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

Data Catalogue

Deliverable No.	D4.1
Dissemination level	Public
Work Package No.	WP4
Main author(s)	Satu Innamaa and Sami Koskinen (VTT)
Co-author(s)	
Status (F: final, D: draft)	F
Version number	3.0
Date	27 January 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(s): Satu Innamaa and Sami Koskinen (VTT)

Work area: WP4

Document title: Data Catalogue

Version history:

Version number	Date	Main author	Summary of changes
0.1	02.09.2014	Sami Koskinen	Draft table of contents
0.2	02.09.2014	Satu Innamaa	Modification of draft
0.3	09.10.2014	Sami Koskinen	State of the art, description of a few datasets and a status report on WP4.3
0.4	27.10.2014	Satu Innamaa	Chapter 3
0.5	30.10.2014	Adrian Zlocki	Input for Chapter 4
0.6	31.10.2014	Satu Innamaa	Final draft for discussion in WP4 meeting
0.7	10.11.2014	Satu Innamaa	Updated draft according to discussion in WP4 meeting
0.8	14.11.2014	Satu Innamaa	Revision based on comments from Sami and Helena (incl. removal of chapters on tools catalogue and application of data sharing framework)
0.9	27.11.2014	Satu Innamaa	Revision based on comments from Helena
1.0	8.12.2014	Sami Koskinen	Improvements to Chapter 2. Submitted to the EC
1.1	18.12.2015	Satu Innamaa	Revision based on work done during 2015
2.0	27.01.2016	Satu Innamaa, Sami Koskinen	Review version
2.1	13.12.2016	Satu Innamaa	Revision based on work done during 2016
2.2	16.1.2017	Satu Innamaa	Updates based on Activity in Dec 2016
2.3	20.1.2017	Satu Innamaa	Reviewed version made ready for language revision
2.4	23.1.2017	Satu Innamaa	Language revision and comments from Helena Gellerman
3.0	27.1.2017	Satu Innamaa	Submitted final version

Review:

Name	**Date**
Helena Gellerman	24.11.2014 v1.0
Sami Koskinen	15.1.2016 v2.0
Sami Koskinen Adrian Zlocki Helena Gellerman	18.1.2017 v3.0 19.1.2017 22.1.2017

Circulation:

Recipient	**Date of submission**
EC	8.12.2014 v1.0, 27.1.2016 v2.0, 27.1.2017 v3.0
FOT-Net Data consortium	10.11.2014 v1.0, 18.12.2015 v2.0, 27.1.2017 v3.0

Table of Contents

Table of Contents 4

Executive Summary 6

1 Introduction 7

1.1 FOT-Net Data Project 7

1.2 Purpose of the FOT Data Catalogue 7

2 State of the Art on Catalogue Services 9

2.1 Open and Big Data 9

2.2 Catalogue Services in Other Disciplines 10

2.2.1 General Observations 10
2.2.2 Polar Data Catalogue 10
2.2.3 DataFed 11
2.2.4 UK Data Archive 11
2.2.5 EUDAT 12

2.3 ITS Catalogues 13

2.3.1 RDE 13
2.3.2 NDW – National Data Warehouse for Traffic Information 13
2.3.3 DLR Clearing House of Transport Data 13
2.3.4 Mobility Data Marketplace 13
2.3.5 NEARCTIS 13

2.4 Conclusions on Existing Catalogues 14

3 Data Catalogue 15

3.1 Catalogue Principles 15

3.2 Catalogue Development Process 21

3.3 Catalogue Content 22

3.4 Future of the Data Catalogue 25

4 First Datasets for Re-Use 26

4.1 Datasets Identified as High Value 27

4.1.1 DRIVE C2X 27
4.1.2 euroFOT 28
4.1.3 FOTsis 29
4.1.4 TeleFOT 29
4.1.5 UDRIVE 30
4.1.6 SHRP2 31
4.1.7 Safety Pilot and Research Data Exchange 31
4.1.8 Australian Naturalistic Driving Study 32

4.2 Other Available Datasets 32

4.2.1 National FOTs 32
4.2.2 Other Datasets 33
4.3 Lessons Learned when Getting Entries to the Catalogue 35
5 Discussion 36
List of Figures 37
References 37

Executive Summary

Datasets collected in Field Operational Tests (FOTs) of vehicle information technology form a valuable resource for further research. An important method for promoting these datasets is cataloguing them. FOT-Net Data has compiled the first European data catalogue for driving studies and field trials, in tight co-operation with similar international efforts and the FOT community. The datasets included in the catalogue are available for further research work and described in reasonable detail for potential re-users. This deliverable summarises the work done on the Data Catalogue during the FOT-Net Data project.

The Data Catalogue work was based on the existing FOT-Net Wiki, which at the start of the FOT-Net Data project contained an FOT Catalogue section describing projects worldwide and a Tools Catalogue section describing available tools (see D4.2 for development of this). The FOT-Net Wiki has now been complemented with the separate *Data Catalogue*.

To begin with, the project identified potential high-priority and otherwise suitable field trial and naturalistic datasets for re-use, and contacted the relevant partners to discuss data-sharing possibilities and to get a catalogue entry to the Data Catalogue. We built templates and experimented with various wiki tools on creating e.g. forms and search capabilities for the catalogue.

The Data Catalogue was launched with a preliminary group of datasets submitted by project partners, and the operation of the catalogue was piloted. After a testing period, external organisations were invited to submit further entries, starting with personal contacts and later involving everyone listed in the FOT Catalogue. From now on the catalogue should be easily maintainable by the FOT community, as it is in wiki format. All included functionalities aim at ease of use; for example, a dataset entry can be inputted simply by filling in a form on a web site.

