6th Transport Research Arena April 18-21, 2016

Data sharing of transport research data

Helena Gellerman [a],*, Erik Svanberg [a], Yvonne Barnard [b]

[a]SAFER Vehicle and Traffic Safety Centre at Chalmers University of Technology, Gothenburg, Sweden
[b]University of Leeds, Leeds, UK

Abstract

With the rapid progress of the development of intelligent transport systems over the last 15 years, the need for testing them in the real world and collecting data about their impact became more and more important. We have seen a fast growth in the number of Field Operational Tests (FOT) and Naturalistic Driving Studies (NDS) performed worldwide. The need to better understand the benefits of safety systems and the factors behind the occurrence of incidents and accidents have been a main driving force and the data has therefore been collected through naturalistic driving by volunteer drivers. As the number of different datasets has increased and so also the awareness of the substantial effort and funding needed to run these FOT/NDS, the interest in data sharing has increased worldwide.

The availability of a common Data Sharing Framework (DSF) could highly facilitate a larger use of the collected FOT/NDS data. The FOT-Net Data project has developed such a framework, in collaboration with a variety of stakeholders from Europe, the US, Japan and other countries. The seven topics addressed by the DSF are (1) project agreements, (2) data and metadata descriptions, (3) data protection, (4) training, (5) support and research services, (6) financial models and (7) applications procedures. Many of the topics are general and can be used for other types of transport research data as well.

There remain challenges to make data sharing possible on a global scale. Some of these are: the project funding schemes, leading to multiple schemas of ownerships of data, and the legal settings in different countries. On a technical level, the documentation of datasets and of the metadata describing the test is not always sufficient. Furthermore, new projects need to be made aware of the importance of inserting the pre-requisites for data sharing into the different project agreements right from the start.

This paper describes the content of the DSF with its hands-on recommendations on how to prepare for and perform data sharing of transport research data. It also presents the status of a use case, implementing the DSF into the European project UDRIVE.

* Corresponding author. Tel.: +46-31-772-1095.
  E-mail address: helena.gellerman@chalmers.se

1. Introduction

1.1. Background

The intelligent transport systems have shown a rapid development over the last 15 years and are now progressing into automation. This development has been accompanied with the need for testing the systems in the real world and collecting data about their impact. We have seen a fast growth in the number of Field Operational Tests (FOT) and Naturalistic Driving Studies (NDS) performed worldwide, where data has mainly been collected through naturalistic driving by volunteer drivers. The need to better understand the benefits of safety systems and the factors behind the occurrence of incidents and accidents have been a main driving force. The availability of technology, such as affordable video capture solutions and decreasing data storage costs, has been a requirement and facilitator for the development of the FOT/NDS area.

The largest datasets have so far been collected in the US e.g. SHRP2 and the Safety Pilot project (SHRP2 2015; Safety Pilot 2015) and in Europe e.g. euroFOT and the on-going UDRIVE project (Selpi et al, 2011; Eenink et al, 2014). In Japan, large ITS datasets have been collected, e.g. ETC 2.0, and also Australia has several interesting datasets, e.g. the Australian NDS project (ETC 2.0 2015; Australian NDS 2015). Data collection is also on-going in China (FOTs in China 2015). These datasets provide a great resource for answering research questions and the interest in data re-use is increasing along with the awareness of the substantial effort and funding needed to do these FOT/NDS.

For many of the previous and on-going projects, the concept of data sharing has been an internal matter and the re-use of data after the project has not been in focus. Therefore, some projects lack the pre-requisites in the project agreements to make it possible to share data. As re-use of data have not been a major project concern, the data has been documented to a level where the project partners can use the data, but for a third party, the documentation level, especially regarding the metadata, is insufficient to understand and re-use the data. Thus, more support services when sharing the data is needed to promote good research results. New laws putting a stronger emphasis on the personal integrity have also enhanced the need for a larger extent of data protection together with training of all personnel dealing with the collected data. At the same time, through-out the community, the majority of the organizations having collected the data face a funding problem at the end of the project, as there are still few funding possibilities to receive a reasonable amount of resources to be able to maintain and provide the data in a consolidated way after the project.

