on any aspect of the data are required to sign a Confidentiality Agreement before commencing work on the dataset.
- Access requests to Data Centre administered by Supervisor (AS-5)	Permission for access to the Data Centre and the ANDS database is managed by the Chief Investigator of the ANDS Project and, if necessary with the support of the Consortium's Project Management team.
- Specified procedures for data extraction agreed and used (AS-6)	Currently, permission for data extraction is managed by the Chief Investigator and Project Management team through the Data Centre. This may be modified if the volume of requests increases markedly to ensure that the process is as efficient as possible and does not present impediments to timely analysis.

FOT-Net Recommendation	ANDS Actions
- Analysis site must not extract or re-distribute data (AS-7)	Further distribution of data extracted will not be permitted for ANDS. This will be covered in the training material and agreement on use of the extracted data.
- Project data must not be used for research areas not covered by consent forms in the project (AS-8)	The ANDS data may not be used for research areas of purposes not covered by the Data access and protection rules including Consent forms.
- Visitors/guests to Analysis site should sign confidentiality agreements (AS-9)	Database users from outside the ANDS Consortium will be required to sign Confidentiality agreements and to undertake the same training as all analysts in the Consortium. Individual and organisation-level Confidentiality agreements will be required.
- All post-project research should be evaluated for need for further ethics approvals (AS-10)	Post-project use of the ANDS database is currently being reviewed. It is likely to require data protection and access procedures and practices to the same level as are required for the Consortium partner researchers. Addressing potential ethical issues will be a required component of the agreement procedure for any new partners including those who are only using the existing databases as we recognise that these issues must be addressed as part of a research plan and before the research is commenced.

FOT-Net Recommendation	ANDS Actions
5. Data and Metadata description	
– Definitions and Data and metadata in general	1. Study design and execution documentation: ANDS has clear study design and methods documentation which was developed collaboratively amongst the partners and is updated as needed and communicated to all parties. 2. Descriptive metadata: ANDS has documentation on all variables contained in the database, as follows: • ANDS data includes Context, Acquired/derived data and Aggregated data • Description of meta characteristics of all data provided in OzSight (VTTI) including vehicle, trip and demographic data includes: – Variable name – Description – Variable type – Source – Units (metric, standard) – Response codes (description of code) – Notes

FOT-Net Recommendation	ANDS Actions
	• ANDS/VTTI Data characterised as: Vehicle data, Trip Summary data, Time series data, Event detail, Driver characteristics (demographics, driving history, Visual and cognitive tests, Physical strength tests Sensation seeking, DBQ, Sleep habits, Impulsiveness tests) • Continue with data description for all new data collected and created variables • Documentation of this process. • Data from the Mobileye unit is described in the same way as the OzSight data using the same meta characteristics. As described above, however, this data is stored separately and treated separately to the VTTI data. • For data that is commercial and currently under development from Seeing Machines the ANDS Team is working with the company to document the metadata for all variables collected. This is currently in progress. 3. Structural metadata: ANDS is currently developing the structural framework for the database to be held in the UNSW Data centre. All aspects relating to the design of the database will be clearly documented. 4. Administrative metadata: ANDS has clear procedures established for access to the database held at VTTI for Consortium partners, although this is yet to be tested fully as the data has only recently become available and only a few analysts have access at the current time. The procedures for access to the Australian Data Centre are currently being developed.

Comment on the Data Sharing Framework

Overall, the FOT-Net DSF provided an excellent benchmark against which the development of the ANDS could be assessed. Given that the ANDS project is reasonably well progressed and has done so independently of the FOT-Net Data project, it might have been expected that there would be considerable differences between the way ANDS has developed and the recommendations from the DSF. It is evident that this is not the case. The DSF provides a check against which the ANDS partners can review and assess the completeness of their own data management and protection and project documentation in order to ensure that this project will achieve one of its important aims, that of data sharing. Furthermore, the DSF will aid the ANDS Consortium in the remaining planning and decision-making for the ANDS project.

The FOT-Net DSF documentation is most comprehensive and in the main readily understandable. One of the few places where problems were encountered was the Financial models section where the models are presented as if they are mutually exclusive. Consideration might be given to modifying this section. As shown in Table 4, the characteristics of each of the models may be highly interchangeable, for example dataset funding may be primarily by the hosting organisation (Model A) but may have costs covered through offering analysis services (Model G). This could be reflected more in the document.

In the main, this investigation has shown good compatibility between the FOT-Net Data recommendations and existing data access and management structures and activities of the ANDS project. A few issues have been identified that need to be addressed by the ANDS Consortium that relate mostly to balancing freedom of access to data and privacy as well as use by third-party or commercial entities. On these issues, the DSF documents again provide a useful starting point for development of extended and specific permissions for access.

Under the current ANDS Agreements, there are limitations on access to data generated from devices owned by industry partners. Further work is likely to be needed to refine these agreements as more in-depth analysis is conducted using these commercially owned devices and the ANDS-owned DAS data and where analysis might be by third-parties outside the Consortium.

As highlighted in other sections of this report, issues are also likely to be encountered in developing further data protection procedures and practices for new users outside the Consortium. The ANDS Governance Group will be the focus of these activities and issues that will need to be addressed will include whether changes are needed and the nature of any changes to the current data access practices and the procedures and timeliness of any permissions to download or extract data.

Investigate joint data analysis possibilities with Europe

There are considerable opportunities for joint research and data analysis of naturalistic driving data between Europe and Australia. There is great potential for research on common topics between Europe and Australia. Viewed from the perspective of the ANDS project, joint projects with Europe will open the possibilities to expand the scope of our research in a number of areas and to validate others. In the areas of normative data and human factors data, the possibility to compare how drivers in Australia and different parts of Europe manage the demands of driving and remain safe and/or within the driving regulations may reveal some differences that help our understanding of the driving task. It may also validate findings on normal driving behaviour that were found in both regions of the world. Joint projects will also provide more information about the nature of exposure to the road environment and how they influence driver behaviour. There are considerable differences in the road systems in Australia and many parts of Europe. Joint projects on the effects of these differences on natural driver behaviour are likely to reveal more than has been known before about effective and less effective aspects of such factors as road system design and road rules.

There is a wealth of common research questions to be investigated re-using already collected data such as the datasets from ANDS or the European NDS project UDRIVE. If planned ahead in project proposals, funding for accessing existing datasets could be included in proposals, which could contribute to solve the issue of funds for maintenance and availability of rich, high value datasets.There are multiple possibilities for joint research in this area. As is the case for most research, the development of joint projects on naturalistic and field-operational data between Europe and Australia will most likely stimulate further ideas for research into areas not previously considered.

Conclusions and future directions

The FOT-Net DSF has been a most valuable resource for establishing a data protection and sharing platform for the ANDS Project. The DSF provides a comprehensive and usable approach to establishing datasets that facilitate data sharing and maintain privacy and ethical standards for access. Overall, the recommendations of the DSF addressed decisions already made for the ANDS project and so were entirely consistent and compatible. On a few issues the DSF documents have highlighted areas where decisions are still to be made. In these cases, the DSF has provided guidance on a number of aspects of the ANDS database that need further development by showing the issues that the ANDS Governance Group will need to address.

For the future, the FOT-Net DSF recommendations and guidelines will enhance the possibilities of collaboration between groups that use them in developing FOT or NDS Databases. This means that it will be possible for ANDS to collaborate readily with other NDS projects such as the European UDRIVE project in terms of researching mutually interesting driver behaviour questions. Applying the DSF recommendations and guidelines is quite feasible and they provide the necessary common approaches to establishing data exchange protocols regarding security, terminology and data structure to make data sharing a real possibility.

UDRIVE

European Naturalistic Driving Study

EUROPEAN COMMISSION
SEVENTH FRAMEWORK PROGRAMME
FP7-SST-2012.4.1-3
GA No. 314050

eUropean naturalistic Driving and Riding for Infrastructure and Vehicle safety and Environment

Application of the FOT-Net's Data Sharing Framework on the UDRIVE dataset

Written By	Helena Gellerman (SAFER) Karla Quintero (CEESAR) Erik Svanberg (SAFER) Clement Val (CEESAR)	02-02-2017
Status	Final	02-02-2017

Table of contents

1 Introduction

FOT-Net Data is a Support Action for international co-operation that targets efficient sharing and re-use of global data sets. It continues European and international networking activities in the domain of Field Operational Tests (FOT) which started with the FESTA project.

The FESTA methodology was developed to provide guidelines to carry out FOTs and naturalistic driving studies. FOT-Net and FOT-Net 2 followed, and aimed at maintaining such methodology and providing the proper procedures in order to help stakeholders implement the proposed methodology and share their experience within an FOT network. The following project, FOT-Net Data, explicitly addresses the need to continue using collected data after their corresponding initial project, and therefore addresses data sharing issues. The prime goal of FOT-Net Data is to develop and promote a framework for sharing and exploiting collected FOT datasets in national, European and other international FOTs (e.g. US, Japan and Australia). FOT-Net´s Data Sharing Framework takes into account the pre-requisites necessary in the FOTs, such as legal agreements, to enable future re-use of collected data. Additionally, it addresses the procedures, templates and services needed for successful sharing of data. FOT-Net Data also developed and builds a detailed catalogue of available data, enabling organizations to easily assess the value of different data sets for their research purposes.

FOT-Net Data includes stakeholder and expert groups playing a key role in previous and ongoing FOTs.

The aim of this report is to summarize the application of the Data Sharing Framework to the UDRIVE project, currently in the data analysis phase. The report covers how the UDRIVE project addressed the challenges identified in the Data Sharing Framework, including a discussion on findings in the UDRIVE implementation that could be used to enhance the DSF.

2 The UDRIVE project

2.1 Overview

UDRIVE is the first large-scale European Naturalistic Driving Study on cars, trucks and powered-two wheelers. The acronym stands for "European naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment". UDRIVE is co-funded by the European Commission, DG Research and Innovation, in the 7th Framework Programme.

The purpose of the UDRIVE project is to increase the understanding of road user behaviour. Its objectives are two-fold: to identify well-founded and tailored measures to improve road safety and to identify approaches for reducing harmful emissions and fuel consumption in order to make road traffic more sustainable.

The UDRIVE project is building on the experiences from previous European FOT projects and aims to contribute to developing this in-depth knowledge by conducting a large-scale European naturalistic driving study (NDS). The project establishes one central database with all the collected driving data accommodating the analysis on road safety and environmental impact remotely and finally leaving behind the collected data to be re-used, subject to legal and ethical constraints, for additional analyses once UDRIVE is finished.

Targeted analyses are performed in four areas of interest, crash causation and risk, distraction and inattention, vulnerable road users and eco driving. The findings obtained in these topics will allow the evaluation of new and promising countermeasures, the potential of simple DAS for monitoring performance indicators over time, the improvement of driver behaviour models for road transport simulation and the possibilities for commercial applications of naturalistic driving data.

The UDRIVE consortium consists of 19 partners and represents a good balance between different EU regions and various stakeholders. The members of the consortium are: Institute for Road Safety Research (SWOV, coordinator), the Federal Highway Research Institute (BASt), Transport Research Centre (CDV), European Centre for Studies in Safety and Risk Analysis (CEESAR), Research and Development Centre in Transport & Energy (CIDAUT), National Research Centre for Aeronautics and Space (DLR), ERTICO ITS Europe, International Automobile Federation (FIA), Road and Bridge Research Institute (IBDIM), Institute of Science and Technology for Transport, Development and Networks (IFSTTAR), Road Safety Board (KFV), Laboratory of Accidentology, Biomechanics and Driver Behaviour (LAB), Loughborough University Transport Safety Research Centre, Or Yarok association for safer driving (OY), Vehicle and Traffic Safety Centre at Chalmers (SAFER), Professorship of Cognitive and Engineering Psychology at Chemnitz University of Technology (TUC), Organisation for Applied Scientific Research (TNO), Institute for Transport Studies (ITS) at the University of Leeds and Volvo Group (Volvo).

The composition of the consortium is such that all aspects of conducting a large-scale NDS are covered: the research and methodology aspects, e.g. related to defining verifiable research questions and performance indicators, an adequate study design, and valid analysis methods; the technical aspects, e.g. related to data acquisition, data management, data transfer and storage, and data processing and analysis; the overall expertise in the safety and the sustainability topics that are covered in the proposed study; and the specific expertise in both the engineering and the human factors related to road, vehicle, and road user as well as their interactions.

2.2 Data management

The data management from data collection to analysis in the UDRIVE project is presented in Figure 1. Data collection is managed by operation sites (OS), each one handling a fleet of vehicles instrumented with a data acquisition system (DAS). Data is stored in hard drives which are changed regularly.

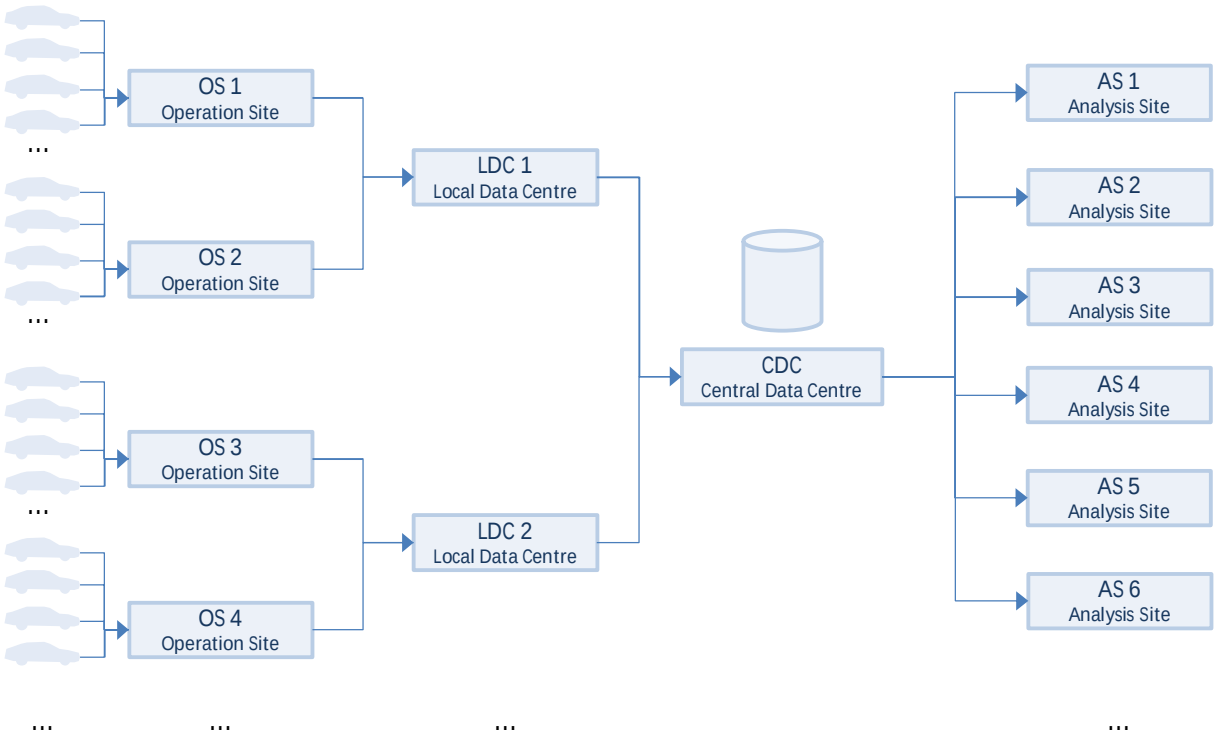

Figure 1. Organisation from data collection to analysis in the UDRIVE project.

The collected hard drives are transferred from the OS to one of three local data centres (LDC). An LDC can manage data from several OSs and it performs pre-processing of the received data. The central data centre (CDC) then collects all of the pre-processed data produced by all of the LDCs and uploads it into the database where the data from the project is stored and managed. The CDC hosts a virtual desktop infrastructure (VDI) which together with the analysis tools allows the analysis sites (AS) to remotely visualize/annotate/query/analyse the dataset. Data query, visualization, and analysis are performed in a joint environment, i.e. developments are available to every user accessing data. UDRIVE's organization includes: one CDC, six OSs, three LDCs, and fourteen ASs. A more detailed description is given in the Data Protection chapter.