This deliverable covers the state of the art of existing catalogue services. It also summarises the work done for the development of the Data Catalogue (structure and content fields) during the FOT-Net Data project, and reviews briefly the first datasets entered into the catalogue that are now available for re-use.

1 Introduction

1.1 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 organises international workshops, publishes a series of newsletters, and promotes FESTA, a European handbook on FOT methodology.

FOT-Net Data was a Coordination and Support Action in the EU 7th Framework Programme for Research, submitted for the call FP7-ICT-2013-10. The name of the project stands for Field Operational Test Networking and Data Sharing Support. FOT-Net Data is a continuation of FOT-Net's activities. For the sake of continuity these activities are referred to in external communications simply as FOT-Net.

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

The project partners were 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 was Dr Sami Koskinen, VTT.

1.2 Purpose of the FOT Data Catalogue

While the FOT-Net Wiki already included a comprehensive catalogue of field trials and naturalistic driving studies carried out in recent years, FOT-Net Data compiled further details regarding available research datasets and tools related to them. This work provided more details on collected data and information on the possibilities and options to access and share data.

The FOT-Net Data project set up a new Data Catalogue and updated the existing Tools Catalogue (see FOT-Net Data D4.2 for details) to facilitate data re-use. These catalogues will support potential data re-users in identifying interesting, suitable datasets, as well as providing tools for their research. The target was to make the catalogues easy to use.

The Data Catalogue features datasets in reasonable detail from the perspectives of different impact assessment areas. A catalogue entry introduces, for example, basics of the study design and describes tested functionalities, groups of test subjects, measured variables and

processed summary data. The catalogue includes contact details and information on conditions for data access and sharing. It is also possible to attach sample data.

The purpose of this deliverable is to describe the work done on the Data Catalogue. The deliverable includes a review of the state of the art on catalogue services, and a description of the content and functionalities of the first version of the Data Catalogue.

2 State of the Art on Catalogue Services

2.1 Open and Big Data

Open data and big data are hot topics in ITS (intelligent transport systems) research today, as well as in other scientific disciplines as evidenced e.g. in international congresses like ITS World. Cities are opening up their data storages such as maps, public transport routes, timetables, and information on parking areas. Traffic status is being collected from mobile phone users, taxi fleets and road infrastructure. The data is then made available to application developers, road and transport operators, data scientists and others.

Open data refers to the idea that certain data should be freely available to everyone to use as they wish. Especially for datasets collected with public funding, this is seen as a reasonable requirement. However, considering the nature of traffic data, such as movement of individuals, opening ITS datasets is restricted by privacy laws. For example, a dataset about average travel times between two camera measurement points can become open data, whereas the licence plate data behind these calculations cannot be made public. The licence plate data may still be available under a contract. FOT datasets and catalogues face similar questions on what parts of the data can become public.

Big data is defined in Oxford English Dictionary (2014) as computing data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges. The term is also used for the branch of computing involving such data. The term covers all areas of application and does not therefore reveal what type of data one is dealing with. Many transport datasets like those collected in large-scale field tests can be regarded as big data. However, it is usually clearer to specify for an audience that the discussion or presentation is about e.g. FOT or certain specific transport datasets rather than ‘open big data’. Some people might also understand ‘big data’ within ITS e.g. as probe vehicle data, covering thousands of vehicles, being logged continuously but usually providing a few signals only. Data from FOTs is usually more complex and from a limited time period, but FOT datasets can still be larger in size, as probe vehicles do not often collect video.

From governmental perspectives, ITS are just another discipline collecting large datasets. In a Finnish roadmap on research infrastructures (Suomen Akatemia 2014), the social sciences, energy, environment, health, materials, biology and mathematics had in total more than 30 national-level research infrastructures to maintain and 25 European partnerships.

The difficulties, tools and financial models needed to upkeep large datasets seem actually rather similar, when comparing the status between disciplines. Existing data management practices from other topic areas provide helpful tips and good practices on how to manage transport data.

From a European perspective, in 2012–2013, the Commission spent €45 million on data infrastructures. This funding targeted all research areas and did not usually consider project-level datasets, but rather levels of single disciplines or support for many. The funding for data infrastructures has been growing, notably in relation to the Horizon 2020 (H2020) programme, with just the first call for proposals on e-infrastructures totalling €55 million. (EC 2012)

Some of these H2020 e-infrastructure projects were tasked to store digital data being recorded across scientific disciplines. The Commission has launched an Open Research Data Pilot, where beneficiaries must deposit valuable datasets in a research data repository before the

projects end. These goals reflect the current view that all valuable datasets collected with public funds should be made available to a wider community than just those who collected the data.

ITS projects have not, as of this writing, faced rigid requirements to submit datasets to a 'data library'. They are, however, required to present plans for sharing and exploiting collected data. Long-term preservation of data is also an important feature in the evaluation of project proposals and its importance is growing in H2020.

FOT-Net Data is among the first ITS projects to specifically target data sharing, but more H2020 projects are likely to join the work.

The APARSEN project report summarises recent developments by recommending that digital preservation should be considered as a core business activity and should thus be embedded into an organisation's overall policies (APARSEN 2014). Basically, organisations receiving public funding will have to consider changes in their data management practices and some cooperation regarding data.