There are also different views on the value of data sharing depending on if you are a data provider or a data user. Substantial efforts have been made to collect data and build up the data infrastructure and tools. It is therefore important to find mutually beneficial business models for both providers and users, for re-use of the data to become a reality.

1.2. Purpose of this paper

This paper presents the Data Sharing Framework (DSF), continuously evolving through capitalizing on the information and needs collected within the European support action FOT-Net Data (Gellerman et al, 2015). The project organizes international workshops, stakeholder meetings and topic-specific workshops, involving a variety of stakeholders from Europe, the US, Japan, Australia and China (Barnard, 2014). The seven topics: (1) project agreements, (2) data and metadata descriptions, (3) data protection, (4) training, (5) support and research services, (6) financial models and (7) applications procedures of the DSF have been further detailed and the enhanced recommendations and guidelines are presented (DSF 2015; Data and metadata description 2015; Data protection

2015). This paper also presents the current status of a use case, the EU project UDRIVE, with its hands-on implementing of the framework step by step in the project.

2. Data sharing framework

The availability of a common DSF has the potential to substantially contribute to a larger use of the collected FOT/NDS data. Such a framework should include data sharing pre-requisites that could be integrated into project agreements from the start, as well as procedures and templates to facilitate easier data sharing. Researchers wanting to re-use already collected datasets could then utilise a more or less standard application procedure, rely on already performed training that is widely accepted, and plan for the costs that using a specific dataset might cause the project.

In the following chapters, the suggested content of the DSF is described. On an overall level, the following seven areas, as shown in Fig. 1, need to be addressed by a DSF. The bottom two topics are addressing the more technical recommendations of the DSF, whereas the remaining five topics are more administrative, focusing on guidelines and content in procedures and templates.
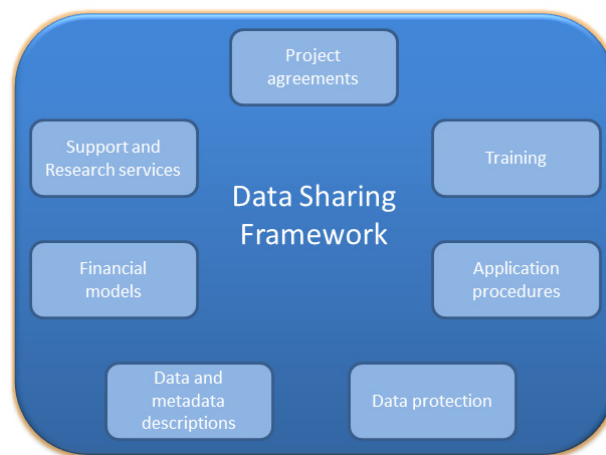

Fig. 1. Data sharing framework.

2.1. Project agreements

The initial process of setting up a project is crucial to the possibilities to share data during and after the project. The main documents to focus on are: 1) the grant agreement, if the project has external funding, including the description of the work; 2) the consortium agreement among the project partners; 3) the participant agreement; and 4) agreements with data providers to the project.

In the grant and consortium agreement, it is important to be aware of the topics and issues to be discussed in relation to data sharing and re-use of data and to focus them already during the project application and a possible negotiation phase. It is especially important to 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 topics that should be addressed are: 1) the ownership and access to data and data tools; 2) distribution of data; 3) the access methods; 4) the research and commercial areas where data usage will be allowed; 5) the post-project (re-)use of data; and 6) the post-project financing.