2.3 Dataset

The UDRIVE project is collecting a dataset originating from six OSs located in: France, the Netherlands, Germany, the United Kingdom, Poland, and Spain. The data collection covers 120 cars, 32 trucks, and 40 powered two-wheelers (PTW). The distribution of the vehicles and also participant overview is described in the following table:

Table 1. Gender and age distribution for UDRIVE participants

	Vehicles	Participants	Gender M / F		Age 20-29	30-39	40-49	50-65
CARS								
DE	20	27	66%	33%	19%	26%	4%	52%
FR	30	43	46%	54%	9%	28%	30%	33%
PO	30	31	71%	29%	6%	48%	39%	6%
UK	30	52	48%	52%	15%	29%	19%	37%
NL	10	33	55%	45%	9%	30%	27%	33%
TOTAL	**120**	**186**						
TRUCKS								
NL	32	48	98%	2%	6%	13%	31%	48%
PTWS								
ES	40	47	74%	26%	9%	55%	34%	2%
TOTAL		AVERAGE						
		281	66%	34%	11%	33%	26%	30%

The acquired data include following types:

- CAN;
- accelerometer;
- Mobileye;
- video:
    - car data have seven views: forward left, middle, and right, driver face, cockpit, cabin and pedals (an example of the video views for car data is presented on Figure 2.);
    - truck data have eight views: forward left, middle, and right, left and right blind spot, driver face, cockpit and pedals;
    - PTW data have five views: forward, back, left and right side, and one view of driver upper-body and head;
- GPS;
- and questionnaires.

Figure 2. Recorded video views for car data

This data is transformed during pre-processing: geographical data with relevant indicators is integrated based on the GPS location (map-matching), CAN data is decoded, data between dissimilar vehicles is harmonized, and some signals are resampled over common and regular timestamps.

The pre-processed data is then sent to the CDC where data processing is performed, based on the need of the analysis. This data is also enriched by manual video annotations. Also automatic video annotations are being explored. The process and the description of the data is presented in section 3.2.2.

2.4 Data sharing approach within UDRIVE

As exposed by its star-shaped organization around a single dataset, UDRIVE is by definition a data sharing project. Data management is centralized since all the pre-processed data is stored and managed by the CDC. The CDC provides remote access to all analysis sites which develop algorithms to process data and subsequently obtain data. All of the new derived data developed by a particular user (in an analysis site) remains in the CDC and are remotely available for all other users. All algorithms accessing the data in the CDC in order to create new derived data can only be run by being stored at the CDC and are also available for all other users. In this sense, UDRIVE is not only sharing data but also analysis tools and scripts.

3 Application of FOT-Net's Data Sharing Framework in UDRIVE

This chapter describes how UDRIVE addressed the challenges identified in the Data Sharing Framework (Gellerman et. al, 2017). First, the agreement-related challenges will be summarized, followed by the more technical topics related to data and metadata description and data protection. The training and support and research services are discussed and finally the financial models and application procedures are addressed.

UDRIVE was developed almost simultaneously with the Data Sharing Framework. The following illustrates the framework put into practice, both identifying its strong points as well as the areas in which further work and adaptations need to be carried out.

3.1 Project agreements

The following table lists the items that are considered as essential in the data sharing framework and that should be covered in an FOT or NDS, as well as the practical answers based on the UDRIVE experience for the applicable items.

3.1.1 Funding agreement including the description of work

The funding agreement together with the description of work is one of the main agreements to pay attention related to possible funding for data sharing to according to the DSF. The complete list of topics brought forward by the DSF are summarized in Table 2 together with the corresponding feedback.

Table 2: Funding agreements in FOT-Net DSF vs UDRIVE

Topics to be addressed according to the DSF	UDRIVE Feedback
• Be aware of the topics and issues to be discussed in relation to data sharing and re-use of data. • Focus on these issues during the project application and during a possible negotiation phase. • Pay attention to the possibilities to provide open data after the project, based on the scope of the project and the data to be collected. • The DOW could include a list of topics of interest (specified in the following text) and should pay special attention to ensuring funding for the post-project phase in which the dataset can remain available for data sharing after the project.	• Many discussions were held during the proposal phase regarding data sharing, as the idea of re-using the data was present from the start based on the project partners´ experiences. Many of the topics were incorporated in the description of work such as third party access and that the data would be available conditioned by the availability of funding to maintain the data. • Open data in the sense that third parties would be allowed to have access to the data was incorporated into the proposal. The data would though reside within the CDC to protect personal data and commercial confidential data. The project funding agreement did not include the post-UDRIVE funding to keep data available.

Lessons learnt:

It is important to discuss the possibilities for funding of the data after the project during the funding agreement preparation. In UDRIVE, this has become a challenge in the current phase of the project and is introducing uncertainties if the data will be able to be available for re-use after the project. Several business models are being studied to investigate funding possibilities.

3.1.2 Consortium agreement

The consortium agreement is the main document stating how the consortium has agreed to handle the main topics allowing potential data sharing to take place. UDRIVE had long discussions regarding the different

topics brought forward by the DSF. Table 3 shows the many questions to ask before starting the project and the answers that UDRIVE partners agreed on during this phase of the project.

Table 3: Consortium agreement in DSF vs UDRIVE

Topics to be addressed according to the DSF	UDRIVE Feedback
Ownership and access to data and data tools • Who owns the data? Separate ownership clause? • How could the data be used and on which conditions? Will all partners have access to all/part of the data? • May the data be licensed to third parties? • May third parties have access to the data and if so, are there special conditions? • Are there constraints on personal data, especially video? • Are there future agreements with data providers to take into account? • Who will own the analysis tools and on what conditions are they licensed during and after the project? • Has a partner included previous work as background in the tools? Who has the IPRs on those?	• All partners own all data jointly. • Data cannot be licensed to third parties. • All partners have access to all data. • Video and GPS data will only be available at partners' facilities. • Third parties will have remote access from their own premises to all data except video and GPS, which can only be accessed via a UDRIVE partner with remote access to the CDC. • The tools are owned by the organisations developing them. Several tools include background from previous developments. The partners have access to the analysis tool during and after the project. Third party has access to a viewing tool. • Some partners include previous algorithms as background and those are therefore encrypted.
Storage and download of data • Where will the data be stored, centrally or distributed? • What are the general requirements on data protection and how are they assured? • Shall all data/part of the data be downloadable for all partners and if so, under which conditions? • Shall all data/part of the data be downloadable for third parties and if so, under which conditions? • Is there a time limit to request data for download? • Is there a time limit for keeping the data?	• Data is stored centrally. No data can leave the CDC except at data download by a UDRIVE partner. • Each partner can download the data and become a host of all or part of the data. They must become a partner data centre (PDC) and fulfil the same requirements as the CDC. • Data protection is addressed in the chapter Data protection and was ensured following the DPC. • During the project, partners can request anonymized data extractions managed by the CDC. These extractions are managed by the CDC. • Data remains stored at the CDC for 3 years after the project's completion. Funding is needed to make it available. • Different data have different time limits before deletion, such as 0, 5, 12 and no time limit.
Access methods • Can the data be remotely accessed at the premises of any partner? • Shall a specific access procedure be used and	• The data can be accessed remotely by partners that have applied and have been validated as analysis site, following data protection recommendations of the DSF. • The access procedures are managed by the CDC.

managed by whom? • What are the requirements on data protection for partners/third parties analysing the data?	• The same data protection requirements are applying to partners as well as third parties.
Areas of use • Shall it be possible to use the data for both education, research and commercial purposes? • Are there special conditions on the use?	• UDRIVE data can be used for education, and research purposes.
Post-project re-use of data • Which partner is responsible for maintaining the data after the project? • Which application procedure shall be used? • Who will grant access to data after the project?	• The CDC remains as data provider in the Post-UDRIVE phase, if sufficient funding is available. • The application procedure is not yet ready as the funding is not yet available data maintenance.
Post-project financing • How will the storage and support services for data re-use be financed after the project? • Known or to be decided? • How will this funding be distributed?	• No external funding is provided for UDRIVE data access after the project • UDRIVE partners are investigating possibilities to fund the data.

Lessons learned:

It is important to be make it clear in the consortium agreement, who owns data and tools and on which terms they can be used. It is also strongly recommended to take the extra time and effort to solve things when discussing the consortium agreement. It makes it easier later on in the project. Furthermore, be clear how the data will be handle during and after the project. It enhances a common understanding of the project and how it will be set up. It also avoids unnecessary discussions in the project.

3.1.3 Participant agreements

UDRIVE paid a considerable effort in aligning the different participant agreements in the project, as the project performed an pan-European collection of data, all ending up in one common data center, the CDC. Table 4 shows how UDRIVE addressed the topics put forward by the DSF for the participant agreement.

Table 4: Participant agreements in DSF vs UDRIVE

Topics to be addressed according to the DSF	UDRIVE Feedback
• Describe the project to some detail • Explain how data is treated and being distributed • It is recommended that the participant actively consent to vital aspects of data sharing, such as re-use areas, what data could be used publicly and if third party researchers could re-use the data	• It includes text on: the project, all of the phases of the study (including data collection), who can access data during and after the project, protection of personal data, deletion processes. • The participants actively consent to some specific questions in line with the recommended content in the DSF. • A template (for all countries collecting data) was developed on common data sharing pre-requisites. It was then adjusted to national legislations and

	translated. The draft final templates were then gathered and checked that the specific data sharing content was still present. Conditions had been inserted in some countries.

Lessons learnt:

It is really necessary to start from a common template if data later should be commonly shared. The project has used large efforts in trying to stick to the commonly agreed template participants agreement. This is something that should be emphasised in the Data Sharing Framework.

3.1.4 External data provider agreements

Based on previous experience regarding restrictions to use external data providers' data after the project, these agreements were handed carefully. Table 5 showv the outcome of these discussions related to the DSF recommendations.

Table 5: External data provider agreements in DSF vs UDRIVE

Topics to be addressed according to the DSF	UDRIVE Feedback
• What is regarded as confidential information and what can be shared? • Can confidential data be anonymised/changed/aggregated, to allow for more open access? • Can the data be accessed by another project partner/third party? • Can the data be transferred to another project partner/third party? • Are there restrictions on what the data can be used for? • Are there special conditions for sharing and re-using the data after the project? • What happens if the external data provider is bought by another company?	• A non-disclosure agreement was signed with Renault regarding the DBC files for the vehicles in the study, allowing CAN data to be interpreted. • For the map matching solution, the tools and geographical database are licensed to LDC partners. The solution is valid for the duration of the UDRIVE project and can be extensible for usage after the project completion with new contracts with the supplier. • The resulting map data is valid to be used by all UDRIVE partners and public research community • An agreement was signed with Mobileye allowing re-use of the data after the project

Lessons learnt:

The necessity to discuss the data sharing issues should be underlined in the DSF. The topic list gave good guidance in what to focus on. If these issues are not solved during the project, it would not be easy to go back after the project and ask for additions to the agreements or new agreements.

3.1.5 General lessons learnt from project agreements discussions

Clear agreements early in the project solved many later discussions.

Agreement discussion together with the Data Protection Concept efforts, enhanced the awareness regarding what sensitive data are and what is needed in the form of detailed data protection implementation among the partners. This indirect outcome should be enhanced in the DSF, as it fostered an understanding of why certain somewhat restraining procedures were implemented.

Legislation in different countries affected the data that could be shared. Be sure to start the legal and ethical process early on to avoid unnecessary time delays in the project.

3.2 Dataset description

Data sharing framework identifies four categories of metadata that need to be properly documented in order to allow successful sharing of a dataset:

- *FOT/NDS study design and execution documentation*, which corresponds to a high-level description of a data collection: its initial objectives and how they were met, description of the test site, etc.
- *Descriptive metadata*, which describes precisely each component of the dataset, including information about its origin and quality.
- *Structural metadata*, which describes how the data is being organized; and
- *Administrative metadata*, which sets the conditions for how the data can be accessed and how this is being implemented.

We describe in this section of the document how each of these aspects is documented within the UDRIVE project.

3.2.1 FOT/NDS study design and execution documentation

UDRIVE is a Naturalistic Driving study, not an FOT, so no experimental plan with e.g. different phases was implemented, and as a result, documenting this is not applicable in that case. For the other items that the Data Sharing Framework recommends documenting, such as research questions, sample selection criteria and description of recruitment, description of data collection environment and periods etc. several deliverables exist, with extensive information. However, information which would be useful to reuse data is spread over a large amount of such deliverables, each of them covering different aspects, some of them covering originals intents and plans, others describing actual implementation and execution. Allowing efficient data reuse would require aggregating all relevant information in a single, synthetic, document. It is expected at the moment that such a document will be produced shortly before the end of the project.

3.2.2 Descriptive metadata

Descriptive metadata is the type of metadata which is most developed in the data sharing framework, and it is also one of the aspects regarding data which has been best documented in UDRIVE. It is recommended to provide *"the characteristics of each measure or component, but also the origin on how the data was generated and collected"*, and to keep that information *"close to the actual data to facilitate analysis"*. The ambition in UDRIVE has been to, as far as possible, use and transform the data in a systematically controlled way. Complete traceability is ensured on any Record (set of continuous data acquisition by an in-vehicle data acquisition system, ideally corresponding to a trip) generated in a vehicle, from data collection to analysis, thanks to several pieces of software, some of them developed specifically for the project. From the moment data have been uploaded in the CDC database, full tracability of any value is supported by the analysis software. This toolchain is presented in Figure 3.

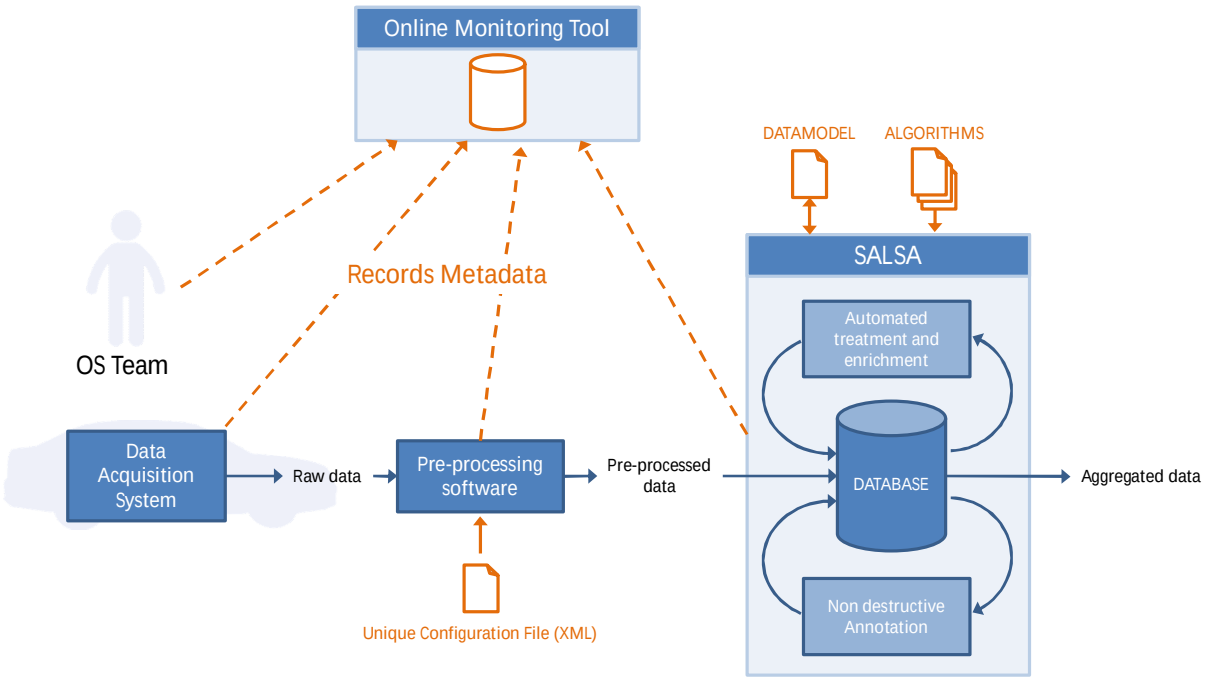

Figure 3. Data lifecycle and tool chain in UDRIVE. Metadata is represented in orange.

- The Online Monitoring Tool (OMT) closely tracks the lifecycle of each Record. From the initial data collection to import in database and through pre-processing to the CDC, the status of each Record but also of each vehicle, driver and hard-drive used for collection or transfer is monitored. The corresponding metadata is centralized in a database managed by that tool. Therefore, for each Record it is possible to know precisely in which vehicle and with which equipment and configuration it was collected. The underlying database also makes it possible to query the current status of a Record or the overall picture in data processing. The two functions has proven to be a valuable tool for technicians at the operation sites, for the local and central data centres to assess how much data that is to be processed, as well as for project management. The OMT is also used to organise the deletion of data.

