Data Management and Data Sharing in Field Operational Tests

Yvonne Barnard\textsuperscript{1*}, Sami Koskinen\textsuperscript{2}, Satu Innamaa\textsuperscript{2}, Helena Gellerman\textsuperscript{3}, Erik Svanberg\textsuperscript{3}, Adrian Zlocki\textsuperscript{4} and Haibo Chen\textsuperscript{1}

\textsuperscript{1}Institute for Transport Studies, University of Leeds, UK
\textsuperscript{2}VTT Technical Research Centre of Finland Ltd, Finland
\textsuperscript{3}SAFER Vehicle and Traffic Safety Centre at Chalmers University of Technology, Sweden
\textsuperscript{4}Institut für Kraftfahrzeuge (IKA) at RWTH Aachen University, Germany

Abstract

In this chapter it will be discussed how data from Field Operational Tests of Intelligent Transport Systems can be managed and shared. The Field Operational Tests, where hundreds of users get to experience the latest systems, aim to assess the impacts that would result from a wide-scale implementation. Evaluation principles of Field Operational Tests will be explained, and a closer look will be taken at the data that is collected for carrying out the assessments. The widely used FESTA methodology for designing and conducting Field Operational Tests and Naturalistic Driving Studies already provides several recommendations for managing data. This methodology will be discussed and illustrated by examples of its use in European projects. As field test projects set out to collect a huge set of data, the projects themselves do not usually have the scope or the resources to analyze the data from every perspective. Therefore re-use of the collected data also by other projects with different research questions has the potential to generate a wealth of new knowledge about what is happening in the interactions between drivers, vehicles and the infrastructure. Data sharing is the focus of a European support action, FOT-Net Data. The support action is working, with international collaboration, to form a data sharing framework, a data catalogue, and provide detailed recommendations for

\textsuperscript{*}Corresponding author: Y.Barnard@leeds.ac.uk
sharing and re-use. Outcomes from this activity will be discussed. Ways of sharing different types of data will be described, including the necessary steps to be taken to open up the data.

1. Introduction

Over the past decades new technologies for Intelligent Transport Systems (ITS) have been researched and developed. In Europe these activities have often been supported by the European Research Framework Programmes as well as national programmes. The short and long term impacts of these systems need to be understood to answer questions which are crucial for market introduction and penetration. By testing the systems on a large scale, in real driving conditions during a significant period of time, Field Operational Tests (FOT) can answer most of these questions. The results of FOTs enable policymakers to establish the right framework for deployment of these systems, and business leaders to make informed decisions about their market introduction. Since 2008 the European Union has supported a number of projects enabling testing of 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. The 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 on, for example, nearby car accidents or approaching emergency vehicles. Field test projects have evaluated the impact of these technologies, and contributed to their introduction. Drivers’ behaviour whilst using the systems has been monitored for continuous periods of up to more than a year, collecting valuable information from traffic.

In the European Commission funded FESTA project a common FOT methodology was developed [1, 2], which is now being widely used as the basis for the planning and execution of FOTs and Naturalistic Driving Studies (NDS). The naturalistic studies concentrate on studying everyday driving behaviour and events that may lead to accidents. Since the original FESTA project, the methodology has been maintained and updated by European FOT-Net [3, 4] support actions.

The FOTs have a crucial need for a platform of knowledge exchange in order to let individual FOTs benefit from each other’s experiences. Without such exchange, it takes long especially for newcomers to learn the several practical steps of arranging field trials. A platform was set-up in 2008 as the FOT-Net support action, funded by the EC DG Information Society. The first FOT-Net ended in 2010 and was followed by FOT-Net 2 from 2011 to 2014, and FOT-Net Data, which will end in 2017. FOT-Net supports and coordinate the FOT community by sharing experiences,
Data Management and Data Sharing in Field Operational Tests 61

providing methodology training, building a knowledge base of past tests, maintaining and updating the FESTA methodology and promoting the FOT results. FOT-Net has also developed strategic networking for national, European and global FOTs (mainly North American and Asian-Pacific). The latest support action in the series, FOT-Net Data, promotes especially the sharing and re-using of data gathered in FOTs. Its main objectives are in development of a global data sharing framework in collaboration with world-wide stakeholders and keeping a catalogue of available datasets. All information and documents concerning the FOT-Net support actions are available at the FOT-Net website (FOT-Net).