The e-IRG (e-Infrastructure Reflection Group, e-IRG 2013) considered in their white paper that agencies running large research facilities should be encouraged to provide a data archive and to open it to the relevant research communities, after a proprietary period allowing the original partners to benefit the most. They also suggested that data publication and sharing should be fully recognised as scientific activities, and that criteria for evaluating the work (quality, impact, etc.) should be established.

The e-IRG also considered that the current uptake of Open Science by the research community is still weak. However, they noted that astronomy, which has long used open data, shows that this changes the way science is done and offers equal opportunities to scientists. (Wiebelitz et al. 2013)

2.2 Catalogue Services in Other Disciplines

2.2.1 General Observations

Catalogue services usually begin within their user communities, rather than as universal data libraries. Various catalogues and community-specific metadata definitions exist, some interesting examples of which are given here.

From the perspective of analysing FOTs and being accustomed to read at least 20 pages of documentation to get an understanding of a field trial, the metadata in many catalogues of other disciplines seem to be brief, or rather well-structured and concise. When the data is of a defined type or in standard format, instead of having been recorded by a research project exploring new technologies, it seems easier to document.

2.2.2 Polar Data Catalogue

As an example of well-structured metadata, the metadata in the Polar Data Catalogue (www.polardata.ca) covers

- Title
- Study site
- Research program
- Example citation

- Authors
- Contact person
- Short abstract
- Keywords
- Time
- Coordinates.

The metadata is available in ISO 19115 standard format to enable exchange with other data centres. The catalogue covers different types of datasets generated by Arctic and Antarctic researchers, from ice core data (PolarData 2014a) to social studies (PolarData 2014b) targeting the regions.

At a glance, brief metadata can pretty well explain e.g. ice core data, given that there is a specific practice on how ice drilling is executed. Regarding specific social studies, however, the researchers would probably need to be contacted to learn more. The social dataset in the Polar Data Catalogue was catalogued but unpublished. The ice core data could be downloaded as an Excel file and was supported with a descriptive document containing definitions and scanned handwritten field notes.

The ISO standard 19115-1:2014 “defines the schema required for describing geographic information and services by means of metadata. It provides information about the identification, the extent, the quality, the spatial and temporal aspects, the content, the spatial reference, the portrayal, distribution, and other properties of digital geographic data and services” (ISO 2014). It would also provide the means to define certain transport data in catalogues based on geographic information.

2.2.3 DataFed

The catalogue service for air quality data by DataFed (http://webapps.datafed.net/aq_ufind.aspx) is an example of effective search capabilities based on well-defined metadata. Air quality datasets can be queried by specifying coordinates and time range. Data can also be accessed via a web service and Internet protocols. It includes viewers for time-series data.

DataFed is a community-supported effort, initially supported by information technology grants from the National Science Foundation and NASA.

In a DataFed publication (Husar & Poirot 2005), it is noted that traditionally air quality analysis was a slow, investigative process that took place months after data had been collected. Internet and real-time pollution detection have changed all that. Analysts can now work in virtual work-groups to share their observations and insights.

2.2.4 UK Data Archive

The UK Data Archive is funded by the (UK’s) Economic and Social Research Council. It is a large collection of research data in the social sciences and humanities. The catalogue service is called Discover (http://discover.ukdataservice.ac.uk/). Searching by ‘traffic’ and ‘transport’ produces various hits, mainly road accident and survey data.

The UK Data Archive provides an XML description of their metadata according to the Data Documentation Initiative’s (DDI, www.ddialliance.org) schema. The DDI attempts to create an international standard for describing data from the social, behavioural and economic sciences.

As a persistent identifier for a dataset, the UK Data Archive uses Digital Object Identifiers (DOIs). Each time there is a change in the data, a new DOI will be assigned and a log of changes will be collected. The DOI system simplifies citation. An example DOI is shown in Figure 1.

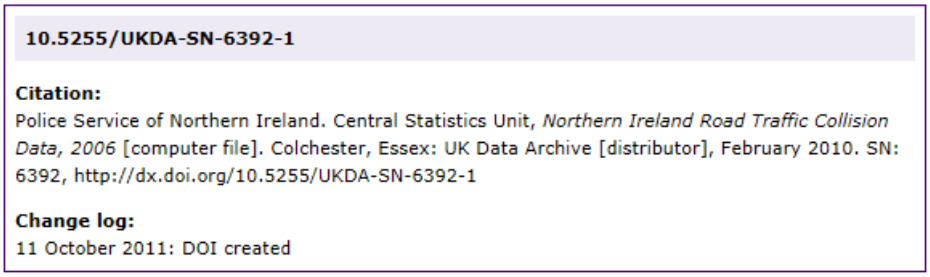


Figure 1. Example DOI from the UK Data Service.

The DOI system is an ISO standard 26324. To get a DOI for a dataset, one must use a service offered by a DOI Registration Agency [http://www.doi.org/faq.html]. One such Agency is the Publications Office of the European Union.

The UK Data Archive also provides guides on how to deposit data. They highlight that data archiving ensures safe-keeping of data in the long term and enables tracking of usage. It also allows data owners to avoid administrative tasks related to data sharing and answering queries.

Much of the academic data comes thanks to the Economic and Social Research Council (ESRC) policy, which states that all award holders are expected to offer data to the UK Data Archive.

2.2.5 EUDAT

The EUDAT (European Data Infrastructure) project has put up a B2FIND metadata service (b2find.eudat.eu) allowing users to search data with a keyword. The metadata is harvested from project communities representing different disciplines. The service translates community metadata schemas to be used in B2FIND.