- The pre-processing software performs various transformations on raw data (CAN data decoding, harmonisation of data between different vehicle types, synchronisation, resampling, map-data integration, quality checks) to generate harmonised files. The pre-processing software also integrates into each Record the most relevant metadata from the online monitoring tool, so that all traceability attributes are directly available in the dataset itself. This software's behaviour is entirely based on a unique configuration file, which defines both the content of pre-processed data files, and the process implemented to obtain each signal. As a result, the original UDRIVE dataset is entirely and unequivocally defined in that configuration file, in XML format, as recommended in the data sharing framework. It is important to note, though, that parts of that configuration file (link between raw CAN signals and their representation in the final dataset) are covered by non-disclosure agreement, to protect intellectual property of car manufacturers and sensor providers. As a result, an expurgated version which only keeps the data description parts was also made, and is named the data manifest. An excerpt of that file is represented in Figure 4 below.

```
    <!-- Record content definition -->
    <record>

        <attribute name="Driver" datatype = "reference" datatype_ref="REF_Drivers"/>
        <attribute name="Vehicle" datatype = "reference" datatype_ref="REF_Vehicle"/>

        <datasegment name="FullRecord">

            <timeseries name="TS_10Hz Signals">
                <preprocessing timesource="RESAMPLE" subsampling_factor="1"/>
                <signal name="Veh_Longi_Speed" datatype = "single" unit="km/h"/>
                <signal name="Veh_Engine_RPM" datatype = "single" unit="rpm"/>
                <signal name="Veh_Steering_Wheel_Angle" datatype = "single" unit="deg"/>
                <signal name="Veh_Steering_Wheel_Rotation_Speed" datatype = "single" unit="deg/s"/>
                <signal name="Veh_Longi_Acc" datatype = "single" unit="m/s2"/> <!--<range: -10 m/s2 to 2.7 m/s2 >-->
                <signal name="Veh_Yaw_Rate" datatype = "single" unit="deg/sec"/>
                <signal name="Veh_Odometer" datatype = "single" unit="m"/> <!--<range: ClioIII 0 to 6553.5 m, MegIII & Clio
                <signal name="Veh_Throttle" datatype = "single" unit="Arbitrary"/>
                <signal name="Veh_BrakePressure" datatype = "single" unit="bar"/>
                <signal name="Mob_PCW_PedDZ" datatype = "reference" datatype_ref="NV_Pedestrian_Collision_Warning"/>
                <signal name="Mob_General_pedestrian" datatype = "boolean"/>
            </timeseries>

            <timeseries name="TS_1Hz Signals">
                <preprocessing timesource="RESAMPLE" subsampling_factor="10"/>
                <signal name="Veh_Distance_Totalizer" datatype = "double" unit="km"/> <!--<accuracy: ClioIII 1km, MegIII &
                <signal name="Veh_Fuel_Level" datatype = "single" unit="l"/>
                <signal name="Veh_External_Temp" datatype = "single" unit="deg C"/>
                <signal name="Veh_Absolute_Time" datatype = "double" unit="mn"/>
            </timeseries>

            <!-- GPS -->
            <timeseries name="GPS_streamOutput_1Hz">
                <preprocessing timesource="RESAMPLE" subsampling_factor="10"/>
```

Figure 4. XML Description of pre-processed data

- SALSA (Smart Application for Large Scale Analysis) is then used to manage the dataset and perform all further steps in data reduction:
    - Calculation of derived measures
    - Detection of relevant events
    - Visualisation
    - Annotation
    - Database querying

A screenshot of the software is shown in Figure 5 below.

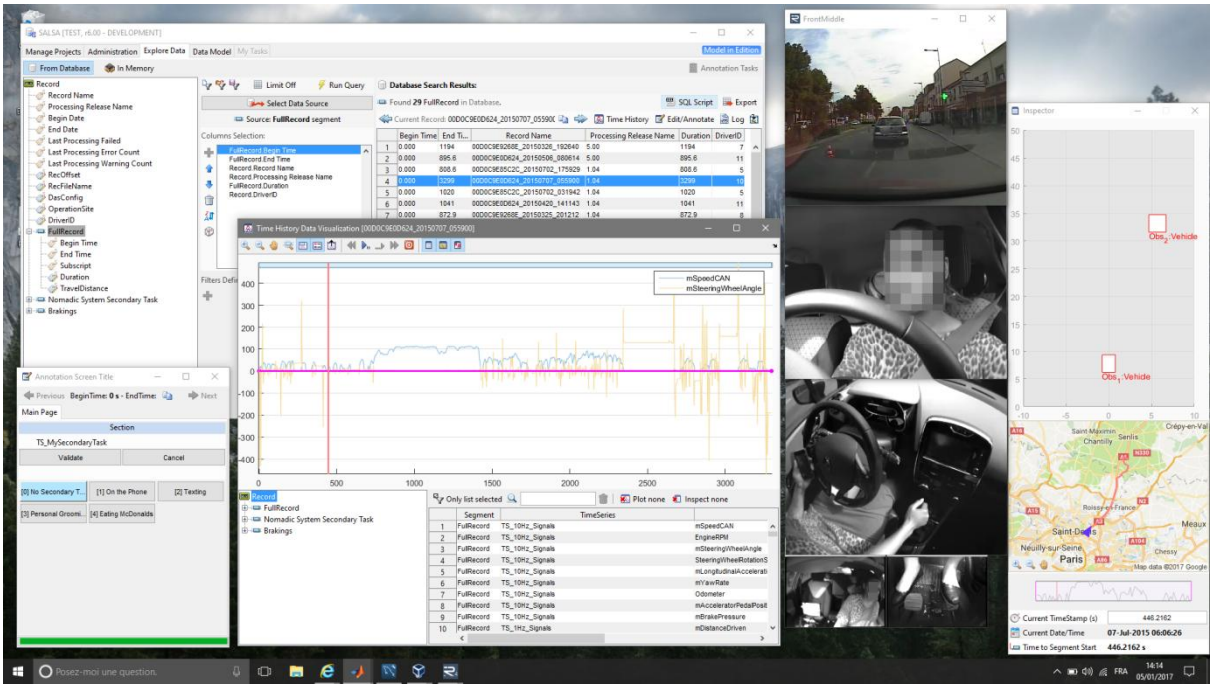

Figure 5. SALSA User Interface, as seen in a typical visualisation/annotation scenario

As SALSA is used to manage both the database and all data reduction processes, complete traceability of each single value is ensured: each value comes from a software-controlled chain of transformation including combination of original data, transformation through algorithms, annotations etc. which can all be browsed within the interface of the software.

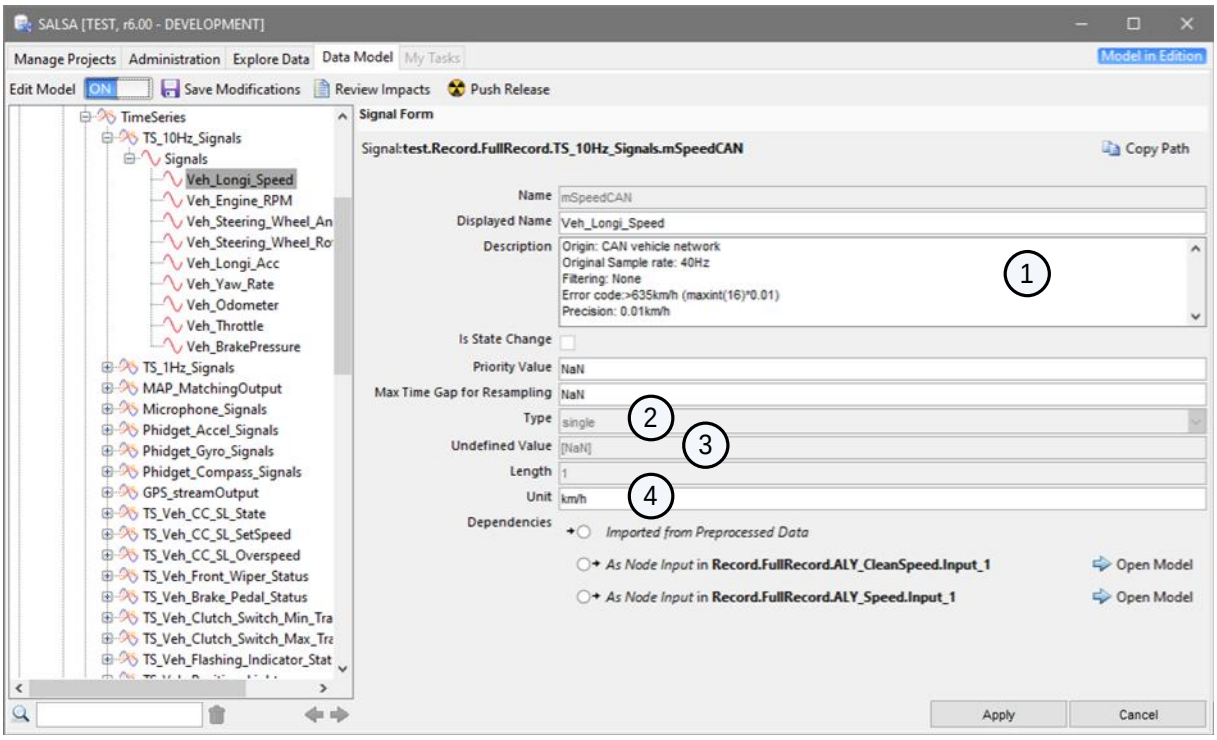

Figure 6. Metadata for a signal. (1) Free Text description, (2) Datatype, (3) Undefined value, (4) Unit.

This also included the integration of relevant descriptive metadata right in the software. Figure 6 above shows the metadata panel for the original longitudinal speed signal. Some of the recommended attributes from the data sharing framework, such as data type, undefined value, and unit directly appear as fields in the software. Other attributes recommended by the data sharing framework such as origin, original sample

rate, filtering, value corresponding to an error and precision do not appear as fields and had to be included in the free text description, which is a much less structured and probably insufficient way of doing that. This should be revised in a future version of the software.

As shown in Figure 7 below, labels (and potentially additional columns) for each value that a categorical variable can take, are documented in the software. Additionally, a code book for all annotations has been written. It comprises for each modality of annotated categorical variable, a precise description of its interpretation, including when applicable (e.g. driving situation or infrastructure), illustrations and/or example photographs.

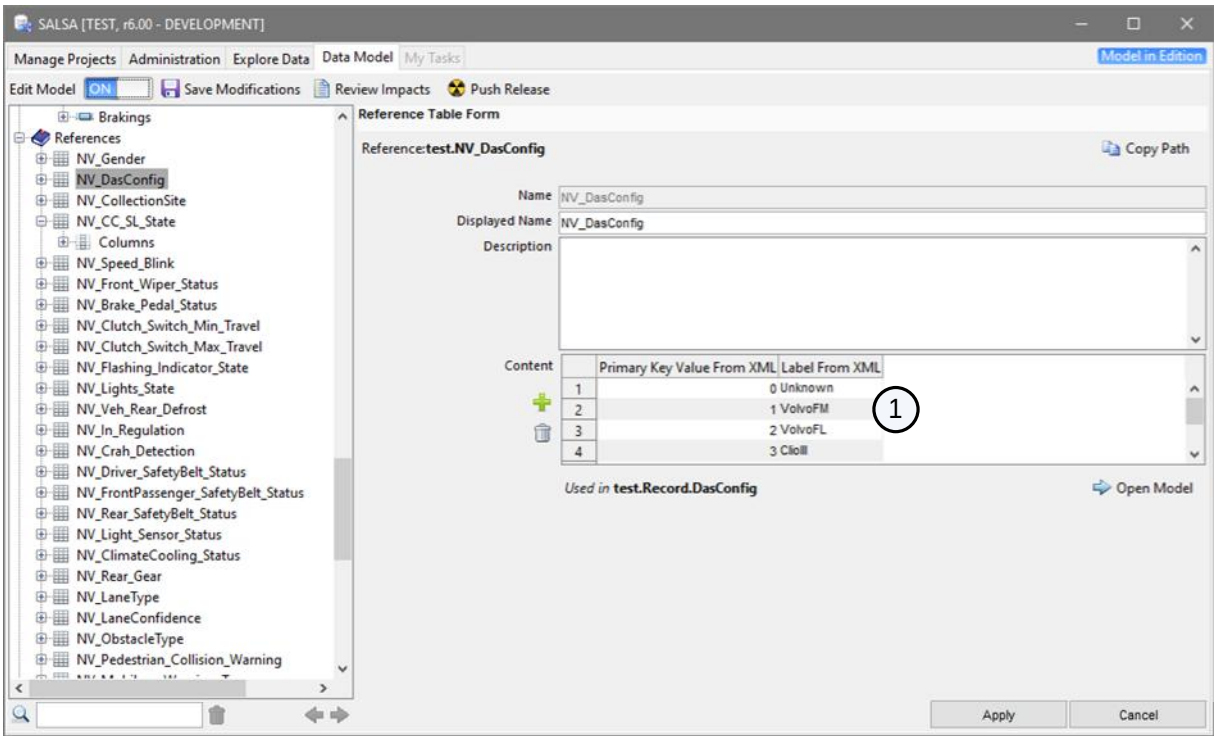

Figure 7. Metadata for a categorical variable, showing value/labels association for each modality.

Metadata is also included for each process: its links (inputs and outputs) with data are again part of the model developed using SALSA, and a free text description is available in the corresponding panel (cf. Figure 8). A template has also been created to document algorithms in their own code file (cf. Figure 9).

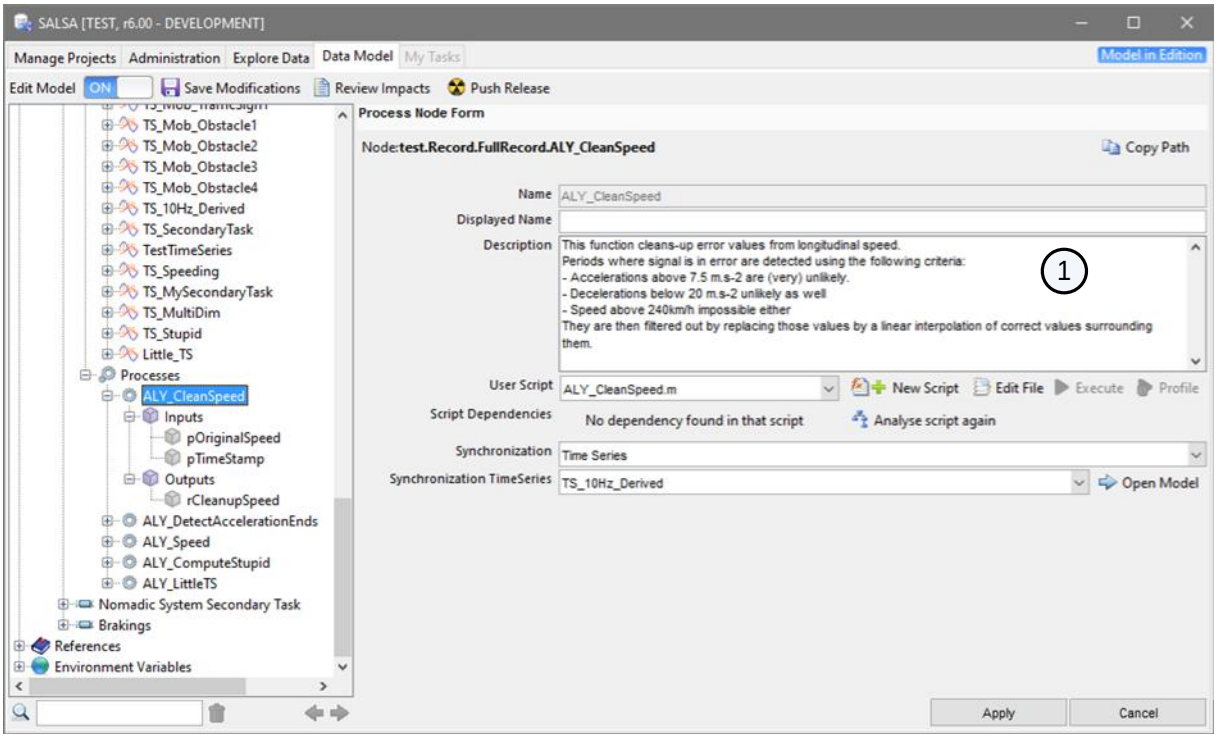

Figure 8. Metadata for an algorithm. (1) shows a free text description of the intent and principles of the algorithm.

ALY_CleanSpeed.m | TemplateForHeader.m

```
% ALY_CleanSpeed
%
% Description: Cleans up the measured speed signal.
%
% Measures needed:
%  - mSpeedCAN
%
% Function parameters:
%  - TimeStamp for synchronisation
%
% Function output: [e.g. dCleanupSpeed]
%
% Other m-files required:
%
% Main (initiating) developer: Clement Val, CEESAR [clement.val@ceesar.fr]
% Revision developer:
% Revision developer:
% Revision developer:
%
% Developed for WP/task: 4.3.3
%
% Code status: 'complete'
% External documentation: []


function dCleanupSpeed = ALY_CleanSpeed( mSpeedCAN, pTimeStamp )

    Acceleration = [NaN; diff(mSpeedCAN/3.6)./diff(pTimeStamp)];
    % Accelerations above 7.5 m.s-2 are (very) unlikely.
    % Decelerations below 20 m.s-2 unlikely as well
    % Speed above 240km/h impossible either
    to_dismiss = mSpeedCAN > 240 | Acceleration > 7.5 | Acceleration < -20;
    dCleanupSpeed = interp1(pTimeStamp(~to_dismiss), mSpeedCAN(~to_dismiss), pTimeStamp);

end
```

Figure 9. Source code for corresponding algorithm, showing template-based header.