The Data Gathered in FOTs and the Knowledge Derived from the Analysis

FOTs are performed to gather knowledge about the impact the large-scale introduction of ITS may have. The socio-economic impacts that are usually considered are:

- Safety: will ITS improve road-safety, for drivers but also for other road-users?
- Mobility: will ITS have a direct effect on traffic flow and mobility, and indirect effects such as reduced crashes and will it change mobility patterns of citizens?
- Environmental: will ITS reduce CO₂ emissions and pollutants, for example, by changed driving behaviours or reduced congestion?
- Economical: what will ITS mean for the costs of systems in vehicles and on the road, and for societal costs such as those caused by congestion?

If one looks closer into how these impact areas are influenced one can use the six areas of impact defined by FESTA [1] based on Draskóczy et al. [5]. Although this approach was originally designed for formulating hypotheses on traffic safety impacts, it is in fact equally well applicable for efficiency and environmental impacts.

The six types of effect are [1, 6]:

1. Direct effects of a system on the user and on driving
2. Indirect (behavioural adaptation) effects of the system on the user
3. Indirect (behavioural adaptation) effects of the system on non-users (imitating effect)
4. Modification of interaction between users and non-users (including vulnerable road users)
5. Modification of accident consequences (e.g. by improving rescue, etc. — note that this can affect efficiency and environment as well as safety)
6. Effects of combination with other systems
In FOTs a variety of complex data is gathered, relative to the specific research questions. Three types of data may be distinguished.

*Data collected by sensors in a vehicle*, for example, vehicle controls such as speed or turn indicators, or position of steering wheel or pedals, GPS traces, vehicle kinematics from accelerometers, or video recorded from one or many cameras mounted in the vehicle, heading to the driver or e.g. the front view. Such data was, for example, collected in the first European large-scale FOT euroFOT [7]. Depending on the different vehicles (from different manufacturers) within the field tests different objective data was collected from vehicle dynamics sensors, environment perception sensors or additional image processing data from cameras filming the scenario, the driver and his interaction with the control elements of the vehicle. Objective vehicle data was processed and evaluated for the research questions of the project [8].

*Data collected by road-side units and back-end systems*, including contextual data, such as traffic flow and weather data, and communication data. For example, the DRIVE C2X project stored messages communicated between cars and road-side units, and traffic flow information derived from inductive loops in the road infrastructure. Internet map and weather data sources were also used in enriching vehicle data, enabling assessment of, e.g., speeding behaviour in different weather conditions [9].

*Subjective data gathered from the participants in the FOT*, such as questionnaire data and travel diaries. For example, the TeleFOT project collected subjective data of participants’ uptake and acceptance with four consecutive questionnaires (before field tests started, twice during the tests and after the tests were completed) [10]. Another source of subjective data in TeleFOT was travel diaries. Participants were asked to fill in a travel diary (timing, length, mode, origin, destination and purpose of travel) for one week similarly four times during the FOT with the same frequency as the questionnaires to collect data on potential mobility impacts [11].

In addition to the collected raw data, FOT datasets often consist of derived data and aggregated data. Derived data usually consists of cleaned-up, filtered and resampled versions of the raw signals. Aggregated data represents computations of different aspects of data segments. It may be the average speed during a trip or the number of passes through a specific intersection.

Another dimension of the data is subjective manual video annotations, often generated when reviewing specific events in retrospect. Real-time observations can also help to provide useful understanding of FOT data.