The project also provides a B2SHARE tool (b2share.eudat.eu) for store and share research data. B2SHARE offers methods to upload and store small scientific datasets into the B2SHARE repository. The success of this sharing is based on mutual trust. (EUDAT 2014)

EUDAT works in collaboration with several communities, currently eight from different research areas. They explore generic technical services that can support multiple research communities. (EUDAT 2014)

FOT-Net Data has collaborated with EUDAT since 2014, exchanging speakers and information. EUDAT also provides the possibility to exchange with other data communities e.g. on personal data and financial models.

2.3 ITS Catalogues

ITS data catalogues are not yet as well established as their counterparts in e.g. the social sciences. However, there are prominent examples, with more being put up or extended.

2.3.1 RDE

The Research Data Exchange (RDE, www.its-rde.net) 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 anonymised, also with the help of dedicated software tools.

RDE metadata uses ASTM 2468-05 standard format. This "Standard Practice for Metadata to Support Archived Data Management Systems" is applicable to various types of operational data collected by intelligent transportation systems (ITS) and stored in an archived data management system. (ASTM 2014)

FOT-Net Data has been collaborating with USDOT about practices and documentation used in RDE, as it is the closest counterpart to our activities in Europe.

2.3.2 NDW – National Data Warehouse for Traffic Information

In the Netherlands, 19 public authorities are working together on collecting and distributing traffic data. The database consists of real-time data from 24 000 measurement sites, status information regarding e.g. road works and bridges open/closed, and historical data. NDW exchanges data e.g. using DATEX II format. (NDW 2015)

2.3.3 DLR Clearing House of Transport Data

DLR, in collaboration with the German transport ministry, has set up a clearing house of transport data (DLR Data Repository 2015). They note that many relevant transport datasets are known only to a small group of users. They provide a service for scientists and traffic planners and policymakers as well. The features datasets are diverse: statistics, empirical studies and mobility research.

Data can be obtained by placing an advance order. Following an agreement with the owner of the data, data is provided via regular mail. (DLR Clearing House 2015)

2.3.4 Mobility Data Marketplace

The Mobility Data Marketplace (MDM) is managed by the Federal Highway Research Institute (BASt). With an internet service platform, it enables offering and searching and subscribing to online traffic data. The marketplace concept aims to simplify business processes. The technical implementation started in 2010. (MDM 2015)

2.3.5 NEARCTIS

The NEARCTIS Network of Excellence (2008–2009) put together an inventory of available resources within their member organisations: http://www.nearctis.org/home/resources-desk/shareable-resources-wiki/. They identified 15 datasets for sharing; mainly loop data, probe vehicle data and camera data. (NEARCTIS D5 2009)

2.4 Conclusions on Existing Catalogues

Existing catalogues and standards on metadata define many basic fields for datasets, such as how to mark an author, date, persistent identifiers and coordinates.

Generally, the practices used in different scientific communities offer good input. Compared to most communities currently sharing data, FOTs are dealing with datasets that contain more personally identifiable information, such as videos. To understand FOT datasets, also a rather extensive set of metadata and supporting documentation is required.

Large open data catalogues like that the EUDAT project is building also contain metadata fields specific to certain scientific disciplines. These fields are extensions to the basic set of information. It seems possible as well as necessary to include such new extensions to better describe and search complex FOT datasets.

The process of integrating a new discipline in general-purpose catalogues starts with the discipline of identifying and creating a consensus on what is important metadata to describe their datasets. After a consensus on the metadata fields is formed, existing open data tools and practices can provide support through the next steps. Also, much work is needed to establish the data-sharing practices and requirements and basically the data sharing culture for a scientific discipline.

3 Data Catalogue

FOT-Net has compiled a Data Catalogue that describes FOT and NDS datasets. The datasets included in the catalogue were selected such that they are available for further research work under defined conditions and that they are described in reasonable detail (as metadata) for potential re-users. The project has provided this first European data catalogue for driving studies and field trials of ITS, in cooperation with similar international catalogue efforts such as the RDE in the US. FOT-Net came up with a format and definitions for the data catalogue by considering current best practices for FOT dataset documentation and existing catalogue standards. Feedback was gathered from the FOT-Net community, especially from dataset owners.

The work was based on the previous version of FOT-Net Wiki, which already contained a FOT Catalogue section describing projects around the world, and a Tools Catalogue listing available tools. The catalogues were moderated by FOT-Net, but as wiki pages they receive input and updates from the FOT community, various organisations carrying out large-scale trials.

3.1 Catalogue Principles

At present, FOT-Net Wiki includes three catalogues: the FOT Catalogue, the Data Catalogue and the Tools Catalogue. The FOT Catalogue is the main catalogue that aims to include all the FOTs/NDSs done in Europe and the most interesting ones internationally. Those FOTs that are willing to share their data and/or tools provide details on those also in the Data Catalogue and the Tools Catalogue.

The main purpose of the Data Catalogue is to support potential data re-users in identifying suitable datasets for their purposes and to facilitate data sharing. The development work for the catalogue was based on three main principles:

1. It is an extension of the previous FOT/NDS Catalogue (wiki)
2. Ease of use
3. Data remains with its owners.

The FOT Data Catalogue works as an extension of the previous wiki on FOTs/NDSs. The new Data Catalogue was added to the wiki's main page in line with the FOT and Tool Catalogues (Figure 2). The wiki functionalities were enhanced also for the existing catalogues to be fully in line with each other.