Most of the important metadata is therefore stored in the database and accessible from the software. Such structured and centralized approach should probably be encouraged in the data sharing framework. Additionally, this would allow automatic generation of documentation, similar to what can be done with, e.g. Doxygen (www.doxygen.org).

It is important to note, though, that important, deeper knowledge about some signals is often missing, at least at the beginning of the project. As a matter of fact, data which is collected from vehicles networks or other sensors exist for specific purpose (typically driving assistance systems), and is rarely designed for general reuse. Accessing precise knowledge from the vehicle or sensor manufacturer about each signal's limitations, in which context and with which purpose it can or cannot be used is generally impossible. Although algorithms developers were encouraged to comment their choices in their own code, no formal process was developed in UDRIVE to share that knowledge within project partners, let alone capitalize it in a deliverable to help with later reuse. This aspect should probably be emphasized in the Data Sharing Framework.

3.2.3 Structural metadata

Structural metadata shall contain the description of the organization of the data. In UDRIVE, although video are stored as a hierarchy of flat files, numerical data (i.e. preprocessed data, derived measures, annotations etc.) is stored in a relational database. This database is managed by the SALSA software and therefore follows consistent rules: entities defined by users (e.g. segments, attributes, signals) to manipulate and store data are automatically structured and written in a repeatable way. Those rules are described in one of UDRIVE's deliverable. A simplified representation of SALSA's relational model is shown in Figure 10 below).

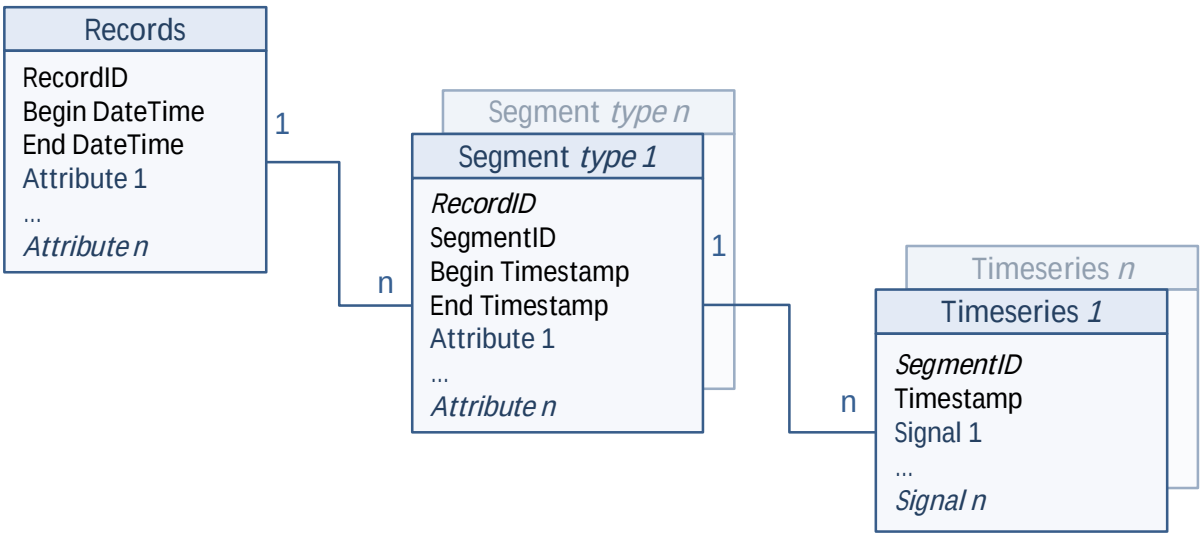

Figure 10. SALSA relational model

Video data files (and the necessary index files) are stored per Record with a reference in the Records table. The structure of the expected 100+ TB file system indicates where and when data were collected: file share / country / vehicle / year / month / record.

3.2.4 Administrative metadata

The intent of administrative metadata in the scope of data sharing could probably be further clarified in the data sharing framework, maybe through more practical examples. However, what seems crucial for data sharing is access rights management. Given the important effort carried-out to ensure proper *within-project* sharing of data, notably through the development of the UDRIVE Data Protection Concept (further described in subsequent section), this aspect has been thoroughly documented in the project and the respective implementations have been certified. The applications, adhering to the data protection concept, from each operation site, local data centre, central data centre, and analysis site, contain extensive information about data protection and access rules, including potential differences about data use conditions (e.g. end of life or specific restrictions) depending on local laws and authorizations. Additionally, non-disclosure agreements protecting intellectual property of, e.g. vehicle manufacturers or sensors providers, have been defined and signed either by the coordination on behalf of the project, or by individual partners who had direct access to protected data. The essence of this scattered information which is important for general data reuse could be consolidated in a single document to facilitate data reuse.

3.2.5 Discussion on data description

A lot of measures have been taken in UDRIVE to ensure consistency of the dataset, and the existence of the corresponding documentation. UDRIVE was designed from the ground up as a data sharing project, and the methodologies which have been developed for the project will certainly help in reusing the dataset. However, it is now clear that despite all those efforts, documentation is still insufficient for efficient reuse *after* the project. Notably, information is spread around a large number of documents which also contain both initial plans – including things which were finally abandoned in the course of the project – and information which was necessary during the project, but less important for data reuse. This information overload makes finding the important bits for reuse difficult, especially for users who did not participate in the project. Compiling those important bits in synthetic documentation was not planned (and as a result not funded) in the initial plan. Comparably, a dataset clean-up and consolidation was not planned either, and the tools, which facilitate access to the data, do not currently include sufficient user management / data protection, to allow new users to use the dataset independently from each-other. The preparation for reuse should therefore be integrated as a dedicated task in future, comparable projects.

3.3 Data Protection

3.3.1 Overview

UDRIVE handle personal data and confidential commercial data. Personal data must be handled according to the laws of respective EU country, while confidential commercial data are governed by contract with OEMs (Renault, Volvo) and external data providers (Mobileye). The requirements of handling this type of data is strict and therefore UDRIVE implemented the UDRIVE Data Protection Concept, to have a structured approach in securing the data.

3.3.2 UDRIVE Data Protection Concept

UDRIVE implemented detailed data protection guidelines of the DSF in the UDRIVE Data Protection Concept (DPC), formally approved by the second General Assembly. The UDRIVE DPC is tailored for the project, whereas the DSF is more general in the recommendations and requirements. Both projects have learnt from each other during the implementation of UDRIVE DPC and the development of the DSF.

The UDRIVE DPC describe general definitions of personal data and data controller, following the EU regulation (European Directive 95/46/EC Art. 2.), and also referred in the FOT-Net DSF. The DPC introduce a role “UDRIVE data supervisor”, as the person responsible for documenting and implementing data protection within each partner organization for the UDRIVE project.

The general processes and procedures also include: encryption, certification and its organisation, remote access, data extraction, data download and training.Where the DSF describes the general structure of Data Centre and Analysis Site, the UDRIVE DPC further describes stakeholders, data to be handled, functional requirements, external requirements and agreements, stated for the data handling steps following the data flow in section 2.4:

- Data acquisition system (DAS)
- Operations site (OS) and Vehicle adaptation site (VAS)
- Local data centre (LDC)
- Central data centre (CDC) and Partner data centre (PDC)
- Analysis site (AS)

The DPC finally states the data protection conditions for post-project data usage. This means that there are requirements stated for any occurance of personal or confidential commercial data in project. The requirements do not detail the exact implementation since there could be differences between partners, executing the same data handling step, as well as adapting to the legal context where data is being managed.

The UDRIVE DPC has procedures for handling data deletion. The on-line monitoring system is used to monitor that data is being removed accordingly. The data deletion procedure covers three scenarios:

1. Deletion request of certain records initiated by the participant.

   The participants are allowed to file a deletion request to their OS. Any organization hosting data is obliged to check for such requests periodically and remove the data.

2. Deletion of records where consent is not given.

   The driver identification process is at the CDC before making the data available for analysis. Any unidentified record will be deleted and thus not available for data analysis. Other organizations hosting data (i.e. the LDCs) are obliged to check for unidentified records periodically and remove the data.

3. Final deletion of personal data.

Depending on legal requirements in each country having an operation site, personal data must be removed after a specific date. The OS, the CDC and the LDC must remove personal data prior to this date.

The possibility to track the current state of each record in the dataset, is enabled by the online monitoring tool.

The UDRIVE DPC was presented at ITS 2015 in Melbourne (Gellerman et. al., 2016) illustrated as in Figure 11:

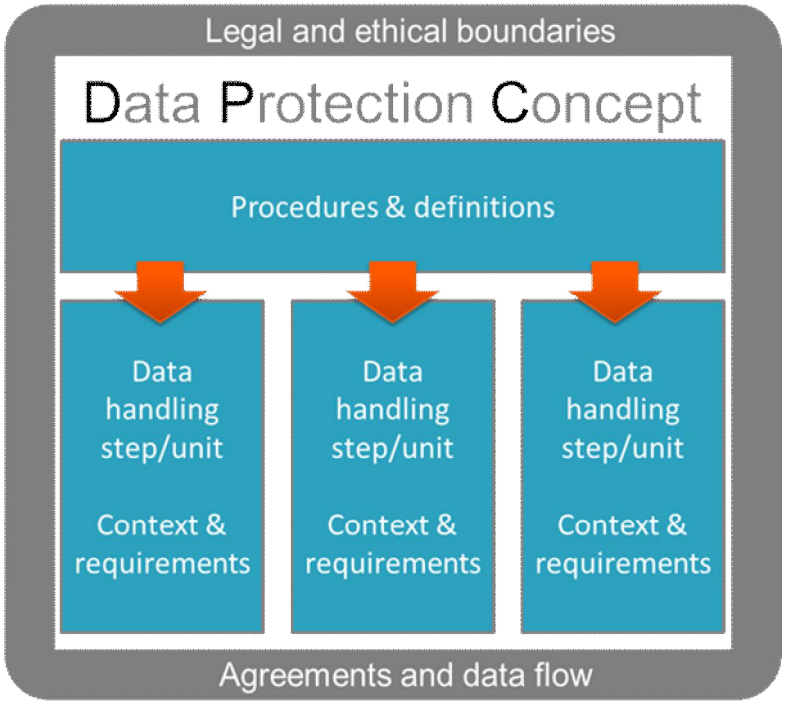

Figure 11: UDRIVE Data Protection Concept

This model describes how legal and ethical boundaries, and project agreements and data flow, will have impact on the overall procedures and definitions in the project, and for each data handling step that is decided. As for the DSF, it is impossible to create project specific requirements. Each project must therefore act within the boundaries, and define the necessary data protection producedures and requirements. It is recommended to do this in an early phase of the project, and even a simplified overview should be taken into consideration in a project application phase. Below the different data handlings steps are presented, with the respective impact on data sharing. The central data centre and analysis site are described more elaborate.

Data acquisition system

The UDRIVE DPC states functional requirements on the data acquisition system but these are not related to data sharing. Instead these requirements are to ensure IP rights of an OEM and the collection of personal data in a vehicle.

Operation site

Regarding data sharing with the LDCs, the CDC and the ASs, the OS condition two major issues: for how long personal data is allowed to be used and content of the participant consent form. The OS is the legal representative of the project towards the participants, and must ensure that the participant consent form agree with the legal and ethical setting of the OS. Any DC must ensure that the usage of data comply with the these two items.

Local data centre

The LDC is the partner managing the raw data of the UDRIVE dataset. The LDC is responsible pre-processing raw data including decryption, synchronization, re-sampling, map matching, harmonisation and data quality checks. The LDCs have the agreements with OEM regarding data usage, as well as the agreement with external data provider for map matching and data attributes.

Central data centre / Partner data centre

The CDC is the partner organization managing the complete UDRIVE dataset. The CDC is responsible for hosting the data analysis tool and providing remote desktop infrastructure, driver identification tool, file servers, and the database. The CDC is also responsible for hosting the online monitoring tool. The CDC is per FOT-net DSF definition a DC. If a partner in the project downloads the whole or parts of the data set, this partner must become a PDC. The PDC must comply with the same requirements as stated for the CDC. Table 6 describe how UDRIVE implemented data protection for the CDC/PDC, very much inline with the DSF's data protection of a Data Centre:

Table 6: UDRIVE data protection for central data centre

Data protection item	Data protection description
Stakeholders	The stakeholders of the CDC were identified as 1) UDRIVE data supervisor; 2) CDC leader; 3) drivers; 4) data administrators (persons managing the data at the CDC, including uploading, processing, and deletion); 5) IT services support personnel (persons operating the database, file servers, remote desktop infrastructure, or other IT infrastructure hosting the collected data); 6) transporter (the company that will ship disks to the CDC); and 7) analysts and annotators.
External factors or requirements	1) Data distribution is administered through the CDC during the project. Partners have virtual access under certain conditions (data extraction restrictions). 2) The CDC needs to maintain at least a basic IT infrastructure for at least three years after the end of the project. Data protection measures must be in place during this time. 3) Data used in research must be archived, stored and deleted for the time period regulated by law.
Data to be handled	The CDC will handle following data types: *1)* Processed data (including derived measures, performance indicators, and events from in-vehicle signals, external equipment (radar, accelerometer), and GPS, and road attributes derived from position); *2)* video (driver, cabin, and external views); *3)* audio (non-personal, just sound); *4)* annotations; *5)* questionnaires; *6)* participant metadata;

	7) non-participant metadata; *8)* vehicle metadata; *9)* audit logs; *10)* online monitoring data; and *11)* user account information.
Agreements	The agreements, if applicable, at the CDC were: 1) Agreement with external IT infrastructure provider (if applicable); and 2) agreement with external data provider (if applicable).
Requirements	DPC-CDC-1: Data stored and processed at the CDC must be protected from unauthorized access. DPC-CDC-2: Data stored and handled at the CDC must be protected from accidental deletion or corruption. DPC-CDC-3: The CDC is data controller for its data and must comply with national laws implementing the Data Protection Directive. DPC-CDC-4: Confidentiality agreements for any involved personnel must be in place. DPC-CDC-5: The CDC must perform data deletions of the data they administer, when required by law. DPC-CDC-6: The monitoring system must employ proper authentication and authorization procedures. DPC-CDC-7: Data protection must be ensured by the CDC organization after the end of the project. DPC-CDC-8: Data sent from CDC to a partner must be encrypted. DPC-CDC-9: The CDC must administer and document all data extractions according to the specified procedures.

Analysis Site

The AS, both during and after the project, is the final user of the data collected in the project. The AS will access the data by connecting to the CDC using remote desktop. As Table 7 shows, the implementation is inline with the DSF for an analysis site.

Table 7: UDRIVE DPC for analysis sites

Data protection item	Data protection description
Stakeholders	The stakeholders of the AS were identified as 1) UDRIVE data supervisor; 2) AS leader; 3) drivers; and analysts and annotators (persons performing annotations or analysis on collected data).
External factors or requirements	Data distribution is administered through the CDC during the project. Partners have virtual access under certain conditions (data extraction restrictions).
Data to be handled	The AS has remote access to data at the CDC including: 1) processed data;

	2) video; 3) audio; 4) annotations; 5) questionnaires; 6) participant metadata; 7) non-participant metadata; and vehicle metadata.
Agreements	The approval from ethics committee for intended research (if applicable).
Requirements	DPC-AS-1: The analysis work stations must be physically protected to prohibit unauthorized access. Monitors must be placed so that contents on screens can be seen only by the analyst or annotator. DPC-AS-2: Analysts and annotators must have received relevant training in data protection and integrity issues. DPC-AS-3: The AS leader administers access requests to be made to the CDC. DPC-AS-4: A confidentiality agreement for any involved AS personnel must be in place. DPC-AS-5: Specified procedures for data extraction must be used. DPC-AS-6: The analyst must not extract or re-distribute data. DPC-AS-7: Project data must not be used for research areas not covered by the consent forms.