The agreements with the participants of the FOT/NDS have substantial impact on the possibility to re-use data after the project both by the project partners and by external parties. Thus, it is important that these topics are addressed in the agreement. It is difficult to reach participants after the project has been concluded to ask for additional consent. The participant agreement explains the project to the participant and it is vital that the participant

understands the use of the data during and after the project. From a data sharing standpoint, it is especially important to describe: 1) what data is collected; 2) where the data will be stored; 3) who is responsible for the data; 4) who will have access to what data and on what conditions; and 5) the access procedures. As each participant allows the project to follow the participant´s private life for a period ranging from a few weeks up to more than a year, it is important that they have a solid understanding of what the data could be used for. The participant should make an active consent to the most vital topics for data sharing.

Datasets are often enhanced with data collected from sensor systems bought from suppliers and put on the vehicles, and data from external data providers such as companies providing map data, weather data, or other services. Non-disclosure agreements and contracts should be signed and it is important to be aware of the topics that can affect future research, due to possible restrictions in data use.

2.2. Data and meta data descriptions

FOT/NDS studies collect a large amount of raw data, especially when continuous data-logging is favoured over event-based data collection. Depending on the aim of data re-use, simply reusing data in its most transformed/aggregated form may be sufficient. In rare occasions, it might be necessary to go back to its original, raw form. In most cases, however, cleaned-up, derived, annotated data will be the most useful. In any of those cases, the core of data sharing is that the data provided is valid or at least is documented to a level where an assessment of the level of validity can be performed. This is potentially problematic if the data re-user has not been part of the project and does not know the way the tests were performed in detail, which sensor/version was used or how the data was processed from raw data. The main problem is usually that the data itself is not sufficiently described. Therefore, good metadata is vital.

As a result, data re-use requires precise knowledge about the data itself, and hence, extensive and high quality metadata: for instance the conditions in which it has been collected, for which purpose, how it has been stored, processed, and how it can be accessed.

We propose the definition of data as "any piece of information whose value might be used during analysis and impact its result". This means that data which may be considered as "contextual", such as, for instance, participants' characteristics or weather, traffic and driving conditions is considered as data, and part of the dataset.

Conversely, we define metadata as "any piece of information necessary to use or properly interpret data".

We divide metadata into four different categories, surrounding the actual data presented in Fig 2.

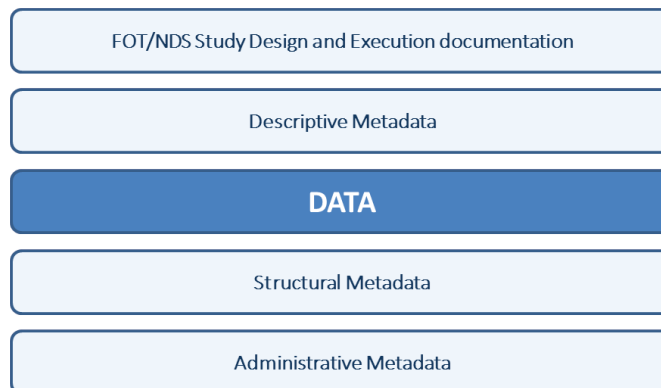| FOT/NDS Study Design and Execution documentation |
| Descriptive Metadata |
| **DATA** |
| Structural Metadata |
| Administrative Metadata |

Fig. 2. Data and metadata categories.

- 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.
- Administrative metadata, which sets the conditions for how the data can be accessed and how this is being implemented.