The analysis of data gathered by FOTs can provide evidence concerning the effects of ITS introduction, supporting the decision making of industry and policy makers. It also provides a variety of data on what is actually happening on the roads, and how drivers behave, knowledge that will be very useful for future transport research.
2. Collecting and Managing Data in FOTs

As FOTs are complex studies, a structured approach is needed to set-up the study, collect and analyze the data and determine the impact of the systems evaluated. In 2008 the FESTA methodology was developed to provide guidance and support for the upcoming EU-funded FOTs. A handbook was produced with many practical recommendations [1]. The basis of this handbook was a methodology, to be followed by the FOTs. This methodology has not only been adopted by FOTs funded by the European Commission but also by many nationally (or otherwise) funded projects, and has influenced FOTs outside Europe. The FESTA methodology is summarized in Fig. 1. It includes several steps, which, although described in linear way, are performed in iteration. The V-shape shows the dependencies between the different steps on the left- and right-hand side of the V. The main steps can be summarized as follows:

- Defining the study: Defining functions, use cases, research questions and hypotheses
- Preparing the study: Determining performance indicators, study design, measures and sensors
- Conducting the study: Collecting data
- Analyzing the data: Storing the data, analyzing the data, testing hypotheses, answering research questions
- Determining the impact: Impact assessment and deployment scenarios, socio-economic cost-benefits analysis

![Figure 1: The FESTA methodology.](image-url)
FOTs produce large amounts of complex data and the main challenge is to make the data manageable and easily available for analysis [12]. As each FOT has its own type of data and data management demands, a generic model that is suitable for all FOTs does not exist. Often the data recorders lay the foundations of the FOT dataset by grouping data into segments, e.g. by trips or events. By establishing these segments it is possible to attach rich metadata that can help filter the dataset.

Documentation is a key requirement. The data need to be described in detail, but also the tools and processes implemented during the FOT. To understand the dataset the study design (the objectives of data collection) is important information, as well as the detailed test protocols with relevant test scenarios. It is also important to describe the structure of the dataset, how it is organized and stored, to facilitate use of the data.

Data quality assurance is another key issue. Procedures ensuring the quality have to be applied from the very beginning of the planning and piloting of the data collection, and the subsequent storage and processing.

3. FOT Data Collection: Projects and Data Catalogues

FOT projects gather large data-sets, usually from the three sources described above: in-vehicle sensors, road-side units (for projects studying cooperative systems) and from drivers. In this section one shortly describes several large FOTs and give an indication of the type of data they collect, but this selection is by no means exhaustive. In the FOT-Net wiki, a catalogue of FOTs may be found, listing over 150 projects [4].

The largest datasets have so far been collected in the US (e.g. IVBSS [13], SHRP2 [14] and Safety Pilot [15]) and in Europe (e.g. euroFOT [19, 20], TeleFOT [21], DriveC2X [16], FOTsis [17], and UDRIVE [18]).

The IVBSS [13] project designed four Integrated Vehicle-Based Safety Systems, and tested the systems in cars and heavy trucks [13]. Data includes vehicle sensor signals, video data, questionnaires and focus groups. The world’s largest naturalistic driving study SHRP2 [14] includes data from normal driving from over 3000 participants during 1-2 years each (SHRP2) [14]. Data includes in-vehicle signals, internal and external videos, radar, questionnaires, weather data and detailed map data. The Safety Pilot is a connected vehicles project where over 2000 vehicles are instrumented and the communication data is collected (Safety Pilot) [15]. On some vehicles, data about the driver behaviour is also collected (GPS, speed, vehicle signals, video data, questionnaires).

In Europe several large-scale FOTs [4] have collected data (some of the projects are listed above). The euroFOT project was one of the first large-scale FOTs of in-vehicle Advanced Driver Assistance Systems (ADAS) performed on the roads of several European countries [19, 20]. Data
collection included vehicle dynamics, data from environment perception sensors and video. The TeleFOT project studied the use of nomadic devices. Data collection included GPS data, video data, travel diaries, questionnaires, focus groups and interviews [21]. FOTsis [17, 22] and DRIVE C2X [16, 23] studied cooperative systems. Data collection included communication data, GPS and HMI data, video data, questionnaires, focus groups and interviews. UDRIVE [18] is the first large European naturalistic driving study on cars, trucks and scooters. Data collected includes video data from eight cameras (including a smart camera), GPS data, vehicle dynamics data and a range of participant questionnaires (UDRIVE) [18, 24].