3.3.3 Data protection in DSF vs UDRIVE

Table 8: Data Protection in DSF vs UDRIVE

Topics to be addressed according to the DSF	UDRIVE Feedback
Data Classification • A series of guidelines in order to classify data and ensure the proper protection according to the level of sensitiveness. • Protection of personal data andconfidential data, and management of non-sensitive data.	• UDRIVE is treating recruitment data, and video and GPS as personal data that need protection. Confidential commercial data is CAN and related documents such as dbf files and also data originating from sensors from commercial actors such as Mobileye. • As video is present in the dataset, all data can be related back to the participants, so there is no non-sensitive data in the dataset. The reference to participants is de-identified by using a driver id.
Data Access Methods The DSF describes data access methods, e.g. public download, conditioned download, remote access, and on-site access.	• Remote desktops were set to connect the pan-European Analysis Sites to one Central Data Centre, hosted by SAFER. Such configuration has its limits since the shield that physical security can give for personal and confidential data is removed. On the other hand, data is readily available and analysis is made on a common dataset. The used Network protocol is PCoIP, making it possible to stream

	video over long distances.
Certfication	• A series of requirements need to be fulfilled for each organisation handling data. • An application describing the implementation by the partner and details on the compliance is sent through the DPC certification process
Data transfer	• Any data transferred between UDRIVE partners must be encrypted (the main bulk of data is transferred on fully encrypted disks). • The remote desktop communication between Analysis Sites and Central Data Centre is encrypted.
Data extraction process	• The data can only be extracted by the partner managing the data, being either Central Data Centre or Partner Data Centre. The extraction process is specified in the DPC. • Extracted data must be fully anonymized and every data extraction is documented by the Central Data Centre.

3.4 Training on data protection

The DSF recommends that training is given to all people handling the actual data. The guidelines says that a description of the data with special focus on personal data and IPR, external data handling requirements, explanation of the consent form content, explanation of data-handling procedures and the information about publication rights should be included in the training. It also states that the training needs to be documented.

UDRIVE has inserted the requirement for training into the DPC, where both the CDC and the AS, relating to the DSF, need to develop training material adjusted to their specific set-up and provide training to all personnel handling the actual data. The DPC demands that this training is documented by the AS and a paper is signed by the people taking the course. Each organisation is also signing off towards the project that they are giving this training, through the implementation documentation that is signed by the analysis leader and a person in the certification organisation.

Lessons learnt:

It is important that this information is given before anyone is given access to the data and that it also could be good to repeat the training some time into the project, to keep the focus on how to treat the personal and the confidential data.

3.5 Support and research services

The DSF brings a large variety of support and research services. The suggested support services consists of information and data provision, supporting tools, assistance with dedicated research needs and data protection and analysis facilities. The function could also provide research services through research advice on methodology, research involvement/research support and complete research performances.

The project is already using remote data sharing as a mean for analysis and some of the proposed services are already available to the analysis centers, such as information and data provision, supporting tools such as analysis tools and all partners working remotely have set up analysis facilities according to similar requirements as set out in the DSF.

As eleven analysis sites already are connected to the CDC, an extensive ability to utilise the tools and perform research remotely is built up in the project. Therefore, when the project has ended and as the same set-up is intended to be used after the project, several partners has the capability to provide research services and also the support functions such as making analysis facilities available. They are also aware of the advantages and drawbacks of the data and can provide assistance in the usability of the data for specific research needs.

Lessons learnt:

The projects that are implementing data sharing on a common dataset already in the project can later easily do additional analysis and provide assistance to new re-users of the data. The whole set-up also get tested in the project, which is a huge advantage compared to the data centers setting up a data provision center from more or less from scratch after the project.

3.6 Financial models

The DSF brings up several things to consider, when aiming at providing the data after the project. The DSF discusses the data management costs, eight financial models and how the cost is distributed among the stakeholders for the different models. Part of the content can also be use as a checklist for a proposal phase, for instance to understand what costs are involved in storing and providing data within a project.

The project consortium promised in the description of work to maintain the data for re-use if sufficient funding was available. There were no discussions in the beginning of the project regarding the funding possibilities, as everyone was focused on solving the more project-oriented issues. In the middle of the project the discussions started on how to maintain the data and investigations started to figure out the possibilities for external funding. Funding for accessing data can be included in project proposals and this possibility was utilised and project got funding. Still, funding from a few projects is not enough to keep the dataset up and running in a sustainable way.

Therefore alternative funding possibilities are discussed, but no solution is reached yet. If no funding is provided, the data need to be stored in a cost-efficient way and after three get disgarded if no funding is found.

Lessons learnt:

The DSF should stress that it is good to bring up the issue of funding up the question of funding already in the GA, to get an understanding of the possibilities to receive funding for data sharing after the project. Large datasets can not be provided on the basis of projects providing funding for minor research activites, there need to be a basic funding to just make the data available. The project research related costs should though be provided by the project. The situation where datasets might need to be disgarded might also have implications for the ability to replicate the research done in UDRIVE.

3.7 Application procedures

The DSF is providing extensive lists of items for both the application procedure and the actual application form.

UDRIVE is still in the analysis phase and has not had any need for an application procedure for new projects. First, the issue of sustainable funding for data maintenance needs to be solved. When a solution has been found, UDRIVE will use the list from the DSF; the list is actually already inserted in the proposal for an organisation to handle the data after the UDRIVE project.

4 Conclusion

The Data Sharing Framework was developed in parallel with the UDRIVE project. For this reason, most of the suggested guidelines were applied in the project and some others were updated with the lessons learned from the project. The recommendations in the DSF definitely provided a reference point when shaping all of the data-related issues in UDRIVE. The DSF does not necessarily imply a single and strict structure on how to ensure proper data sharing for every possible project but it can be considered as a starting point and could evolve into different adaptations depending on specific needs. Based on lessons learnt in UDRIVE, we have also proposed improvements to the DSF.

Several challenges have emerged within UDRIVE when trying to ensure data sharing. A very time-consuming task has been getting the acceptance and clear requirements to share data from a specific country according to the specific legislation.

Another great challenge which continues to be time-consuming and that will have an important financial impact is ensuring the post-UDRIVE data & tool maintenance. It has been difficult to find a business model involving a monetary risk partners are willing to take based on the overall uncertainty of the availability of project that can eventually fund the costs. Finding a sustainable solution is crucial to ensure the longevity of the data.

Overall, the UDRIVE project can recommend the DSF as providing excellent guidance in what pre-requisites and procedures to implement in the project.

References

European Directive 95/46/EC Art. 2. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en, accessed on December 27 2016.

Gellerman, H., Svanberg, E., Kotiranta, R., 2016. UDRIVE Data Protection Concept. In: Proceedings of the ITS World Congress 2016, Melbourne, Australia.

Gellerman, H., Svanberg, E., Kotiranta, R., Heinig, I., Val, C., Koskinen, S.,Innamaa, S., Zlocki, A., Bakker, J., 2017: Report from the FOT-Net Data Sharing working group, In preparation.

List of abbreviations

AS	Analysis site
CAN	Controller Area Network
CDC	Central data centre
DAS	Data acquistition system
DPC	Data protection concept
FOT	Field operational test
GPS	Global position system
LDC	Local data centre
NDS	Naturalistic driving study
OEM	Original Equipment Manufacturer
OMT	Online monitoring tool
OS	Operation site
PDC	Partner data center
PTW	Powered two-wheelers
VAS	Vehicle adaptation centre
VDI	Virtual desktop infrastructure

List of Figures

List of Tables

ANNEX 3

Division of Transportation Engineering,
Dep. of Civil Engineering
Aalborg University
Thomas Manns Vej 23
9220 Aalborg Ø
www.civil.aau.dk

Aalborg, 31 January 2017

Aalborg application of the FOT-Net's Data Sharing Framework on the ITS Platform dataset

Background

Since the millennium, there has been a fast growth in the number of Field Operational Tests (FOT). Floating Car Data (FCD) has mainly been collected under real driving conditions by volunteer drivers and datasets have been used to answer the research questions in their respective original projects. Many FCD sets have only been analysed to a small extent. FOT FCD is expected to provide significant benefits for research societies as well as the society in a broader perspective, when fully analysed.
Data sharing allows enterprises, researchers and students to gain further knowledge from the re-used data sets. It is much cheaper than collecting new data, and it allows the re-user(s) to stand on the shoulders of the work other researchers/organisations have conducted. This is also one of the main reasons why the researchers behind the ITS Platform North Denmark, cf. below, have been in dialog with the FOT-Net Data consortium. In cooperation and with guidance from the experiences gained in this consortium, it was expected that the collected FCD could be shared, and that research and innovation cooperation can be the outcome of this data sharing.
Aalborg University therefore contributed to the further development of the FOT-Net Data Sharing Framework (DSF) through applying the framework to our own FOT FCD dataset to make it ready for data sharing and at the same time, provided the DSF document with comments.

Description of the dataset

Data were collected in 2012–2014 in the Danish research and innovation project 'ITS Platform North Denmark', also known as the ITS Platform (FOT-Net Data 2016). ITS Platform consisted of an On-Board Unit (OBU) installed in 425 cars and enabling the mobile network to communicate with a backend server and WLAN connection with the neighbourhood.
The platform had a back-end server, OBUs and communication infrastructure. An application thus consisted of a number of programs in the OBU/ back-end server, Floating Car Data (FCD) provision through back-end server, and web/smartphone-based applications as user interface. The OBU connected the ITS back-end server, the individual vehicle and the driver with each other. The OBU was a mobile unit, placed under the car's dashboard. The OBU was active, when the ignition was on.
The OBU was developed aiming at mobile ITS services. Dedicated services and applications could run on the OBU through a set of open interfaces. It was easy to install and was updated via GPRS frequently. The back-end server was the backbone of the project. It collected, processed and transmitted FCD, parking data, and traffic information, and was the centre of the infrastructure, and communicated with all OBUs via a secured, closed network. The back-end server's main task was to be the platform for all applications in the project. It thus ensured that an application could communicate with the OBUs, receive FCD from the cars, and return control messages.
Partners were Aalborg University, Gatehouse and the at the time start-up company Inntrasys. The project was co-supported by EU regional Funds and the North Jutland development funds and had a total budget of 34.7 million DKK. It ran from April 2010 to December 2013, while data collection continued one more year.
Each OBU collected FCD with 1 Hz ID, position, map-matched position, direction, speed and a number of other attributes, which are mostly related to position reliability. In addition, acceleration data was collected with 10 Hz in three dimensions. Data consists of about $1.4*10^9$ positions. The recorded distance driven is about 15 million km in total. The number of accelerations recorded approximates $42*10^9$.

Head of Division, Associate professor, PhD – Niels Agerholm – na@civil.aau.dk - +4561780455

Extensive review and use of the Data Sharing Framework

As part of the cooperation with the FOT-Net Data consortium, an extensive review of the Data Sharing Framework (DSF) was carried out. It had a two-fold purpose: 1) To review the DSF in order to have other "expert eyes" looking into it, and 2) to maximise the learning from this document in order to raise the quality of the planned data sharing with the ITS Platform FCD. The main outcome from the DSF review with regard to the shared FCD is summed up in table 1. Detailed comments were inserted directly into the DSF document and provided separately.

Table 1. Summary of the outcome of DSF review with regard to the shared FCD.

Headline from DSF	Lesson learnt	Any proposed amendments to the current DSF
General data sharing project documents	It is smart to take all possible opportunities into account when the project documents are written – especially with regard to participant agreements – as it is almost impossible to reach all participants later.	None
Data and metadata descriptions	The content of the data from ITS Platform and metadata description is made and follows the proposed content in the DSF with a high-level description of a data collection; descriptive metadata, which describes each component of the dataset. Also, it consists of structural metadata, which describes how the data is being organized; and administrative metadata, which sets the conditions for how the data can be accessed and how this is being implemented. Everything is presented in a downloadable short scientific report.	None – the proposed design in the DSF has been a good scenario to follow.
Data protection recommendations	The raw data that needs protection is password and firewall protected. AAU works as Data Centre and actions to ensure data protection has been taken. Also, the database access is prepared for any unexpected additional requirements from the upcoming EU regulation regarding privacy. As the possibilities are small that the data centre can ensure sufficient data protection and responsible data handling at the Analysis Sites, it was decided to make the data sufficiently anonymised to meet the required protection of privacy even regarding a full publicly available data set.	None – The data protection recommendations have guided the work very well, as it has clarified the frame of possible useful solutions.
Training on data protection related to personal data and IPR	Written description of the data with special focus on personal data and Intellectual Property Rights (IPR) has been provided to all internal users of the FCD.	None

Head of Division, Associate professor, PhD – Niels Agerholm – na@civil.aau.dk - +4561780455

	Originally, vague description on data protection was made. As part of the cooperation with FOT-Net Data more effort was made: Meetings were held to gain information about the national and other applicable laws, regulations, and rules, while little was written down in this regard – an absence which later might cause some extra administration (in case of replacement of key persons etc.). Here the DSF could have been followed more carefully, but the task was anchored on a person with the main responsibility in other fields of research.	
Support and research services	It has to be simple and easy – else it will limit the data sharing, as persons who has to support the research services often are busy with other businesses, and difficult procedures can extent the time of sufficient service very much.	It might be good to highlight the inertia, that will often be associated with such services – especially when it is provided by an expert, who doesn't have it as his/her central interest. In some cases, the support etc. is transferred to a new research assistants or a general research support organisation. Often such divisions have a very high workload and then it might be difficult for an external to push the organisation to deliver the promised support.
Financial models	As it is yet unclear and rather uncertain, if it is possible to make re-users pay for data access, it is wise to place the data management/backup etc. in bodies in an organisation, where they see it as a basic task (e.g. as research support in a central IT unit so the Digital preservation is a part of organisations' core activity). As to keep the administrative burden the lowest possible in order to find a viable financial model it has been chosen to anonymize the FCD dataset sufficiently and to make the setup self-explanatory, to reduce future time use for data sharing.	It is suggested to add the value of citations as paying part of or the total cost for the operation of providing a dataset for sharing.
Application Procedure	In order to make ITS Platform North Denmark data available for research and education without too much effort, a simple approach to access has been prepared: 1. Enter this web page: http://FCD-share.civil.aau.dk 2. Fill in the form with contact info and description of purpose 3. Read and accept the principles for access 4. Download data	None – it is stated clear in the DSF, that high time use can be expected.

Head of Division, Associate professor, PhD – Niels Agerholm – na@civil.aau.dk - +4561780455

Development work

The data and metadata description for the pilot dataset are available on the homepage, http://fcd-share.civil.aau.dk/, where data can be accessed.

The procedure for development of the description was 1) Dialog with FOT-Net Data experts, 2) Scrutinising the DSF and extract what was suitable for the particular data set, 3) A running continuous adaptation to the development of the final data set meant for public access. A final version of the metadata is entered into the FOT-Net's Data Catalogue.
A training manual for data re-users has been developed. It has been decided to make it simple also c.f. the financial model decided. The basic input to re-users are available on the homepage and the more thorough information is in the scientific report available from the homepage. It is expected that the training manual will be treated in an iteration and updated concurrently with an increased number of re-users become acquainted with the available data set. The development of the accessible FCD and its related infrastructure has its basic in the original project, but central tasks carried out as part of the cooperation with FOT-Net Data are described here.

The development and description of the Financial model behind the data access was time consuming. Initially dialog with internal IT experts on the basis of the recommendations in the DSF was conducted. After many discussions and concerns on having an operational set-up for data sharing, counselling and payment system, it was decided to make a system, where the payment for use of the published data is based on citations. It was partly because citations is expected to have an increasing value in the Danish higher educations and partly because it was uncertain if the data in practice can generate sufficient income to cover the costs of such a set-up.

The design of the user interface has been developed in close cooperation with the internal IT unit and based on the proposed recommendation from the DSF suited to the actual data set and requirements.

The development of the anonymisation procedure was performed in the following steps:

1. Removal of all socio-economic and questionnaire data. Only the FCD are the objects for the anonymization procedure.
2. Simplification of data: Only attributes of relevance were kept in this step. It included removal of time, date, and ID, the latter which would have allowed for connection of drivers to different trips.
3. In order to be able to analyse single time-space movements of individual trips, each trip has its time substituted with an integer, where a trip is always initiated at the time 1 (sec), and sequential. Hence, the trip length, its time use and the connection within the single positions recorded can be investigated without violating the time/space privacy challenge.
4. Each single trip is truncated near its initiation and its finalisation after these principles:
    a. All initiation and finalisation of each trip are registered.
    b. A circular space with a radius of 500 m is located around the initiation and finalisation positions with a random distance from this point to the centre of the circle of 0–500 m.
5. By these procedures for anonymisation it should be ensured that reverse engineering cannot be carried out.

The test procedure for anonymization is made, but might be subject to amendments, and consists of the following steps so far:

1. Check for any unforeseen deviation through data extractions and visual inspection
2. Manual random check of driving in remote areas – if there is a specific location that cannot be anonymised
3. Careful consideration if data are too anonymised, as Lu et al. (2014) highlighted, the fundamental challenges regarding usability of big data are "If data are not authentic, new mined knowledge will be unconvincing; while if privacy is not well addressed, people may be reluctant to share their data."