Data can take many successive forms, from raw collected data to very high level aggregated data, and many steps in between. A dataset is not only the result of a data collection, but also the result of an iterative process, comprising pre-processing, integration of different data sources together, calculation of derived measures and manual and automatic data reduction. A data re-use case will typically require a combination of all those different forms of data. The data can take subjective or objective forms of contextual, acquired (or raw), derived or aggregated data. Context data corresponds to all information which doesn't change with observations during the study, but helps explain them, or inform their values. It may either be directly collected or generated for the purpose of the experiment, or existing and retrieved from external data sources. It contains for instance background information such as infrastructure characteristics (e.g. map data) or vehicle and drivers' characteristics, including questionnaire results. Acquired and derived data is all data collected for the sole purpose of the analysis, during the course of the study. This includes the acquired data from vehicles, infrastructure or participants, derived time-history data from in-vehicle sensors, media (video or audio), subjective data (from annotations), data from road side equipment, or experimental conditions (e.g. traffic density or weather conditions). For the purpose of the analysis it can be relevant to analyse the data for a delimited period in time or location segments e.g. journeys, events (as defined in FESTA (FESTA 2014)), or road segments. They are defined by a combination of several conditions being met and characterized with some attributes summarizing it, some of which are automatically computed or manually annotated from video. Aggregated data is created when using relations between segments, reduced data (e.g. segment attributes) that is typically aggregated into smaller, more usable tables, suitable for data analysis.

Descriptive metadata needs to define the dataset and include detailed description on measures, performance indicators, time and locations segments, and their associated values. Also external data sources, subjective data from self-reported measures and situational data from video coding must be described in detail. Not only the output of the data must be described; how the data was generated and processed is equally important and this is where you build trust for the dataset. The more thoroughly the origin of a measure is being described, the greater the trust. The data is often processed several times and the final product might consist of more than one original signal. The need of detail description is crucial to create trust for data re-use. The output of the data processing must be documented and include information on data type, precision, unit, sample rate, enum specifications, error codes and quality indicators. Metadata must also include how the data was processed e.g. synchronization policies, re-sampling filters and harmonization rules.

Structural metadata is used to describe how the data is being structured in relation to other data. Data is organized in a system (e.g. a database or a file system), in a structure or database schema and in a data content format. The aim of structural metadata is to facilitate the initial phase of data re-usage by providing the necessary documentation on how the data is organized. The description should include file system, file structure and how to interpret the content of a data container. This means that all components of the dataset need to be described to facilitate re-use of the data. Over time it is important to also describe and preserve tools that can read the data. This is something that is highlighted when it comes to data archives, storing data for a very long time. But even five years after a project has ended the knowledge about specific tools might have been lost and the cost for building up the competence again might exceed the data value. It is therefore recommended to also describe the tools, the platform and prerequisites, even more in-depth if using a non-standard data container, file format or file structure.

Administrative metadata is information that is collected for effective operation and management of data storages and catalogues. Administrative information is stored along with the datasets, covering various topics. From a FOT data re-use perspective, the key role of administrative metadata is to cover access conditions, rights, ownership and constraints. The administrative metadata has a role also in data protection, defining processes, personal data management, access rights and keeping track of e.g. periodic backups.

The study design and experimental procedures must be well documented to enable analysis also by those persons and partners, who did not take part in executing the test. The main purposes of this documentation is to describe in a free form the purpose of the data collection, the experimental procedures and the important details of the actual execution including description of the test site, which are important to know before interpreting the data. As a result,

this documentation should not only contain initial plans, but also the final details of the study. The document should give an overview of purpose for which the field tests or data collection were made, research questions, sample selection criteria and overall description of recruitment, overall description of used equipment, date and timing of different phases of study, test plan and execution, how contact was kept during the study.

2.3. Data protection

Data protection is the key to create the trust needed between the data provider and the researcher to make the data owners provide access to their data. If the data provider knows that the researchers have good, proven procedures in place to keep control of who is accessing the data and that the researchers have knowledge in the legislation surrounding the handling of personal and IPR data, the more data they are willing to share. This section first explores data classification that will affect the data protection level, then describes different roles, and finally presents a categorisation of data access methods. The actual requirements on data protection, and also proposed implementations, are to be found in FOT-Net Data protection draft report.