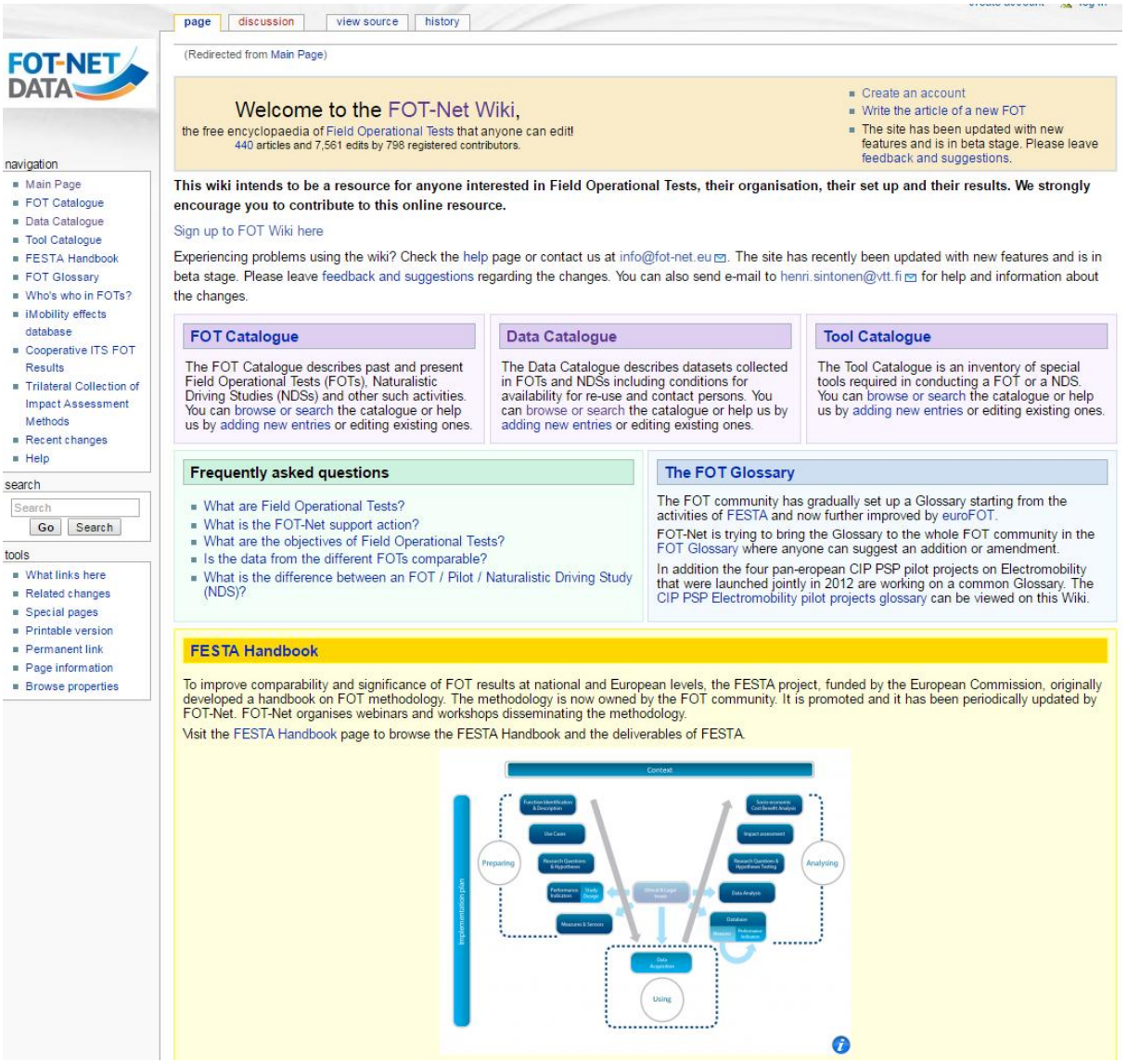


Figure 2. FOT-Net Wiki main page with links to the three catalogues (purple background).

Ease of use was a key principle in development of the catalogue. For someone willing to share data this included ease of information provision and update:

- Clear instructions
- Forms with which information can be submitted to catalogues
- Easy update functionalities (with the same form as submission).

For someone willing to re-use data, ease of use included essential information on datasets and their re-use conditions, as well as ease of search among datasets in the catalogues:

- Enough information (and datasets)
- Easy to make intelligent searches
- Easy to download information.

In practice, to facilitate the ease-of-use principle, new functionalities were added to the catalogues. MediaWiki was chosen as the engine as it is popular, used e.g. by Wikipedia. Popularity ensures a variety of add-on components, extending the normal editing possibilities of a wiki.

For a newcomer, the wiki syntax and how to create new pages can take a few hours to learn. Efficient guidelines and wiki templates can help, but to a limit. When only wishing to add information about a FOT, any learning process can be discouraging. For this reason, FOT-Net Data experimented and developed wiki forms to enter new data or edit the current information about a trial (Figure 3) without the need to utilise wiki syntax.