In Japan large data sets based on event recorders have been collected, such as the Tokyo University of Agriculture and Technology (TUAT) event recorder dataset (GPS, speed, accelerometer, brake pedal, turn indicator, and forward video), and Australia as well has several interesting datasets, e.g. the Australian NDS [25] collecting naturalistic data from the normal driving of 360 participants (video, GPS, accelerometer, radar, questionnaires) [26, 25]. Data collection also takes place in Korea and China.

In most of these studies, equipped vehicles were used by ordinary volunteer drivers for over a long period (months or even years) in real world traffic conditions and large amounts of data have been gathered by sensors, capturing driver and vehicle behaviour, as well as the driving context. The data has been used to answer a wide range of research questions. These data enabled the analysis of the impacts of ADAS and cooperative systems on safety, efficiency and the environment, as well as the usability and acceptance of the systems by the users [27, 28, 21, 29]. The data is also used to analyze crash causation [30, 31]. The size of the datasets varies, from gigabytes (GB) to several petabytes (PB), depending mainly on whether the data is collected continuously, and whether it includes video.

Although datasets may be huge, and contain valuable information, often not all data collected is analyzed within a project, either due to lack of time or resources, or because more data is collected than is needed to answer the selected and specific research questions. This data can be of value for research purposes other than those of the project that collected them, but this requires an awareness of its existence, the willingness and capability of its owner to share it.

In Europe datasets reside with the project or the partners in consortia that gathered the data. FOT datasets are not stored in public repositories. The FOT-Net Data support action [3, 4] (see below) is currently working on a central catalogue, where researchers looking for FOT data can find what data is available, and how they may obtain it.
The closest counterpart to the European FOT catalogue activities is the Research Data Exchange [32], which is a core part of USDOT’s Data Capture and Management Program. The RDE is a tool to access transport research data, especially on connected vehicle technologies. The data shared on RDE is anonymized, partly with the help of dedicated software tools.

While the FOT-Net wiki [4] already includes a comprehensive catalogue of field trials and naturalistic driving studies carried out in recent years across the world, the FOT-Net Data support action is compiling further details regarding available research datasets and tools related to them. The catalogue is moderated by FOT-Net [3, 4], but as a wiki it continues to receive input and updates from the FOT community – various organizations carrying out large-scale trials.

The main purpose of the FOT Data Catalogue is to support potential data re-users in identifying suitable datasets for their purposes, and to facilitate data sharing. The main principle of the catalogue is that the actual datasets remain with their owners. Data providers will make the final agreements with interested organizations, and can offer support to new analysts regarding the details of the study. The catalogue will include information on data and contacts but not the data itself. It is, however, possible to add anonymized sample data for allowing re-users to get a practical example.

This division of content between the FOT Data Catalogue and data owners is presented in Fig. 2. The figure also lists common roles: catalogue

![Diagram](image_url)

**Figure 2**: FOT Data Catalogue relationship with data owners.
operators collect statistics of information searches, and can advertise the datasets available from a community. Catalogue operators also offer basic support to researchers, while the data owners and persons who executed the test are the ones who have in-depth knowledge of it.

4. Data Sharing: Benefits, Challenges and Recommendations

Data sharing means that not only the research community, but also industry and public authorities, will have access to a wealth of information. Data from one project may be of value to other projects. For example, most FOTs collect data about the speed of vehicles. This is not only of interest to a project investigating the effect of speed adaptation systems on safety, but 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.

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 one will be able to generate new knowledge out of these existing datasets.

Opening datasets allows other institutes to verify impacts and calculations. This is an important part of scientific processes and publications. Access to data also allows for calculations from larger populations, assessing the situation in several countries. Data sharing should also be seen as a method of cooperation: the group that originally collected data obtaining complementary results from new researchers. These results will often be interesting for the original funders, and may boost new projects and consortia.

However, sharing poses challenges, which may be of a technical, legal and organizational or practical and financial nature [33].

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

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

Practical and financial issues concern questions about who is paying for the access, are there resources to be allocated to support data sharing, the training of new data analysts so they can understand the data, its limitations, tools, and the physical access to the data?