The data protection level needed depends on the harm the revealed data could do. Two types of data need specific protection: personal data and proprietary data. Personal data is classified as sensitive personal data, and more general personal data (European Directive 95/46/EC Art. 2 1995). Sensitive personal data include, on a high level, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life. In the vehicle and traffic research accident databases often include sensitive personal data. The requirements on this type of data are beyond the scope of this section. Personal data could include video of driver or video of persons outside a vehicle, GPS, or other attributes that can help identifying a person. Indirect identification is a more difficult topic: for example no long-time and detailed travel diaries containing addresses can be made public, even though the person making a single trip in the diary can actually be anyone living or working at the addresses. Proprietary data, if revealed, could potentially harm a commercial company. The provision of proprietary data is usually accompanied by agreements, stating the conditions for access and use. Other data is by his classification considered non-sensitive data. The data protection requirements of this data are of course less restrictive. If the dataset consist of personal, sensitive personal or confidential data, it is mandatory to take actions to ensure data protection, even for a minor dataset. If the data is classified as non-sensitive there are fewer requirements that are mandatory, but it is still recommended to investigate all requirements.

An organisation can handle the data and act in the role of a Data Centre (DC) and/or analyse the data in the role of an Analysis Site (AS). Even if the DC and AS reside within a single organisation it might be practical to use this distinction, especially if managing personal and/or confidential data. An organization will per definition be considered a DC when managing any FOT/NDS data and making it available to more than one user. A person downloading a dataset to her computer is therefore not considered a DC, unless the data will be provided for more persons. An organisation will per definition be considered an AS if it will perform analysis on FOT/NDS data.

This chapter presents data access in four different ways. The data could be downloaded via a public or restricted website, transferred on hard drives to the research organisation, remotely accessed at the data provider or only accessed at the premises of the data provider. Each method has its own implications and it is usually the data type that has a large impact on the conditions for selecting a method. Public download means that the dataset is downloadable from a public accessible space (e.g. a web or an ftp server). This data must be non-sensitive as it is not possible for the data provider to follow-up on the use of the data. The dataset could be under a license that sets conditions or restricts the usage of data. The organisation downloading the data will by definition be considered a DC, and also as an AS, if performing analysis on the data. Conditioned download means that the dataset is transferred between two (or more) parties that agree on the conditions. The data will be transferred from the data provider to the requesting organisation using portable disks or by an agreed Internet protocol. There are no restrictions on data categories but it is mandatory for the data provider to consider all related agreements. The dataset could be under a license agreed between the parties that sets conditions or restricts the usage of the data. The requesting organisation downloading and by that managing data, will by definition be considered a DC and an AS if later performing analysis. Remote access means that the data will not leave the DC and that all analysis is performed within the data provider's IT-infrastructure. There are no restrictions depending on data classification but it is mandatory for the data provider to consider all related agreements. The dataset could be under license agreed

between the parties that conditions or restricts usage of data. The requesting organisation will be considered an AS. When remote access is not possible due to legal, contractual or bandwidth limitation, on-site access might be the only option. Then it is up to the data provider to, on their premises, allow access to the data for an external partner. In this case the data provider will be acting both as a DC and an AS.

The different data categories can be shared given the different access methods. It is important to emphasize that sensitive personal data must be shared with great care.

2.4. Training

Most datasets collected in FOTs, NDS or pilots includes personal data or data from vehicles ruled by Intellectual Property Rights (IPR), such as video data, GPS or data collected from the vehicle. All personnel handling this data need to receive training, to be aware of the level of data protection needed to handle the data securely. Training on personal integrity issues and IPR needs to accompany the general training on the data security measures put in place to protect the data. The latter includes data protection procedures, the analysis environment set-up together with more general information and rules following the specific dataset at hand. This material can be created by the data provider or the data AS. It is recommended to organize a personal training session in order to be able to answer questions and make sure that all staff members know their responsibility. As a complement, there are different kinds of personal integrity training available, e.g. the US NIH training course (US NIH 2015), where the data user gets a certificate at the end of a web course. It is recommended to document all training through an analyst's information sheet, which the participant should sign and to keep records of persons who have undergone training.