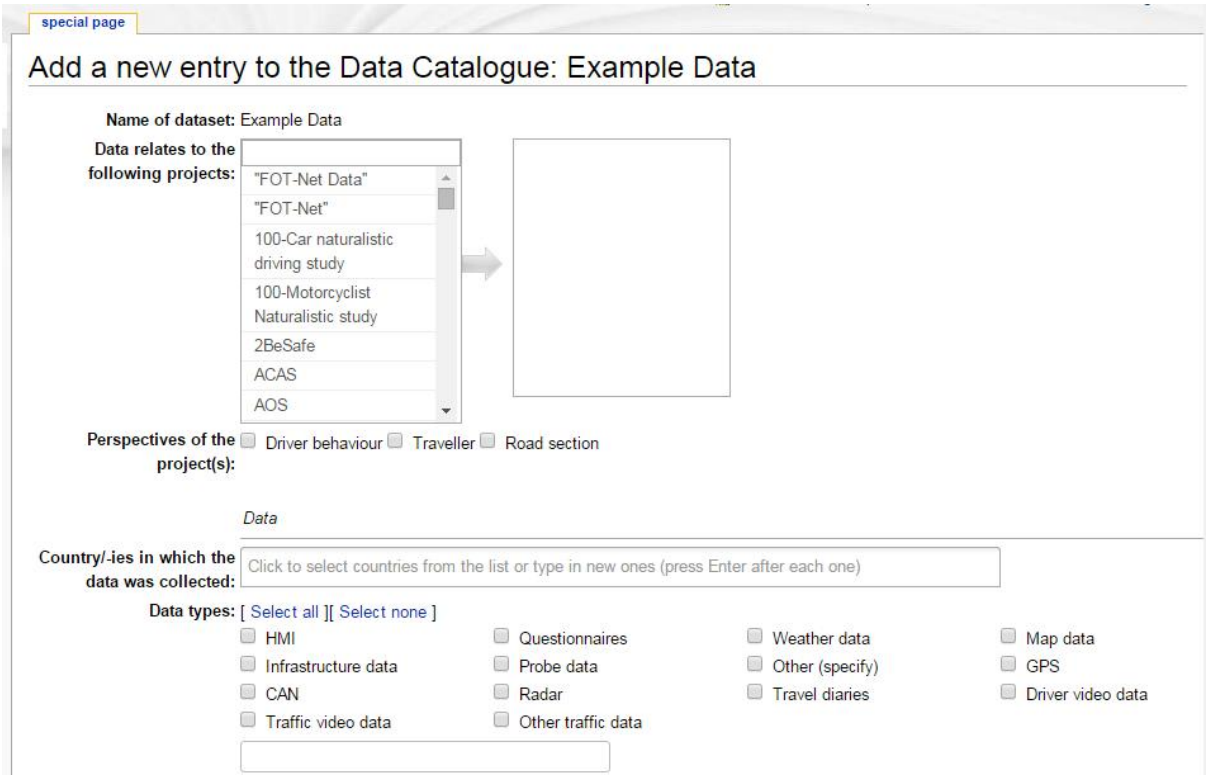


Figure 3. Part of the form with which an entry can be made to the Data Catalogue.

To facilitate re-use, search (Figure 4) and browsing (Figure 5) engines were also set up. To enhance their performance, predefined attributes were used for as many catalogue objects as possible (instead of free text entries used previously in the wiki). Automatically updating listings (Figure 6) were set up to ensure that all new contents can be seen by wiki-users, to avoid manual work and to facilitate easy maintenance.

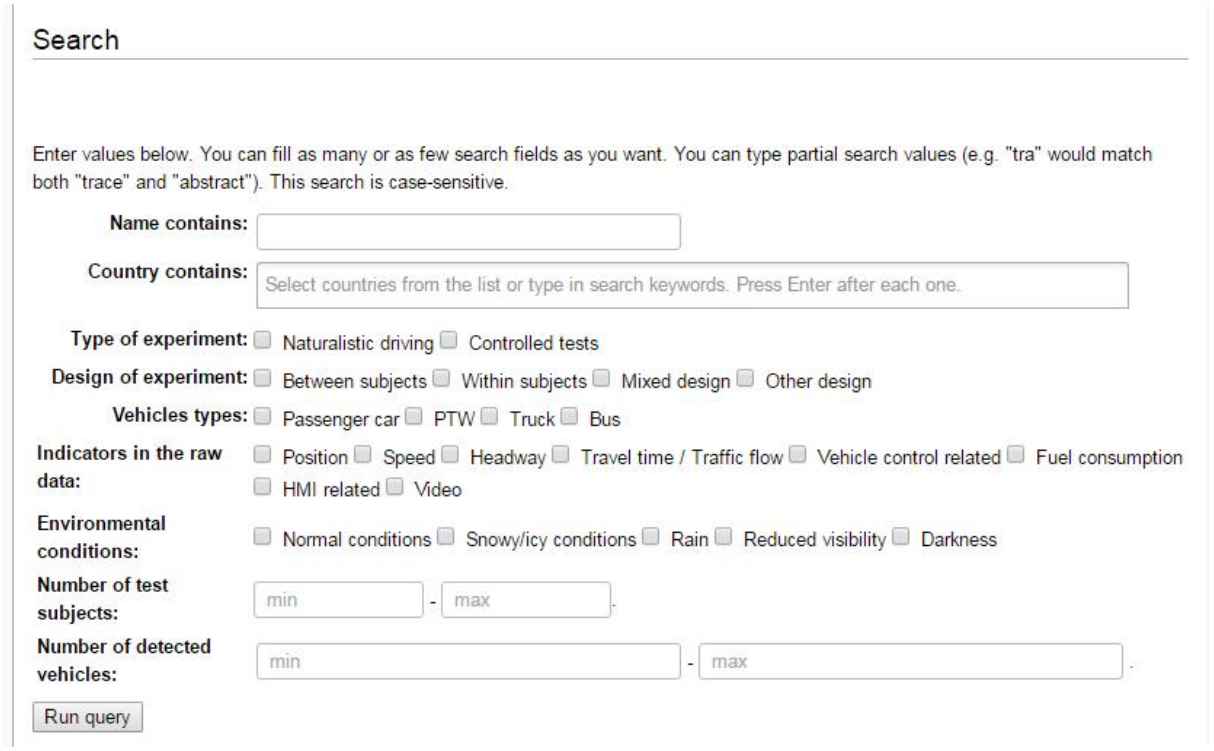


Figure 4. Search function in the catalogues.