Head of Division, Associate professor, PhD – Niels Agerholm – na@civil.aau.dk - +4561780455

Anonymisation of the data set is ongoing. Although the first part of data was anonymised satisfactorily, a larger data sample is under investigation in January 2017, to be sure that the published data is sufficiently anonymised and at the same time as useful as possible.

Conclusions

AAU has opened up its FCD for the public and ensured that the data is sufficiently anonymised. The main source of inspiration and guidance, especially with regard to data protection is from the Data Sharing Framework developed by FOT-Net Data. Also, the nature of the data, the local conditions and resources, but in particularly the valuable contribution from the FOT-Net Data society with recommendations and especially many good scientific discussions, have shaped the quality of the publicly available FCD.

Reference

FOT-Net Data 2016. ”ITS Platform - FOT-Net WIKI”. http://wiki.fot-net.eu/index.php/ITS_Platform. Accessed 20/01 2017.

Head of Division, Associate professor, PhD – Niels Agerholm – na@civil.aau.dk - +4561780455

Test Case – Trafisafe Annex 4

In the beginning of WP4.3, VTT and the University of Leeds carried out a test case for sharing a dataset. The main purpose of the test was to clarify required agreement details for accessing a dataset with the help of a real re-use case and two lawyers. The case also served as an example for assessing the level of additional efforts required in documenting a dataset, when re-users were not familiar with the study that collected the data.

The dataset was from a Finnish FOT called Trafisafe[1], where young drivers with their parents (the whole family) received driving style feedback for about a year. The feedback consisted of web reports supported with e-mail reminders to view the reports, and real-time reporting service on another website optimised for mobile devices.

Novice drivers, especially young men, are over-represented in road accident statistics. They are also a group whose risks are very difficult to reduce. The received feedback was considered to speed up the learning process of becoming a reliable driver. The feedback for the whole family was supposed to initiate family discussions on driving styles and to start a thinking process (e.g. “why do I actually drive so fast on our home street”). In total, 200 drivers took part in the tests, 75 of them being novice drivers who had received a driving license only a few months back. Trafisafe partners were the Finnish Transport Safety Agency, EC-Tools and VTT.

The dataset consists of GPS, tri-axial acceleration data, OBD-based fuel consumption and engine RPM values, and questionnaire responses. The whole dataset but questionnaires was shared. The shared dataset does not contain contact details as they had been stored separately. After the original study, direct links between names and data had been deleted.

Agreements

Trafisafe partners were interested in re-use possibilities already from the start of the project, as further analyses were seen to benefit the original study. Therefore, Trafisafe participant agreement included clauses enabling wide re-use of data.

Data description

Mainly the existing data description documents had to be translated into English and further comments to be added to avoid misunderstandings. As no test site description document was available, two publications were made to provide an overview. As the test design was naturalistic, it didn’t involve e.g. lengthy measurement plans.

The shared documents aimed to provide similar information as what had been used in recent FOTs (DRIVE C2X and TeleFOT) between partners, to introduce new people to a dataset: a test site description and a document about dataset details. The dataset was otherwise relatively easy to describe, since the data types are common in FOTs and the used database management system (PostgreSQL) is publicly available.

[1] Trafisafe – Feedback for novice drivers. Tarkiainen, Mikko; Peltola, Harri; Koskinen, Sami; Schirokoff, A. 10th ITS European Congress, 16–19 June 2014, Helsinki, Finland (2014), 7 p.

Data protection

The dataset was encrypted before transferring it over FTP, after which it was deleted from VTT's FTP server.

According to a basic non-disclosure agreement (NDA) agreement, it's up to the recipient to take reasonable measures to protect confidential information. Further specifications were considered, but only a clause about destruction of the data, after a specified date, was included.

Since the dataset can fit well into a single laptop, access was given to named persons only and as the dataset was considered reasonably anonymous for research purposes, only basic data protection measures were taken.

Training

A PowerPoint presentation related to installing a relational database management system was provided, with further instructions on how to import the shared dataset. The presentation also gave examples on database queries that had been used in previous analyses.

Support and research services

The provision of the training material could be considered a support service, although in future re-use cases, it would already be part of the general dataset documentation. When comparing to DSF's classification of support services, VTT provided 1) information and data as well as 2) supporting tools, but no further services were necessary.

One extra publication was provided about earlier work on the same topic, as well as a small software source code example for processing acceleration data.

The re-users also requested further information on used sensors, which proved somewhat difficult to provide. During summertime, getting further specifications from a product takes e-mailing efforts.

Financing

In this case, the financing for data sharing came from the FOT-Net Data project. However, sharing simple datasets where both parties see benefit might be possible without specific funding (i.e. both organizations cover the needed resources for their behalf). It's likely that lawyer services will be needed.

As the original dataset is small and stored with several others of VTT's datasets on a dedicated server, no direct financing is required for its upkeep. Nevertheless, backups and occasional server maintenance are necessary as long as the server setup is in use.

Application procedures

Negotiations on an NDA between VTT and University of Leeds raised up interesting discussions on

– Liability clauses: what could be the level of damages (financial and reputation), if parts of the dataset would leak out? The dataset contained no direct identification (e.g. video, names), but had full GPS data (potential for indirect identification). A sum for total liabilities was eventually agreed.

– Common confidential data protection clauses were included in the contract to ensure for instance that those having access to the confidential information receive advise and that reasonable measures are taken in the handling of the dataset. Further, according to good scientific practises, evaluation reports should not show full GPS tracks of single persons.
– A PhD candidate made a request that VTT would oversee allocation of the main research questions, when/if data is shared for other re-users. This was to avoid a potential issue with two very similar research papers being written at the exact same time. The request was found reasonable in this case, although it would not fit completely open datasets that are posted on the internet. A date was agreed, after which VTT could share the data for exactly the same research questions.

The Trafisafe dataset is posted on the FOT-Net's Data Catalogue, with contact information. There is no formal application procedure but sharing is handled case by case.

Conclusions

The efforts to share the dataset were mostly related to discussions and numerous e-mails from both sides, involving two lawyers. In comparison, translation of key documentation took only a day to complete and required support for re-users maybe two days.

University of Leeds found the dataset to be easy to use and has planned further use for it in student work and research projects.

ANNEX 5

FOT-NET AND THE RDE – LEARNING FROM EACH OTHER

January 20, 2017

Review comments on the FOT-Net's Data Sharing Framework

on behalf of the ITS Programs Connected Data Systems Program, U.S. DOT

Table of Contents

1. INTRODUCTION AND BACKGROUND

Numerous countries on a number of continents have hosted or are currently hosting Field Operational Tests (FOTs) and Naturalistic Driving Studies (NDS), as well as pilots, demonstrations, and deployments. There is general recognition that by sharing analytical methods, data, tools, and results, those engaged in FOTs, NDS and other types of field studies can benefit from lessons learned, templates for gathering information, technical advice, resources that save time and money, and the ability to potentially replicate, modify, extend or refute research findings. To this end, the European Union has funded the Field Operational Tests Networking and Methodology Promotion, known as FOT-Net, for nine-years of support actions, ending in December 2016. FOT-Net goals will continue in the new project Coordination of Automated Road Transport Deployment for Europe (CARTRE), whose network activities have a key focus on vehicle automation and which will pursue FOT-Net's interest in developing a repository somewhat like the Research Data Exchange (RDE), described below.

One of FOT-Net's concerns is establishing a defensible research methodology for FOTs and NDS that is based on the scientific method and that focuses on ITS functions and systems. This methodology is broadly applicable, can be applied or tailored to different sites, and can provide a variety of support ranging from reusable data to tools for preparing and analyzing data.

1.1 FOT-NET DOCUMENTS

FOT-Net has prepared and updated a guidance document, known as the FESTA Handbook, to assist FOTs and NDS to organize and conduct their research. The handbook was originally developed in the European project Field opErational teSt supporT Action (FESTA). The fundamentals of the methodology are based upon a V-shaped diagram that serves as the organizing scheme for the handbook. In the preparation phases for conducting a FOT or NDS (signified by the left and downward portion of the V) are the following steps: identification and description of the functions that are to be addressed, examination of use cases, determination of research questions and hypotheses, identification of performance indicators (an outgrowth of the study design), and looking at measures and their sensors. These steps point to data acquisition and then the analysis. Analysis (signified by the right and upward portion of the V) is supported by a database with measures/performance indicators, involves data analysis, answers research questions and tests hypotheses, undertakes impact assessment, and conducts socio-economic benefit cost analysis. Consideration of ethical and legal issues is a central issue of FESTA and informs many steps. Indeed, the concern in the FESTA methodology and other forms of research support offered by FOT-Net that relate to human experimentation and research appear to have its roots in the original Helsinki Declaration (FESTA, p.14)

FOT-Net also prepared a *Draft Data Sharing Framework.* Two chapters of the Framework are presented as separate files that have been circulated for review and are intended for integration within the framework. One such chapter is entitled *Draft Data and Meta Data Descriptions* and the other is

entitled, *Draft Data Protection Recommendations.* The *Draft Data Sharing Framework* is offered as a set of recommendations, but portions of the framework, especially in the data protection chapter, are put forth as requirements, and may result from an impulse to prepare regulatory language that could be adopted by different countries, especially in the European Union (FOT-NET DATA, January 16, 2015)

FOT-Net has also developed three interrelated online catalogues: the first lists and describes FOTs and NDS, the second concerns data, and the third pertains to tools.

1.2 DATA COLLECTION IN THE US

FOT-Net can be compared with similar resources and forms of support in the United States to conduct research involving FOTs and early Connected Vehicle (CV) deployments as well as NDS. As the US nears adoption of Dedicated Short Range Communication (DSRC) for inclusion in light and heavy vehicles, most likely required by regulation, the US DOT has funded a number of pilots, demonstrations or early deployments of connected vehicles that use DSRC, including those in Ann Arbor and Detroit, Michigan; Manhattan, New York; Tampa, Florida; and the I-80 corridor in southern Wyoming. These deployments involve vehicle-to-vehicle (V2V) communication of a basic safety message every tenth of a second to allow calculation of trajectories and to determine if a crash is likely and initiate crash avoidance. For vehicle-to-infrastructure communication (V2I), message sets may include the basic safety message, signal-phase-and-timing (SPAT) data at signalized intersections, intersection layouts represented by MAP data, and Traveler Information Messages (TIMs). A third class of connected transport is referred to as V2X, and is exemplified by the transmission of messages between vehicles and nomadic devices of pedestrians. For the Tampa and Wyoming pilot deployments, an Operational Data Environment (ODE) will be collecting all the data which will be transmitted to a Traffic Operations Center and may also be transmitted to the RDE. The RDE, which allows world wide access and downloading of CV data over the internet, contains many large data sets from various pilots, demonstrations and deployments. The RDE is described more fully below.

Guidance for the connected vehicle pilot deployments at (http://www.its.dot.gov/pilots/)– which could be thought of as being somewhat similar in spirit to the FESTA Handbook -- was prepared by the ITS Joint Program Office (ITS JPO) of the US Department of Transportation. This guidance, in the form of numerous documents, is characterized by a systems engineering approach to deployment, performance based management and evaluation, protecting personal information and commercial Intellectual Property Rights (IPR), partnership coordination through various agreements, achievement of long term financial sustainability of the pilot deployments, and performing outreach that will help promote national rollout of the connected vehicle technology.

In contrast to FOT-Net, the CV safety and pilot deployments are focused more on technology deployment than answering research questions. However, the Manhattan, Tampa, and Wyoming deployments require that the participants develop performance measures and targets to track the deployments and pay attention to important considerations of experimental design, such as avoiding confounding variables in the analysis. In addition, there will be an independent evaluation for each pilot that will require close attention to the principles of experimental design. The self-performance evaluation and the independent evaluation for the three pilot CV deployments have many concerns in common with the research and experimental thrust of FOT-Net.

1.3 RESEARCH DATA EXCHANGE

Other noteworthy background concerns the Research Data Exchange (RDE), mentioned earlier. As a key piece of the USDOT's Connected Data Systems Program, the RDE supports research, analysis, application development and testing that are useful for creating operational systems concerning multimodal transport and frequently involving connected vehicles and mobile communications. The RDE is a large data repository accessed through a web site (www.its-rde.net/) and may be linked to external federated data. Data in the RDE has a hierarchical organization consisting of data environments containing data sets composed of data files. The RDE contains archived and real time data, probe data from field tests, and research project data including simulations. The RDE can be used for single mode and multimodal research, development, testing and demonstration of safety, mobility, environmental, and weather applications. The RDE captures data from connected vehicles, mobile devices and the infrastructure. RDE data has undergone data cleansing and quality checks, and has been thoroughly documented.

The RDE also reveals implications for standards, intellectual property rights (IPR) data ownership, and privacy. The RDE data sets do not include any private or sensitive data, and the data is in the public domain because the US DOT owns the distribution rights and data providers have signed agreements. Among the features of the RDE are advanced search, multi-file download, Frequently Asked Questions (FAQs), external links, metadata, and a map-based interface for viewing data environments.

New data sets are added to the RDE from USDOT projects and outside sources as they become available. Researchers send submissions, upon invitation, to the RDE Data Administrator and if the data is judged to have sufficient value and quality it is uploaded. The RDE is fairly heavily used. By the first quarter of 2016 there were over 1400 registered users of the RDE (Booz Allen Hamilton, April 12, 2016).

A few examples of data environments on the RDE are:

- Basic Safety Messages (BSM) – Orlando. BSM data collected every 0.1 second from transit vehicles at the 2011 World Congress Demonstration in Orlando FL
- Connected Vehicles and Roadside Data – Two months of connected vehicle, roadside equipment, and contextual weather data from the Safety Pilot Model Deployment in Ann Arbor, MI, as well as a one-day sample.
- Integrated Mobile Observation (IMO) data --- real time and archived data from Minnesota maintenance vehicles containing location data, road weather observations from vehicle-mounted sensors and engine data directly from the vehicles' Controller Area Network (CAN) bus

2. PURPOSE OF THIS DOCUMENT

There are multiple reasons for preparing this document. Each reason could have been addressed in a separate report. Instead the reasons have been addressed in an integrated fashion here because of the parallel and overlapping nature of the subject matter that invites comparing and contrasting the FOT-Net program and similar field studies in the US. The purposes of this document are as follows:

- First, the US DOT connected vehicle deployment program and related activities can directly or indirectly use the reports, catalogues, data, tools and lessons learned from FOT-Net. Of considerable interest is protecting PII and IPR and making good use of material that could enhance the RDE as it evolves, for example by providing improved guidance.

- Second, there is an opportunity to review and provide comments on the FOT-Net materials and resources. Participants in FOT-Net are receptive to feedback regarding their products. Also, by describing some of the more noteworthy things that the US has done, particularly regarding the CV safety and pilot deployments and the RDE, FOT-Net staff can reflect on what might be valuable for their own activities.

- To some extent this write-up concentrates on the FOT-Net data protection document and the umbrella Draft *Data Sharing Framework* because of the common concern regarding PII and intellectual property rights.

- There is strong interest within the ITS JPO regarding data collection. Accordingly, much of the discussion regarding data sharing emphasizes data collection.

- Technological change is opening new windows for management and exploration of FOTs, NDS, demonstrations, deployments and so on. Much of this technological change is related to advances concerning the handling of huge amounts of data which manifests itself differently in different situations. Here some of the implications of Big Data for the next generation of the RDE and connected vehicle deployments are discussed.

The goal of this study, therefore, is to exploit the information and resources of FOT-Net and related ones in the United States for their mutual benefit and to strengthen their respective programs. As stated in the background, the emphasis of each is different. FOT-Net is more concerned with research and the United States is more focused on deployment, although both research and deployment are essential concerns of each. This difference in emphasis is illuminating, and this document attempts to shine a light on important differences that not only may be of interest but compelling to adopt as a whole or in part.

3. THOUGHTS FOR ENHANCING FOT-NET'S DRAFT DATA SHARING FRAMEWORK

FOT-Net FESTA documentation does an admirable job in providing guidance on data collection. One of the main reasons is that FOT-Net maintains a consistent focus on the importance of collecting data from FOTs and NDS. The focus is on performance indicators that pose underlying data needs for analytic inputs to explore outcomes that can be used to accept or reject hypotheses and answer important research questions.

FOT and NDS research, ranging from cooperative systems to driver behavioral studies of crash avoidance, goes through a number of basic steps that researchers and practitioners are likely to appreciate fully:

- Select the functions (applications) to be tested
- Define use cases to test the functions
- Identify research questions that correspond to the use cases
- Formulate hypotheses and link them to performance indicators or measures that have underlying data collection requirements.