The training material is suggested to include; (1) Description of the data, including anonymization, and data structure; (2) what is personal and IPR data; (3) information on data ownership and access rights; (4) Information on applicable laws, regulations and rules and their implications on the data handling; (5) the participants consent relating to data handling; (6) explanation of data handlings procedures; (7) practical rules and procedures for data access; (8) information on publication rights; and (9) contact persons/ data protection responsible.

2.5. Support and research services

The result of re-using third-party data is dependent on the level of documentation and usually also on the available support, especially if the researcher is not so familiar with the type of data. Support services are there to assist the researchers during the analysis, whereas research services are more targeted towards doing part of the analysis work, such as extracting usable datasets for the data re-user or even perform the whole research itself and provide results.

The support starts already at the application stage, where the need for support during the analysis is established. The services could be divided into following four categories; (1) information and data provision; (2) supporting tools; (3) assistance on dedicated research needs; and (4) data protection and analysis facilities. The information and data provision category could include services such as information on available datasets including metadata and tools, data usability and established analysis procedures. Basic aggregation of data, extraction and shipping of data are other support services that could be provided. The supporting tools service deals with providing potential viewing and annotation tools, scripts to extract useful datasets or entire frameworks for retrieving and processing of data. As researchers use different ways of analysing the data, it is important that the analysts can choose what tools to use and that they are not dependent on complex frameworks or other constraints other than the raw data formats and data descriptions. It is also important that the dependency on proprietary software for access is kept to a minimum. The known European FOT tools are available online on the FOT-Net Data website. Assistance on dedicated research needs can be provided in terms of dedicated advice on analysis methods and minor modification of tools to be better suited to the analysis at hand. The final part of the support services is analyst training and support on personal integrity issues as part of the data protection and analysis facilities together with advice on data protection measures and provision of secure facilities for analysis work for data with high data protection demands.

The research services go beyond the initial start-up provided by the support services. They could be divided in the following levels; (1) research advice on methodology; (2) research involvement / research support; and (3) complete research performance. Research advice on methodology focuses on the actual research questions, hypothesis and the

method to answer these, including services such as review if a dataset can be re-used or investigation of appropriate scientific approach/method for the analysis. On the second level, research involvement / research support, the service performs part of the research such as data handling, development of specific tools, data analysis or data evaluation. Finally, a complete research work performed by the data provider itself or by a third party is the most advanced stage of research services.

2.6. Financial models

The collection of large datasets involves huge amount of efforts and resources. To further benefit from these datasets and make better use of the invested resources, the datasets could be made available for further re-use. Also minor datasets could be interesting for further use, focusing specific areas. Selection criteria for which data to store and extend the use of after the project ends, could be; (1) potential for re-use from scientific and also maybe business perspectives; (2) efforts needed to store the dataset; (3) quality and amount of data; (4) funding requirements to plan long-term data preservation and management; and (5) availability of funding for providing the data. If the dataset is large, the funding becomes crucial, as the cost of maintaining and providing the data becomes substantial.

The following data management costs have been identified; (1) data selection, enhancement of documentation (metadata), creation of entries in relevant data catalogues; (2) management & coordination personnel costs; (3) IT operations; (4) analysis or data handling facilities; (5) analysis support services; (6) promotion and advertisement; and (7) standardisation and collaboration regarding dataset formats (optional). The cost items of FOT data management have many things in common with open data efforts and large-scale user tests in various scientific disciplines. As FOT and NDS datasets generally though are collected from a variety of sources and usually are very complex to handle and retrieve, the documentation and support during data re-use are more in focus then with other open transport data. In addition, strict requirements to uphold user privacy and product IPR may require secure facilities and processes and thereby raising management costs of such datasets when compared with fully open datasets.

The provision of funding for the above-mentioned data management costs could be divided into three different types of funding; direct governmental funds, project-based funding and end user funding e.g. through membership fees or licenses. There are several financial models based on the three funding sources, see Table 1. Different funding sources might be required to keep data available and provide services for third parties. Therefore the financial models can also be used complementary.

