A PLATFORM FOR SHARING DATA FROM FIELD OPERATIONAL TESTS

Yvonne Barnard
ERTICO – ITS Europe
Avenue Louise 326
B-1050 Brussels, Belgium
y.barnard@mail.ertico.com

Sami Koskinen
VTT Technical Research Centre of Finland
PO Box 1300, 33101 Tampere, Finland
sami.koskinen@vtt.fi

Helena Gellerman
SAFER Vehicle and Traffic Safety Centre
PO Box 8077, SE-402 78 Gothenburg, Sweden
helena.gellerman@chalmers.se

ABSTRACT

In this paper we will describe how a platform is being developed for sharing data collected in Field Operational Tests (FOTs). In these tests often a huge amount of data is gathered about driver behaviour and the interaction between drivers, vehicles, ITS, infrastructure and the traffic environment. These datasets are often only partly analysed, because of targeted research goals and lack of resources. They may be of great value for other organisations and researchers in order to answer further research questions and to determine the impact of wide-scale use of ITS. The paper will describe a how a European platform for enabling data sharing and data re-use is being developed.

Keywords: Field Operational Tests, Data Sharing, Data Re-use, Data Analysis

INTRODUCTION

In recent years, a large number of Field Operational Tests (FOTs) have been conducted. In Europe, the European Union has since 2008 supported a number of projects testing the latest vehicle information technology in large-scale field trials. Thousands of drivers have been able to test the most promising prototypes or products just entering the markets. Drivers have been testing systems such as adaptive cruise control, forward collision warning, navigators, and most recently, warning systems based on short-range wireless communication between vehicles. The communication can provide information, for example, on nearby car accidents or approaching emergency vehicles. FOT projects have evaluated the impact of these technologies, and have also contributed to their introduction. Drivers' behaviour whilst using the systems has been monitored for continuous periods of up to over a year, collecting valuable information from traffic.

In FOTs, large amounts of data are being gathered. These data may be derived from the vehicle itself, from the CAN-bus, GPS, video cameras pointing at the driver or at the outside environment, or from ITS that are being tested. Other data sources may be mobile and nomadic devices such as smart phones and navigation systems, or subjective data sources such as questionnaires and travel diaries. The recent emphasis in ITS projects on cooperative systems, vehicle-to-vehicle and vehicle-to-infrastructure communication, and the use of floating vehicle data and other probe data leads to even more data becoming available.

DATA SHARING

Most datasets gathered in FOTs are huge, and often not enough resources and time are available to analyse all the data in detail within the frame of a single project. Moreover, projects often focus on a special set of research questions and impact areas. For example, a project may investigate the influence of Intelligent Speed Adaptation on traffic safety. However, data gathered about the speed of vehicles may also be of interest to researchers and stakeholders interested in very different questions, for example, the speed patterns of different age groups, eco-driving behaviour, etc. Re-using data for further and new analyses seems to be the obvious answer, as collecting new data, in a new project, is very costly.

However, re-using data sounds simple, but isn’t. In order to re-use data in a new project, the original project must be willing to share the dataset, and make it available in a way ensuring that it can be re-used. Sharing data poses many challenges.

Technical challenges concern issues such as the quality of metadata, descriptions on implementations, how field tests were run, how the data was collected, the tools used to collect and store the data, the standards and formats used.

Legal and organisational issues concern ownership, data protection and privacy issues. For example, a permission from the FOT participants is needed to allow third parties access the data. Re-engineering tested services or used sensor systems is also seen as a worry.

Practical and financial issues concern questions about who is paying for the access, the training of new data analysts so they can understand the data, its limitations, tools, and physical access to the data.

Finally, there are ill-defined issues such as attitudinal issues like trust, willingness to share, and resources to be allocated to support data sharing.

Sharing and re-using data can be a major step forward in our understanding of the behaviour of transport users and systems. It will be very hard to find sponsors for large numbers of FOTs that start from scratch with their own data collection. In Europe the European Commission is stressing this need for data to be open and shared, as they are one of the main sponsors of data collection. A wealth of information is hidden in the datasets that have been collected in recent years, so the question is how we will be able to generate new knowledge out of these existing datasets.

Re-using research datasets is seen to have several benefits on a societal level: it will yield further research results, support education, improve collaboration and e.g. contribute to market introduction of new systems by enabling several organisations to assess benefits. For such reasons, across different fields of science, publicly funded research projects will be required to share more of their collected data in the future. The practices are being set up and tested e.g. in EC's eInfrastructure and Open Research Data Pilot projects as well as in targeted projects such as FOT-Net Data, addressing FOT data sharing.