The *Draft Data Sharing Framework* is a useful document, but it does not have a focus as direct as FESTA. In the section, "Why data sharing and re-use of data," the main justifications given are:

- To obtain additional funding for further analysis – the highest motivating factor.
- To perform meta-analysis across FOT and NDS sites -- offers the prospect of better conclusions than can be obtained from a single data set
- To strengthen International collaboration and data sharing and the flow of ideas and knowledge
- To foster research collaboration that creates trust and promotes the willingness to share data
- To achieve broader availability of data to facilitate doctoral work and other research.

3.1 ALIGNMENT OF FOCUS ON RESEARCH

FOT-Net may consider aligning the Data sharing framework to focus more on the research aspects of data collection, as does the FESTA Handbook. If this alignment were to happen, there would be a number of benefits:

- The focus of the FESTA Handbook regarding data collection would cascade more directly to the *Draft Data Sharing Framework.* Even though the topic of the data sharing document is somewhat distinct and elaborates on material in the FESTA Handbook, material could reinforce what the handbook seeks to accomplish.

- Potential users of a data set would see whether the goals of a particular data collection effort could be more fully achieved. The organizing scheme of the FESTA Handbook is based on a business process represented by a series of activities displayed in the shape of a "V." It might be useful if the beginning of the *Draft Data Sharing Framework* explicitly referenced this same "V" diagram and called attention to the steps that might be of significant interest to a researcher or others who may wish to re-use the data from a particular FOT or NDS. In other words, what additional research questions might be answered through research based on shared data, and what additional hypothesis tests might be conducted to provide added insight at the target site or a combination of sites.

- Lessons learned regarding an experimental design for a FOT or NDS may be useful for teasing out additional results. The crux of a FOT and to some extent a NDS is the experimental design. Researchers may be able to develop new research results, extend existing ones, or fully augment them if they had some in-depth insight into the experimental design used in an FOT or NDS. For example, the experimental design may have resulted in collecting data that captures interaction effects, but these were never analyzed, so the data represent an opportunity to obtain new research results. Analysis for an FOT may have omitted certain explanatory or confounding variables. By estimating new equations, performing a richer analysis of variance, and so on, it might be possible to extend the results. Finally, it may be possible to augment the results by performing sensitivity analysis, conducting Monte Carlo Simulation or doing some other type of analysis to determine how certain outcomes respond to the variability in key inputs.

- The *Draft Data Sharing Framework* would benefit from more explicit attention to potentially useful taxonomies of performance indicators or measures. The FESTA Handbook brings performance indicators into the foreground as hypotheses are formulated. The definition of a hypothesis used in FESTA is the following:

> "a specific statement linking a cause to an effect and based on a mechanism linking the two. It is applied to one or more functions and can be tested with statistical means by analyzing specific *performance indicators* in specific scenarios. A hypothesis is expected to predict the direction of the expected change." (italics added).

3.2 IMPACT AREAS

In the FESTA Handbook, impact areas, fertile for forming hypotheses, lead to performance indicators, which imply data needs. Approaches to identifying potential impacts are grouped into top-down consideration of impact areas and bottom-up specification of scenarios – use cases covering a specific situation. The first of two top-down areas of impact comes from an approach for identifying safety impacts:

- Direct effects of a system on the user and driving
- Indirect effects of a system on driver behavior
- Indirect effects of a system resulting in imitating behavior
- A change in the interaction between users and non-users
- A change in accident consequences, for example, effectiveness of traffic incident response
- Interaction with other systems.

A second top-down categorization of impact areas offered in the FESTA Handbook concerns Efficiency, Environment, Mobility, Safety and User Uptake. With the exception of User Uptake, these categories closely match the types of impact areas addressed in the solicitation for the Connected Vehicle Pilot Deployments in the United States. These impact areas are highly suggestive of different sets of performance indicators, and therefore potential types of data that may need to be collected.

The bottom-up approach to impact identification identified in the FESTA Handbook consists of developing hypotheses concerning specific scenarios derived from a combination of use cases and situations. Once hypotheses are identified, then one can explore what performance indicators to consider and the required data.

An alternative taxonomy of performance indicators widely used in the United States that might be considered for inclusion in the *Draft Data Sharing Framework* consists of inputs, outputs, outcomes and measures of value added that typically enter benefit-cost analysis. Results can be quantitative or qualitative, monetary or non-monetary. This categorization boils down to outcomes and value added and what produces them. These types of results have corresponding performance indicators that can be used for hypothesis testing. Inputs can be thought of the levers that change the results (including the delta represented by change in resources that contributes to the difference between the control or baseline conditions and the results).

3.3 FORMULATION OF HYPOTHESES

An important point in the FESTA Handbook that should be carried forward into the *Draft Data Sharing Framework,* is the large number of hypotheses that may be formulated, the corresponding large numbers of performance indicators (inputs and outcomes), and the large data collection demands. It is important to acknowledge there are limited financial and other resources that force choices upon researchers in deciding what data to collect and ultimately what can be shared.

4. LESSONS FROM THE FOT-Net DATA SHARING FRAMEWORK FOR THE UNITED STATES

In the United States the Research data Exchange is the primary repository for sharing data concerning FOTs, ITS demonstrations, and likely some Congestion Management Advanced Technology Deployments funded through the FAST Act as well as the Smart City Challenge being conducted in Columbus, Ohio. The RDE is being upgraded to a new platform to serve those who wish to mine its data more efficiently and effectively. It will continue to use open source software, draw on cloud storage, and use a flexible approach to performance metrics to track and gauge RDE usage.

The *Draft Data Sharing Framework* addresses many key concerns that the ITS Joint Program Office must address regarding data sharing supported by the evolving RDE. It is worthwhile calling attention to the strengths and insights of the FOT-Net's *Draft Data Sharing Framework* that may be relevant to the RDE. The RDE Guidance under preparation should examine whether to incorporate a number of sections that cover FOT-Net topics.

4.1 DATA SHARING AGREEMENTS

The Draft Data Sharing Framework begins by emphasizing the importance of forging the agreements that serve as the foundation for data sharing. These agreements pertain to both the partners involved in a FOT or NDS and to the experimental subjects, the participants. FOT-Net asserts that establishing these agreements at the outset cannot be emphasized strongly enough.

In thinking about the RDE and other data sharing repositories in the United States, consideration should be given to enabling users to download the agreements that pertain to a data environment, data set, or data file – to use the terminology of the RDE. Also, useful would be the ability to download checklists, key topics, or FAQs to consider regarding data sharing. Such resources can come from any useful source of guidance whether it be from FOT-Net or guidance written expressly for the RDE. The *Draft Data Sharing Agreement* speaks to the following:

- The Funding Agreement and the description of work – such an agreement may address who owns the data, third party access, who owns and can access to analytic tools, location where data will be stored during the project, data maintenance responsibility, access to data (including who makes the decisions), post-project funding, and legal and ethical constraints.

- The Consortium Agreement -- The *Draft Data Sharing Agreement* does not explicitly define the meaning of "Consortium," but presumably it refers to the partners in a FOT or NDS. It may encompass selected stakeholders because of their importance and could possibly include third

parties who would participate under certain contingencies, for example to perform special analysis of data. This subsection of the *Draft Data Sharing Agreement* includes a thoughtful table that lists data sharing topics that should be considered for inclusion in a consortium agreement. This list of topics helps focus one's attention when examining an actual agreement for an FOT or NDS. It is also valuable as guidance for the RDE or writing a section of the scope of work for a FOT or NDS concerning data sharing. A sampling of topics and corresponding questions from the table is presented here:

Topic	Related Questions
Ownership and access to data and data tools	Who owns the data? The analysis tools? Conditions of data use? Does a partner have intellectual property rights to the analytic tools? Are there constraints on use of personal data, especially video?
Storage and download of data	Will the data be stored centrally or distributed? May the partners download all or only part of the data? May third parties download the data?
Access methods	Remotely downloadable? Just on partner premises? What are data protection requirements?
Areas of use	For research and commercial purposes? Commercial uses (e.g. safety, mobility)?
Post-project re-use of data	Who maintains the data after the project is over? Who grants access to the data post-project? Are there legal or ethical constraints on re-use?
Post-project financing	How will storage and support services be funded after the project?

Source: FOT-NET DATA, Sept 30, 2015

- Participant agreements including consent forms – protection of the interests of participants in an FOT or NDS is of paramount concern because of legal and ethical concerns. Disclosure of Personally Identifiable Information is to be strictly avoided. In the United States, working with an Institutional Review Board (IRB), similar to an Ethical Review Board in the European Union, helps to ensure proper procedures are followed and. Standard procedure is to obtain written consent of participants if they will be the subject of human research or experimentation. The *Draft Data Sharing Framework* acknowledges that participants in an FOT or NDS are likely to be involved for periods ranging from a couple of weeks to a year, and that it is important to be clear on the use of data during and after the project. The *Draft Data Sharing Framework* points out that data may be shared with different countries. This has not been a major concern in the United States, even though data from the RDE can be downloaded via the Internet from anywhere in the world. An interesting question is whether there are differences among states concerning data protection just as there appears to be across countries in Europe. The answer is

that only three states have laws or significant court rulings regard PII: California, where the state Constitution declares privacy an inalienable right, Nevada and Massachusetts (Wikipedia, Personally identifiable information)

- External data provider agreements – the *Draft Data Sharing Framework* states that companies could provide sensor systems, map data, weather data or other sources to enhance the core data. It is suggested that non-disclosure agreements and contracts should be signed and that FOTs and NDS pay attention to issues that could affect future research such as what is regarded as confidential information, the feasibility of aggregating confidential data to allow data sharing, and accessibility and transferability to a third party.

4.2 DATA AND METADATA DESCRIPTIONS

All data environments submitted for posting on the RDE must provide metadata that follows the metadata guidelines established by the USDOT and available on the RDE website (https://www.its-rde.net/rdeabout/faq). The metadata guidelines are based on the Standard Practice for Metadata to Support Archived Data Management Systems (E 2468-05) written by the American Society for Testing and Materials (ASTM). Additional metadata may also be provided on the RDE if available.

The *Draft Data Sharing Framework* devotes considerable emphasis to how data and metadata should be described. This material is provided by reference in a chapter that is a separate document entitled, "Draft Framework for Data and Metadata Description." However, there is no common metadata standard among FOT-Net data since FOT-Net does not yet have data from FOTs and NDS that users can download from a common archive or repository. If someone is interested in the data from research done under the FOT-Net umbrella, it is necessary to go to the lead partner, or perhaps other participants in a specific FOT or NDS, to request their data and metadata.

In the future it may be worthwhile to compare and contrast the two documents discussing metadata and consider borrowing approaches for improving handling of metadata.

4.3 DATA PROTECTION RECOMMENDATIONS

The *Data Sharing Framework* purposely omitted the chapter concerning data protection recommendations, electing to first circulate it for comments, revise it and then re-integrate it into the framework. The chapter on data protection currently exists as a standalone document, and is entitled, "Data protection recommendations."

FOT-Net gives strong guidance regarding protection of personal data, commercial data, and Intellectual property rights. The motivation is several-fold. First are the 1964 Helsinki Directives and revisions that require consent from a participant in a human experiment and make clear that human rights take priority over scientific progress. Second, are European directives that define personal data and set out the conditions for using such data. Personal data is defined as data able to directly or indirectly identify an individual via an identification number or one or more physical, physiological, mental, economic, cultural or social identity factors. Member states may not process personal data unless the data subject provides explicit consent. However, the laws of a member state may prohibit the lifting of data protection of a subject. Third, law protects commercial data as confidential. On the one hand, it may be possible to share confidential commercial data among partners of a consortium. On the

other hand, it may not be possible because the data provider regards the commercial value of the data too high to share. (FOT-NET DATA, January 16, 2015, p.6-7).

Fourth, it is important to protect Intellectual Property Rights (IPR). While the FESTA Handbook addresses IPR, the data recommendations do not. The introduction to the *Draft Data Sharing Framework* might make clear that IPR is addressed elsewhere. Also, IPR needs to be defined more clearly, preferably in the FESTA Handbook. IPR includes copyrighted material, trademarks, and patents. At first blush, IPR do not appear to be relevant to data protection. However, data is increasingly recognized as consisting as structured, unstructured or some hybrid. Video data can include sensitive personal data (i.e. a face) that can be used to establish identity. But the video could also include information tantamount to GPS traces such as mileage markers and road signs with entrance and exit names.

The data protection document's main strength is the procedures for protecting data at Data Centers and Analysis Sites. This document, with the word "recommendations" actually in the title, sets out eight data protection requirements for Data Centers and ten for Analysis Sites. As mentioned earlier these requirements may be offered for consideration as national regulations. A Data Center is defined as a site where more than one person may download data from an FOT or NDS. This may be an overly stringent definition because characteristics of the site should vary with the number of people at a data center but this relationship is not spelled out. Analysis sites upload data from Data Centers.

The Data Protection Recommendations include a figure that attempts to show the data access pattern in four layers: the data layer, the analysis layer, the access layer, and the credential layer. Unfortunately, these layers are not defined. The relationship between the data layer, which presumably coincides with the Data Centers and the analysis layer, which corresponds to the Analysis sites, is unclear. This report would benefit a clearer explanation of the figure. Another way to provide more clarity would be to provide some diagrams defining the data management architecture and the relevant business process.

Here are the eight data protection requirements of the Data Center (DC), taken directly from the Data Protection document, along with selected supporting points. These requirements may be considered for adoption by the US connected data systems program.

4.3.1 Data Center Data Protection Requirements

DC-1. Data stored and processed at a DC must be protected from unauthorized access.

- Cover analysts and administrative staff
- Use group-based privileges instead of providing individual access
- Work stations need special privileges for uploading

DC-2. Data stored and handled at a DC must be protected from accidental deletion

- Must backup data, perhaps in different locations, for disaster prevention
- Consider keeping original data, such as hard drives in antistatic sleeves in a fire resistant safe, in a different building than the data center
- Have controlled procedures for deleting data

DC-3. The DC must document its data protection implementation

- Do so before allowing downloads to occur
- Fulfill legal requirements; be transparent
- Appoint a data supervisor responsible setting out, implementing and following data protection requirements

DC-4. Confidentiality agreements for any involved personnel must be in place

- The Data Center must require signed confidentiality agreements with involved personnel before they start handling FOT/NDS data

DC-5. Data protection must be ensured by the DC after the end of the project

- The same data protection level must be provided after the data collection project
- Anonymized data can reduce overall data protection requirements in a project's aftermath
- Secure data erasure – applies to all media (i.e. storage and backup systems, portable storage and even paper)

DC-6. Data sent between DC and Analysis Sites (AS) must be encrypted

- Data can be transmitted electronically or physically but may not be accessed during transfer

DC-7. Data downloads (for reuse) are regulated by the Project Agreement(s) and the informed consent of the driver

- Should address the possibility of downloading all or part of the data
- Should address downloading to those not partners

DC-8. Data extractions (i.e. the output from analysis) for specific purposes must be in accordance with the consent forms and project agreement and the extractions must be documented

- Agree to and implement a process for data extractions
- Determine if participant consent is required and if signed agreements needed, if applicable
- Assess if the inclusion of data requires the owner's approval

4.3.2 Analysis Sites Data Protection Requirements

The following are the requirements for Analysis Sites (AS) taken directly from the data protection document along with a few explanatory points. These requirements may be considered for adoption by the US connected data systems program.

AS-1. The AS organization must document its data protection implementation

- Appoint an AS Supervisor
- Compile (or prepare as necessary) documentation that satisfies AS requirements
- Data Centers and third parties should agree to the requirements

AS-2. The analysis work stations must be physically and logically protected

- Protect against remote or virtual unauthorized access to analysis work stations
- Provide physical protection against unauthorized access; possibly restrict analysts from bringing mobile phones or portable computers into the analysis rooms
- Prohibit access to network services and detect and automatically report unauthorized attempts to redistribute data

AS-3. Analysts must have received relevant training in data protection and integrity issues

- Require proof that each analyst has received mandatory training
- Web or tailored training on data protection and integrity issues is necessary if data includes personal or confidential data.

AS-4. A confidentiality agreement for any involved AS personnel must be in place.

- Determine the extent that any relevant employer confidentially statement can satisfy the confidentiality requirements for an analyst.