Table 1. Financial models based on the funding source.

Source for funding	Financial model
Direct governmental	Funding to the organization collecting the data.
	Funding to data infrastructures serving multiple organizations.
Project-based	Project budget allows for archival in commercial services.
	Project extension focusing provision of data stored at project partners.
	One or several new projects finance further data provision from the data provider.
End-user	Network of organisations using partner fees to fund collection and maintenance of data.
	Commercial analysis services, based on several valuable datasets. Fee includes data maintenance.
	Data integrators – companies acquire and market transport research datasets along with other related datasets.

Depending on the financial model and activities set up for data sharing, the costs are shared differently between the project that collects data, the organisation(s) that owns the data and finally the re-users. The costs could be divided here into three classes; dataset finalisation, continuous costs coming from management and upkeep of data and costs when data is shared, e.g. selection of data and user support.

2.7. Application procedures

The project collecting the data should at an early stage of the project agree on the conditions for re-using the collected dataset and on an application procedure for re-use. This will facilitate that new research applications which want to utilize the data, will have taken the data application time and potential costs for re-using the data into consideration already during the proposal phase, before the application is sent to the targeted call.

The application procedure shall at least address the following items: 1) where to apply; 2) what information is needed to be provided to be able to evaluate the application; 3) who can approve an application, response times, and conditions to be taken into account in the approval decision; 4) requirements on mandatory training in data protection and integrity issues; 5) information on the data access procedure; 6) requirements on data protection; 7) potential costs for data access, support, and research services; 8) requirements on acknowledgements on publications, reports, and presentations; and 9) documentation of data applications and the related approval decision(s).

To facilitate the actual application, a template could be developed guiding the applicant in what information is needed. The following information is recommended: (1) applicant details; (2) short project description; (3) requested data set; (4) use and expected results; (5) information on the intended publication of the data; (6) list of persons to get access and the related access time period; (7) need of training in data protection and integrity issues; and (8) need of support and research services. The application form together with the application procedure could be provided on a website. The availability of the data could also be made public through data catalogues.

3. Sharing UDRIVE data in line with the DSF

The development of the DSF and its hands-on implementation in the EU project UDRIVE has progressed in parallel. UDRIVE is a large European NDS, collecting sensitive data such as video and GPS, which requires strong data protection measures.

UDRIVE is the first EU project that has taken the data sharing pre-requisites and implemented them cross Europe and the project is the first test bench for the DSF. In 2012, during the grant agreement and the consortium agreement negotiations, numerous issues was discussed and settled upon, such as who can re-use the data, where could it be stored and how the access should be performed. In the consortium agreement, only the overall issues are solved. At present, the consortium is discussing the details of the procedures for accessing the data for partners and third parties. A concept of remote access has been developed and tested in the project, which will also be used after the project.

The data is currently being collected and will be described in detail, to facilitate for up-coming data re-use. As the data consists of both video and GPS together with IPR data, the data protection was in focus from the start of the project. A concept was developed, covering the data protection from the data collection in the vehicles to the data protection to be put in place at the partners before analysis using personal data can be started. During the whole process of handling the data, the different partners are responsible for training their personnel in data protection, personal integrity and IPR handling.

Currently the EC is opening several calls, many where UDRIVE data could be of interest. Analysis organisations are planning to utilize the data for their research and therefore it is urgent to decide on the conditions on which the data can be re-used. Many conditions were already decided upon during the consortium agreement negotiations, but one important factor influence the larger scope of data sharing. The project and EC has agreed that the data should be available for data re-use after the project. Still, the financial pre-requisites for such a platform is not solved. If there is no funding for the maintenance cost of the data, the data will be archived for three years and then probably deleted. This is a major concern of many projects with large FOT/NDS datasets, and a decision regarding the funding needs to be taken soon, to set the rules for the up-coming projects, who would like to re-use UDRIVE data.