And finally, there are issues such as trust and willingness to share that may depend on earlier experiences and on how the challenges mentioned above are met. 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. The implementation of the solutions to the challenges that is chosen is therefore crucial.

To support FOT organizers and data providers in addressing these issues, the FOT-Net Data support action provides a framework providing guidelines and recommendations [34, 35]. The Data Sharing Framework consists of seven different areas, which should all be addressed in order to provide or re-use data.

1. Agreements within the project collecting data, including consortium agreements, participant agreements and agreements with third party data providers
2. Availability of valid data and meta data, including a “standard” description of the documentation of the data
3. Data protection requirements both for the data provider and the analysis site
4. Security and personal integrity education for all personnel involved
5. Support and research services to facilitate the start-up of projects and offer research capabilities
6. Financial models to provide funding for the data to be maintained and available and for access provision personnel to be available
7. Application procedures and data sharing agreements

Each chapter gives hands-on advice on what to implement, and includes checklists and topics to discuss both in the original project and in the data sharing agreement. The availability of a common Data Sharing Framework will importantly facilitate a larger use of the collected FOT/NDS data. The organizations setting up new FOT/NDS projects will not need to develop the specific data sharing content for a certain project, but will be able to focus on the project specific questions such as research questions, study design, and data acquisition requirements. Also, researchers wanting to re-use already collected datasets will be able to make use of 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 may imply for the project.
5. Conclusions: The Future of Open and Shared Data

While FOTs [4] mostly focussed on testing systems implemented in vehicles and/or infrastructure, the next large step will be the field testing of automated driving, including driverless vehicles. This means that there is an even stronger need for data sharing. As fundamental changes in traffic will start to take place, all kinds of insight in the impacts are needed to support public authorities, industry and the general public in forming their opinions and strategies. Only if data are shared widely will it be possible to find evidence on the effects of the introduction and use of automated vehicles. Industrial competition may be a main obstacle, but it is in everyone’s interest to ensure that the introduction of automation does not have negative consequences on impact areas such as safety and traffic efficiency.

Not only the sharing of new data is of interest for automation, also the existing FOT data are very useful. Data on relevant driving situations will be necessary for the design and development of automated systems. Also other evaluation methods, such as virtual simulations, need this data. This “relevant situations” data cannot be collected by one company or consortium only, so sharing is a necessity.

Data sharing and openness have long been wished for. The benefits are clear and numerous, but there are many difficult steps to be taken. Guidelines, agreements, and even changes of opinion are needed. The strongest force for effecting these changes is the setting of new conditions for public funding. When data is gathered with public money, certain levels of openness are required.

In the ITS field, during the last fifteen years, public transport data and timetables have increasingly become available via open internet interfaces required by funding organizations. Additionally, spatial map data and city databases are opening up. These allow for large data analyses to be carried out. Emerging new research topics will benefit from new analyses of existing data.

Datasets from single ITS research projects will remain diverse and will require detailed documentation to be analyzed. These datasets can, however, be increasingly open, and can be collected according to established practices. Standardization of certain key parts (position, map, communication, etc.) of data may be foreseen, and may make analysis and documentation of data more straightforward. However, up to the present moment even GPS data has always been logged in slightly different formats. The varying formats require flexibility from analysis software, requiring first setting up reader components to gather the information required for analysis.

FOT-Net Data [3] and similar efforts will define the minimum level of documentation and data to collect in FOTs, and set up public data
catalogues. These being available, some major hurdles, such as absence of information on what is available, of detailed documentation of the contents of a dataset, of the conditions for its re-use, and even on whether or not the participants of the study have given their consent for re-use, are no longer in the way.

When projects can refer to existing documentation, the effort of data collection is likely to be reduced, instead of placing further burden on them. Harmonized processes and methodology also contribute to the collection of high-quality data that is of real interest for further work. Finally, re-use of datasets is gradually gaining ground in scientific work, and it should no longer be a principle that all scientists gather their own dataset to work with.

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, Grant Agreement number 610453.

References

systems in the European field operation test “euroFOT”. 8th International Workshop on Intelligent Transportation (WIT 2011), Hamburg.