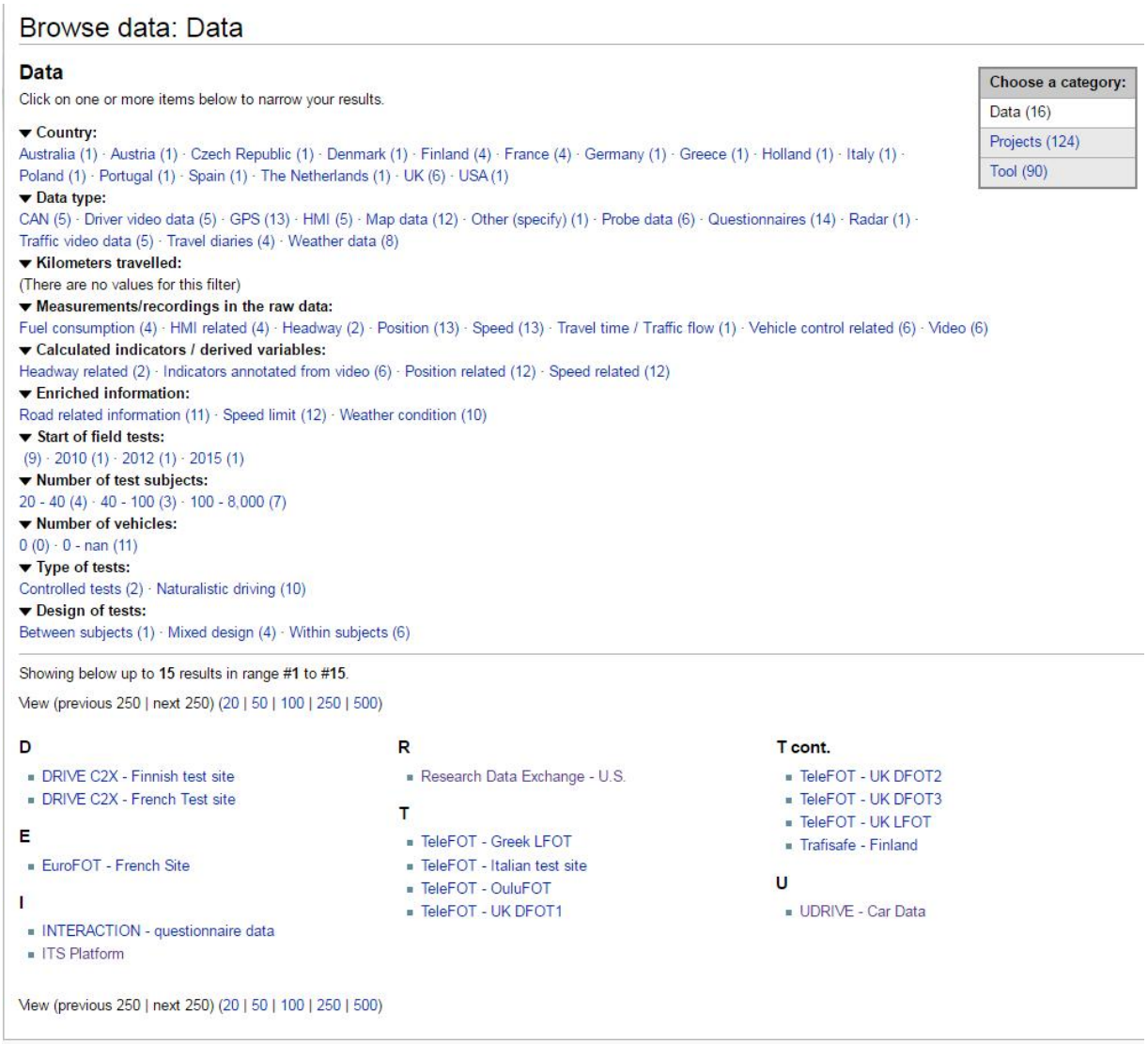


Figure 5. Browse function in the catalogues.