AS-5. The AS supervisor administers access requests and forwards those to the DC data access dispatcher

- Restrict the number of individuals that may request accounts for data to better track users, accounts and privileges

AS-6. Specified procedures for data extraction must be used

- The Data Centers and Analysis Centers should agree on a process for data extractions, which is defined as outcome and other types of non-raw data)
- Extraction requests should address intended use, description, data types, amount of data, and files or folders to extract

AS-7. The analyst must not extract or re-distribute data

- The analyst must not circumvent data extraction regulations or release AS data in any not prescribed (Note: this document appears to be elements of or be a full draft regulation on data protection for use of FOT-Net member countries or the EU as a whole)

AS-8. The project data must not be used for research areas not covered by the consent forms in the project

- Data may be used only for purposes stated in the consent form
- Care should be taken in identifying what research may take place during and after the project; otherwise approval of an Ethical Review Board may be required, which depends on whether national legislation adequately addresses the necessary protections

AS9. Visitors/guests to the AS should sign a confidentiality agreement

- Visitors must sign a confidentiality agreement to look at confidential data.

AS10. All post-project research must investigate the need for approval

- In early phase of a project establish the limitations on potential re-use of the data, i.e. post-project research
- May need approval from appropriate national authority or from Ethics Review Board

4.4 TRAINING ON DATA PROTECTION

Training is a key aspect of data protection. The *Draft Data Sharing* Framework discusses the nature of a training program adequate to protect personal data and IPR. An effective training program addresses these elements among others: personal data and IPR; applicable laws, rules and regulations; consent forms to inform FOT/NDS participants about data collection, purpose, handling, storage, access, and reuse; and how data is made anonymous or encrypted. Although on-line or video training is not sufficient for protection of data they can be very helpful. Examples can be found at the following links:

- Videos: Data Protection Act training video
- Case Studies: https://data protection.ie./docs/CASE-STUDIES-2013/1441.htm#CS6
- US NIH Online training: http://phrp.nihtraining.com/

4.5 SUPPORT AND RESEARCH SERVICES

The *Draft Data Sharing Framework* identifies and describes different levels of support to help researchers use the FOT and NDS data and meet their research objectives. Data managers are cautioned that raw data from all the projects should be readable in "raw" and clearly defined format from main source of data storage. These are the layers of support:

1. *Information and data provision.* Researchers need a way to become aware of what types of data is available. The FOT Catalogue and the Data Catalogue described below are good starting points. Being able to inspect metadata can provide substantial insight. Also, developing an understanding of what types of data transformation and handling is possible may be important, for example, data aggregation, extraction, and transfer.

2. *Supporting tools.* There are a wide variety of such tools and they are frequently indispensable to examining and analyzing data. Among the tools are those used for viewing and annotating data, scripts for extracting data sets from a database or licensed software, and frameworks for downloading, processing and returning data to a database. The Tools Catalogue described below distinguishes among "Tools for Preparing", "Tools for Using," and "Tools for Analyzing."

3. *Assistance on dedicated research needs.* Advice and assistance is offered on selecting, modifying/adjusting and obtaining methods for specific research.

4. *Data protection and analysis facilities.* Helps to protect PII data and IPR and provides recommendations on how to secure analysis sites.

The other key part of the support the *Draft Data Sharing Framework* discusses is research services. There are three levels:

1. *Research advice on methodology.* This type of advice focuses on how to answer the research questions, the hypotheses to be tested, application of the scientific method, and if re-use of data is productive in these regards.

2. *Research involvement/research support*. More active involvement including identification and selection of data for re-use, developing specific tools for data handling, evaluation, and analysis, and carrying out a portion of the analysis. The latter may consist of transforming research questions into hypotheses, using performance indicators to drive data selection and analysis, and conducting statistical analysis.

3. *Complete research performance.* The highest level of research support would be to conduct the research from beginning to end: identify and acquire the needed data, choose and apply the analysis tools, and finish and document the analysis.

4.6 FOT-NET CATALOGUES

A good portion of the support is provided in the form of three catalogs available on the Internet. These are (1) a catalog of field operational tests and naturalistic driving studies; (2) a catalog of data resources, and (3) a catalog of tools. The URL is: http://wiki.fot-net.eu/index.php/Welcome_to_the_wiki_of_Field_Operational_Tests .

FOT Catalogue. This catalogue provides information on close to several hundred FOTs and NDS. There is a brief description of each FOT/NDS, frequently including the study objective and key study questions. There may be a link to a summary of the study or a research report. For each FOT/NDS there is a pane that shows the type, the tested system/service, the participating countries, partners, the number of vehicles involved, period of performance, and point of contact. The "FOT Catalogue" may list data sets used and may show if there is a link to the catalog of data resources. Similarly, there may be a statement regarding whether any tools were used and possibly a link to the catalog of tools.

Data Catalogue. This catalogue has relatively few entries. Almost half consists of questionnaire data. The others have more diverse forms of data. A table introduces each entry indicating whether each type of data can be shared as raw or aggregated data. Other information includes the name of the FOT/NDS, the perspective or goal of the study, the countries involved, indicators (e.g., speed, weather conditions), data formats and possible standards, frequency of logging, quality checks, who the test subjects are and the test set up.

Tool Catalogue. A large number of tools are referenced in this catalogue. In many instances the tools are linked back to the FOT Catalogue and the specific field tests and studies in which they were used. The basic information available about each tool varies but typically includes the name of the tool, the project in which it was used, category, purpose, type of analysis, if it is open source, if it has a GUI, the type of operating system, inputs, outputs, hardware and software. This catalog organizes tools according to whether they are useful for planning, monitoring, control, data acquisition, testing, data transfer, data base, driver communication, field testing, post processing, risk assessment and simulation.

Because FOT-Net does not have a single repository or archive comparable to the RDE, the data catalogue is relatively weak. However, the FOT-Net suite of catalogues provides information and tools that can help provide different layers of support as described above that the RDE does not offer. The

interlinking among the catalogues of FOT-Net framework is a strength, for example between the FOT Catalogue, the Data Catalogue, and the Tools Catalogue. Those responsible for enhancements to the RDE should consider providing comparable information and functionality (e.g. linkage from data sets to tools, or links from a tool to all the RDE environments in which it is used).

4.7 FINANCIAL MODELS

The RDE and similar data stores should be thought of as a financial enterprise that needs to be sustainable over both the short and the long run. The *Draft Data Sharing Framework* expresses a related thought: "If data sharing is not economically feasible for data owners and potential data re-users, re-use of data does not take place and the benefits of data sharing won't be achieved." The data sharing framework recognizes that consideration of appropriate business models and implementation of the right one is essential for data sharing to be viable. The posture of the FOT-Net on this topic is not unlike the posture of the US DOT in regards to insisting that the Connected Vehicle Pilot Deployments be financially sustainable. The *Draft Data Sharing Framework* provides a thoughtful set of financial models that could potentially sustain a data sharing enterprise. Those managing the RDE may wish to consider these models, which are further explained and whose pros and cons are spelled out in the data sharing document, and are excerpted directly below:

A. Organizations' core activity – Digital preservation becomes a part of organizations' core activities
B. e-Infrastructure – Public funding is directed to data infrastructures serving multiple organizations and disciplines
C. Archiving included in project budget – Project budget allows for dataset finalization and archiving as commercial services
D. Project extension – a project is awarded a continuation phase to maintain its data
E. New project funding – new projects finance maintenance or revival of a dataset
F. Established network – a network of organizations with participation fees arranges data management jointly
G. Analysis services – an organization with several valuable datasets creates business using them, offering both data and related services
H. Data integrators – companies acquire and market FOT datasets along with transport and other related datasets

4.8 APPLICATION PROCEDURE

Those interested in a data set need to go through an application procedure to acquire the data. The procedure needs to be straightforward and understandable. The *Draft Data Sharing Framework* identifies the content of the application procedure and information to collect in a form, typically on-line. Clearly the content and form will be specific to a particular approach to data sharing, such as the RDE or FOT-Net. Because of its emphasis on the scientific method, suggested questions for the form include "what research questions are the data expected to answer?" "What are the expected results?" "Would you like the research facilities to do part or all of the research for you?"

The RDE does not require such information to be submitted before a data file can be downloaded, but it would like to know such information for evaluation purposes. The RDE may consider ways to encourage data downloaders to provide this type of information.

5. TECHNOLOGICAL CHANGE IN DATA COLLECTION, STORAGE, AND SHARING

Rapid technological change not only is affecting the objects of CV and NDS research and field tests in the European Union, United States and other countries but also it is affecting data collection, storage, dissemination and data analysis as well as the design and development of data systems that carry out these functions. The exponential growth of data in the transportation sector due to CV, NDS, demonstrations, and pilot deployments parallels the enormous growth in data in other fields and necessitates the perspective, platforms and tools of Big Data. Moreover, open systems is becoming an increasing imperative for both the public and private sectors, and data is more and more likely to be stored in servers in the Cloud.

5.1 SOFTWARE DESIGN AND SYSTEM ENGINEERING

Traditional databases no longer are adequate for storing the terabytes or petabytes of data collected in a typical CV or NDS field test or deployment. A large amount of both structured, semi-structured and unstructured data is collected and stored. Traditional object-oriented software design is giving way to systems architectures more able to address elements that can handle Big Data.

5.2 ARCHITECTURE

For example, in the United States the architecture for CV deployments may include an Operational Data Environment (ODE), which gathers V2V, V2I, and V2X messages and communicates them in real-time to one or more Traffic Operations Centers. The ODE may also transmit data to a repository with shared data sets such as the RDE or some similar archive that supports reuse of data, same as envisioned by FOT-Net. In the case of the CV pilot deployments in the US, the Security Credential Management System will be operating in the background to ensure that the messages are well-formed and not corrupted, security and key aspects of privacy are protected, and deployment is fortified against malicious behavior that can undermine a CV system and spread over a corridor, region, state or country.

The architecture needs to provide for a platform for the repository that can easily accept added software functionality of modules or plugins. The types of architectures being used for Big Data repositories can scale clusters involving a few to thousands of servers, has the ability to detect failures at the application layer, and has a great deal of resilience and backup support.

5.3 PLATFORMS

While there are proprietary platforms for Big Data management and analysis, most of the larger software and cloud vendors embrace the family of open source software from the Apache Software Foundation. These players include IBM, Oracle, Microsoft, and Amazon. There are two Apache-related software approaches that frequently work together, but have distinct characteristics and different operating speeds. The first consists of four fundamental elements of Hadoop:

1. Hadoop Common: Utilities that support the Hadoop modules.
2. Hadoop Distributed File System (HDFS™): High throughput access of application data
3. Hadoop YARN: Framework for scheduling jobs and managing clusters
4. Hadoop MapReduce: Parallel processing of large data sets.

The second major member of the Apache family is Spark, which supports big data analytics including near real-time processing. For example, it is possible to process the message sets communicated among connected vehicles in near real-time. The core Hadoop family is regarded as comparatively simple, tightly integrated, fault tolerant, and highly flexible, but Spark runs 100 times faster than MapReduce if one is pursuing various types of big data analytics in many environments (see Apache, Hadoop reference).

5.4 DATA TYPES AND ACQUISITION

Hadoop and its accompanying modules, as well as other data management platforms such as HDF5, can overcome the traditional problem of not being able to process, store and analyze structured and non-structured data. Hadoop can handle many different types of data and interact with different data management platforms including traditional SQL but also those that deal with different types of data including csv, documents, graphs, video, and maps. Data acquisition can also take many forms including sensors, DSRC, CCTV, smart cameras, and voice recordings. (Datamani, May 15, 2014).

5.5 ANALYTICS

There is a wide range of analytic and visualization tools that can be used to tease out inferences and extract valuable information ensconced in big data sets. The suite of tools numbers in at least the hundreds. Common methods draw upon the fields of statistics and econometrics, artificial intelligence (expert systems, fuzzy sets, genetic algorithms, neural networks), biometrics including facial and gait recognition, natural language processing, data visualization, noise processing, predictive models, and a rich variety of graph and related visualization techniques that help identify hierarchy, anomalies, edge patterns, communities and so on. The Hadoop ecosystem is a rich source of visualization capabilities, especially MapReduce and Spark. A large number of Hadoop software providers offer visualization.

5.6 OTHER TECHNOLOGICAL CHANGE

Technological advances are accelerating in the transportation field, particularly in their integration. This is visible with regards to connected vehicles, autonomous vehicles, security, energy management, and data processing.

Sending messages through DSRC transponders is the backbone of connected vehicles. These messages at the minimum include the Basic Safety Message, but at signalized intersections will typically include Signal Phase and Timing (SPAT) data as well as MAP data to represent intersection geometry.

New ways to address vulnerability to hacking based on the Public Key Infrastructure are manifest in the Security and Credential Management System that is part of the CV Pilot Deployments.

Autonomous vehicles use different sensors such as LIDAR, radar, sonar, cameras with pattern recognition and many sophisticated algorithms to provide assistance braking and staying in a lane.

Monitoring by internal sensors and hundreds of microprocessors is constantly occurring within cars and trucks and the data can be tapped through the CAN bus.

Augmented and virtual reality have entered the scene and their applications will multiply, first most likely in areas such as transportation and land use planning, but eventually in real time operations management.

The growing market share of hybrids and all-electric vehicles is evidence in improvements in battery technology and energy conversion. Indeed, rapid progress regarding hydrogen fuel cell technology is occurring.

Another gauge of scientific and engineering advance was the recent announcement of Google and the National Aeronautics and Space Administration of the first Quantum Computer. This computer appears to have been tailored to address optimization such as the Traveling Salesman Problem and is 100 million times faster than a single processor (PC World, December 19, 2015).

What do these technological changes mean for the RDE and efforts of the FOT-Net to enhance their data collection, storage, sharing and analysis capabilities? Some answers follow:

1. Traditional platforms cannot handle enormous data sets with a variety of complex data types. To handle unstructured Big Data solutions varying capabilities exist ranging from HDFS to offerings of the Apache Software Foundation, especially the core elements of Hadoop.
2. Big data analytic tools, as FOT-Net recognizes, are powerful for hindsight, insight and foresight. Many traditional and advanced analytic and visualization tools can backcast, expand current understanding, and offer greatly improved predictive capabilities.
3. Constantly improving architecture of servers -- for example cluster organization around nodes instead parallel computing -- machine-to-machine communication, hardware advances of traditional types of servers, and the first generation of quantum computing (even if it has narrow functionality) ensures an evolving capability to collect, store and extract value from Big Data mined from the transportation network.
4. Many of these technologies are disruptive and will turn certain parts of the transportation sector on its head, including those that USDOT and FOT-Net are accustomed to dealing with.

Something as simple as crowdsourcing and its cousin, hackathons could result in unanticipated improvements to Big Data platforms for the reuse and analysis of data from CV, NDS and related field tests.

6. CONCLUSIONS AND RECOMMENDATIONS

FOT-Net and sister programs of the ITS JPO have fundamentally similar goals. Their approaches to data sharing, data protection, experimental design and implementation of vehicle and infrastructure applications overlap to a great extent but differ in many ways. The practices identified in this review make clear that FOT-Net (as it is embraced by CARTRE) and the ITS JPO can benefit from continuing to learn about each other's policies, practices, and procedures regarding data sharing, protection and related topics. It is recommended that this comparison of practices periodically be updated and shared with both parties and the broader international community.

7. REFERENCES

1. Apache, Hadoop, http://hadoop.apache.org/index.html , (n.d.)
2. Booz Allen Hamilton, RDE Analytic Report, January 1, 2016-March 31, 2016, presentation dated April 12, 2016
3. Datanami, May 15, 2016, https://www.datanami.com/2014/05/15/three-key-data-types-hadoop-analytics/
4. FOT-Net, FESTA Handbook, Version 5, 28 February 2014
5. FOT-Net, Wiki, the free encyclopedia of Field Operational Tests, http://wiki.fot-net.eu/index.php/Main_Page (n.d.)
6. FOT-NET DATA, Draft Data protection recommendations, January 16, 2015
7. FOT-NET DATA, Draft Data Sharing Framework, September 30, 2015
8. FOT-NET DATA, Draft Framework for Data and Metadata Description, September 25, 2015

9. PC World, "NASA, Google reveal quantum computing leap that leaves traditional PCs in the dust," December 9, 2015, http://www.pcworld.com/article/3013214/hardware/nasa-google-reveal-quantum-computing-leap.html
10. Research Data Exchange, https://www.its-rde.net/ (n.d.)
11. Wikipedia, Personally identifiable information, https://en.wikipedia.org/wiki/Personally_identifiable_information(n.d.)

List of Acronyms

AS	Analysis Site
CAN	Controller Area Network
CV	Connected Vehicle
DC	Data Center
DSRC	Dedicate Short Range Communications
FAQ	Frequently Asked Questions
FOT	Field Operational Test
IMO	Integrated Mobile Observations
IPR	Intellectual Property Rights
IRB	Institutional Review Board
ITS JPO	Intelligent Transportation Systems Joint Program Office
NDS	Naturalistic Driving Study
ODE	Operational Data Environment
PII	Personally Identifiable Information
RDE	Research Data Exchange
SPaT	Signal phase and Timing
TIM	Traveler Information Message
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
V2X	Vehicle to Anything