THE FOT-NET AND FOT-NET DATA PROJECTS

Over the past six years, the European support action FOT-Net has worked on building a community of people involved in FOTs, both on national and European levels, with further international connections. The people come from research organisations, car manufacturers, system and service providers, user organisations, and public authorities and road operators. FOT-Net promotes and maintains the so-called FESTA methodology and the corresponding handbook. This methodology provides detailed support for designing and conducting FOTs. Also workshops, a website, and a wiki containing a catalogue of all known FOTs have been developed to ensure wide sharing of knowledge and experiences.

By building a FOT community common understanding and approaches, trust and networking between organisations have established the pre-conditions for willingness to share. Now we are taking the next steps, addressing the technical, practical and organisational issues. In the project that started in January 2014, FOT-Net Data, we are further developing the European framework into a global one, aiming to gain consensus on a global level for a data sharing strategy and platform, taking into account the work that has been done in the US and Japan.

DATA SHARING PLATFORM AND STRATEGY

A data sharing platform consists of sharing principles and data access procedures, data descriptions (metadata) and standards, and descriptions of technical tools and methods for data sharing. Such a platform is to be used in two major phases. The first concerns the steps to be taken and issues to be resolved from the very beginning of a FOT, such as defining participant agreements for allowing data re-use and resolving data ownership issues. The second is to address the actual data sharing and the procedures, templates and services needed for successful research on data gathered in earlier projects.

Principles and procedures

The main principles and procedures in data sharing concern procedures for application and approval, and templates for application, data re-use in consent forms, data sharing agreements and data sharing texts for consortium agreements. In addition, support and research services, financial models, and personnel education on legal and ethical issues need to be developed.

A data-sharing platform needs to take into account the views of different kinds of stakeholder, serving research, commercial or policy purposes. Data protection, privacy and security are crucial prerequisites to data sharing. Different types of levels of protection are required depending on the sensitivity of the data. Guidelines are needed for each level for how data could be practically accessed and shared.

Data descriptions and tools

The descriptions include details about the test execution and collected log data. The data descriptions cover the data collected in the tests, enhancing data such as map data and the equally important meta-data, that helps understand the background to the collected data. Special attention should be paid to information about video annotations, the subjects and vehicles collecting the data, and the experimental protocol used, and study design applied.

Another issue is the harmonisation of data of the same type across datasets. Ideally a minimum list of data types with minimum requirements for a harmonised dataset should be set-up, which all future FOTs are recommended to collect. In this way the research community would be able use a common set of data for a variety of purposes.

Tools and methods

The technical issues concern tools and methods for data access, processing, enrichment, and analysis. Tools that can be used to support data sharing as well as data post-processing methods and data aggregation types are of interest. Data processing tools and scripts are generally very helpful for an analyst when made available together with FOT data, allowing reading the original raw data, interpret it, enhance it and calculate indicators.

In the FOT-Net project a tool catalogue was developed and made available at the FOT-Net wiki. However, more information will be collected about tools that are helpful for data sharing purposes, tools to alter data to make it anonymous and availability of tools specific to a dataset.

DATA CATALOGUE

To make FOT data sharing successful on a large-scale a third action is needed next to community building and platform development. Knowledge is needed about what data are actually available, who the owners are, and how it can be reached. This knowledge needs to be brought to the attention of researchers and organisations who could exploit the data, either for research or product development. FOT-Net Data will build a detailed catalogue of available data, connected to the existing FOT catalogue, enabling organisations to easily assess the value of different datasets.

The first thing that organisations looking for to analyse FOT data need, is a thorough description of potentially available datasets. For example, what was the number of drivers, what was the study design, what was logged, is video available, and what has been analysed so far. When a dataset seems interesting, it's necessary to know under what conditions it could be available, would there be tools and data viewers to make analysis easier, and who to contact.

The catalogue will be in the form of a public website with linked content. The datasets will be visible on the FOT catalogue, but the data remains in original storages. The data providers make the final agreements with interested organisations. The project will arrange an option to share public samples of datasets through the Data Catalogue. FOT-Net Data will experiment with examples of data-sharing. Resources are reserved to make a few interesting and high-priority datasets consistent with the data sharing framework.

CONCLUSIONS

There are two basic options for data management and access after an FOT project ends: 1) the data stays with the organisation that collected it 2) the data is also submitted to an archival service. In both cases, a research project needs to document collected data in detail to enable re-use. A catalogue of available datasets, containing main information on test execution and collected data types, is needed as well for performing searches.

FOT-Net Data creates a Data Sharing Framework including guidelines on data management, documentation and sharing agreements. FOT-Net Data also creates an index of available FOT datasets by contacting and working together with several organisations. Although FOT-Net Data does not host FOT data as a data archive, the work will also support future archival efforts where publicly funded projects could be obliged to submit their data before the projects end. FOT-Net Data promotes data re-use and continues to offer a forum to discuss all aspects of running a FOT.

Acknowledgements

FOT-Net Data is a Support Action in the seventh Framework Programme Information and Communication Technologies. It is funded by the European Commission (EC), DG Connect, under Grant Agreement number 610453.