The offers of support and research services is not yet detailed, but will be elaborated closer to the end of the project. Also the application forms and procedures will be developed next year, well before the end of the project.

4. Conclusions

The availability of a common DSF has the potential to substantially contribute to a larger use of the collected FOT/NDS data. The framework includes data sharing pre-requisites that could be integrated into project agreements

from the start, as well as procedures and templates to facilitate easier data sharing. Data providers are given hands-on guidelines and are made aware of the importance of data and metadata descriptions and data protection as well as different financial model possibilities for making the data available after the end of the data collecting project. Researchers wanting to re-use already collected datasets could then utilise a more or less standard application procedure, rely on already performed training that is widely accepted, and plan for the costs that using a specific dataset might cause the project.

The recommendations and guidelines in the DSF will need continuous revisions to be valid for the evolution of the data sharing and re-use of data. There remain challenges to make data sharing possible on a global scale. Some of these are the project funding schemes, leading to multiple ownerships of data, and the legal settings in different countries Furthermore, how do we make new projects aware of the importance of inserting the pre-requisites for data sharing into the different project agreements right from the start?

Many of the topics in the DSF are general and can be used for other types of transport research data as well. If the use of the DSF is extended beyond FOT/NDS data, the later combination of these datasets for research is facilitated if the same application procedure and conditions for re-use are in place.

Data owners and researchers searching for data need to be brought together and the FOT-Net Data project has developed a data catalogue on available data from earlier FOTs/NDSs to facilitate this exchange of data.

References

Australian NDS 2015, http://www.tars.unsw.edu.au/research/Current/Australian_Naturalistic_Driving_Study/australian_nds.html, retrieved August 2nd, 2015.
Barnard, Y. (Ed) 2014a. FOT-Net 2 Deliverable D2.2 Report on FOT Network activities.
Data and metadata description 2015. Draft Framework for Data and Metadata Description. Available at: http://fot-net.eu/Documents/draft-framework-for-data-and-metadata-description/.
Data protection 2015. Data Protection Recommendations draft. Available at: http://fot-net.eu/Documents/fot-net-data-data-protection-recommendations/.
DSF 2015. Draft Data Sharing Framework. Available at: http://fot-net.eu/Documents/data-sharing-framework/.
Eenink, R., Barnard, Y., Baumann, M., Augros, X., Utesch, F., 2014. UDRIVE: the European naturalistic driving study. Proceedings of the Transport Research Arena, Paris.ETC 2.0 2015, https://www.go-etc.jp/english/etc2/index.html, retrieved August 2nd, 2015.
European Directive 95/46/EC Art. 2 1995, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31995L0046, retrieved August 2nd, 2015
FESTA 2014. FESTA Handbook, Version 5. Available at: http://wiki.fot net.eu/index.php?title=FESTA_handbook.
FOT in China 2014, http://thinkinghighways.com/913/, retrieved August 2nd, 2015.
Gellerman, H., Svanberg, E., Bärgman, J., Barnard, Y. 2015. Data Sharing Framework for Naturalistic Driving Data. Proceedings of the 22nd World Congress on Intelligent Transport Systems in Bordeaux.
Safety pilot 2015, http://safetypilot.umtri.umich.edu/, retrieved August 2nd, 2015.
Selpi, Borgen, S., Bärgman, J., Svanberg, E., Dozza, M., Nisslert, R., Norell, C., Kovaceva, J., Sanchez, D., Saez, M., Val, C., Küfen, J., Benmimoun, M., Metz, B., (2011). Deliverable D3.3: Data Management in euroFOT. http://eurofot--ip.eu/download/library/deliverables/eurofotsp320110831v10dld33_data_management_in_eurofot.pdf, retrieved August 2nd, 2015.
SHRP2 2015, http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/SHRP2FocusAreas.aspx, retrieved August 2nd, 2015.
US NIH 2015, http://phrp.nihtraining.com/, retrieved August 2nd, 2015.