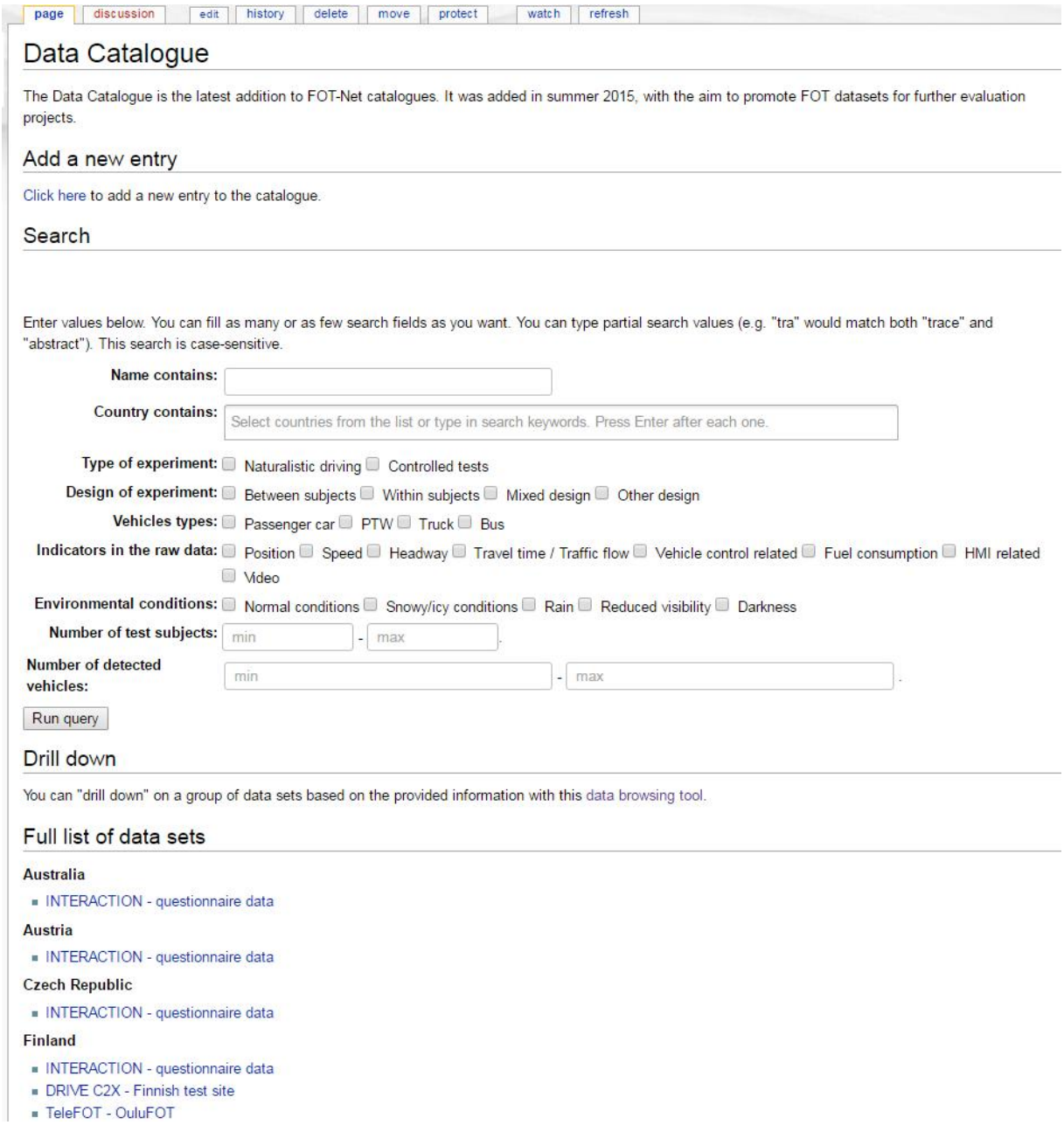


Figure 6. Data Catalogue front page with automatically updating full list of activities at the bottom.

Another main principle for the catalogue development was that the actual datasets remain with their owners and the FOT-Net Data Catalogue is not a data repository. Data providers will make the final agreements with interested organisations and can offer support for new analysts regarding the details of the study. The catalogue includes information on data and contacts but not the data itself. It is, however, possible to add anonymised sample data for allowing re-users to get a practical example.

3.2 Catalogue Development Process

The process in developing the Data Catalogue was the following:

1. Definition of desired catalogue functionalities
2. Enhancement of (FOT-Net FOT/NDS Catalogue) wiki tool with additional software packages to enable the desired functionalities
3. Definition of catalogue contents
4. Deployment of the catalogue (Version 0.1)
5. Pilot on the catalogue content with two datasets owned by FOT-Net Data partners (First pilot)
6. Update on the catalogue (Version 0.2)
7. Pilot on the catalogue content and functionalities with more datasets owned by FOT-Net Data partners (Second pilot)
8. Update on the catalogue (Version 1.0)
9. Public launch.

A summary of the process is presented with a timeline in Figure 7. In practice, the public launch of the catalogue meant that the FOT-Net community was invited to provide information on available datasets to the catalogue. At the same time, they were asked to provide information on FOTs/NDSs to the wiki or update the information currently available. FOT-Net Data continued building and promoting the catalogue throughout the project in all workshops organised by the project.

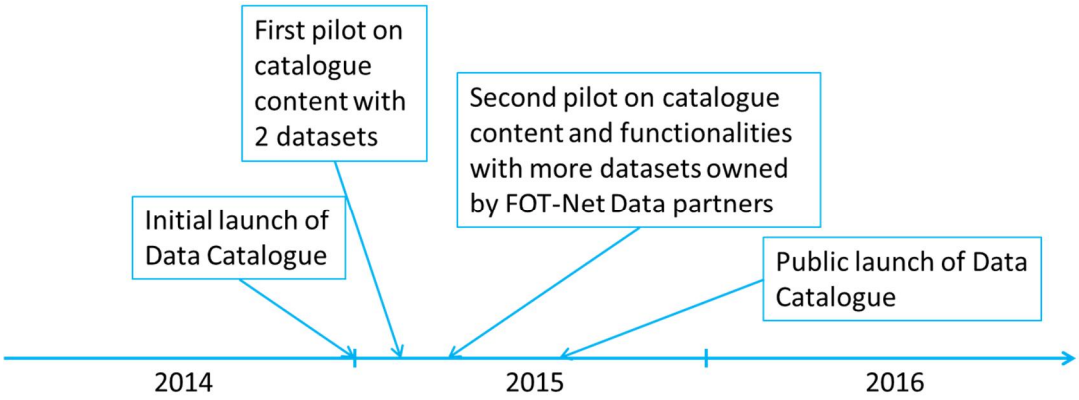
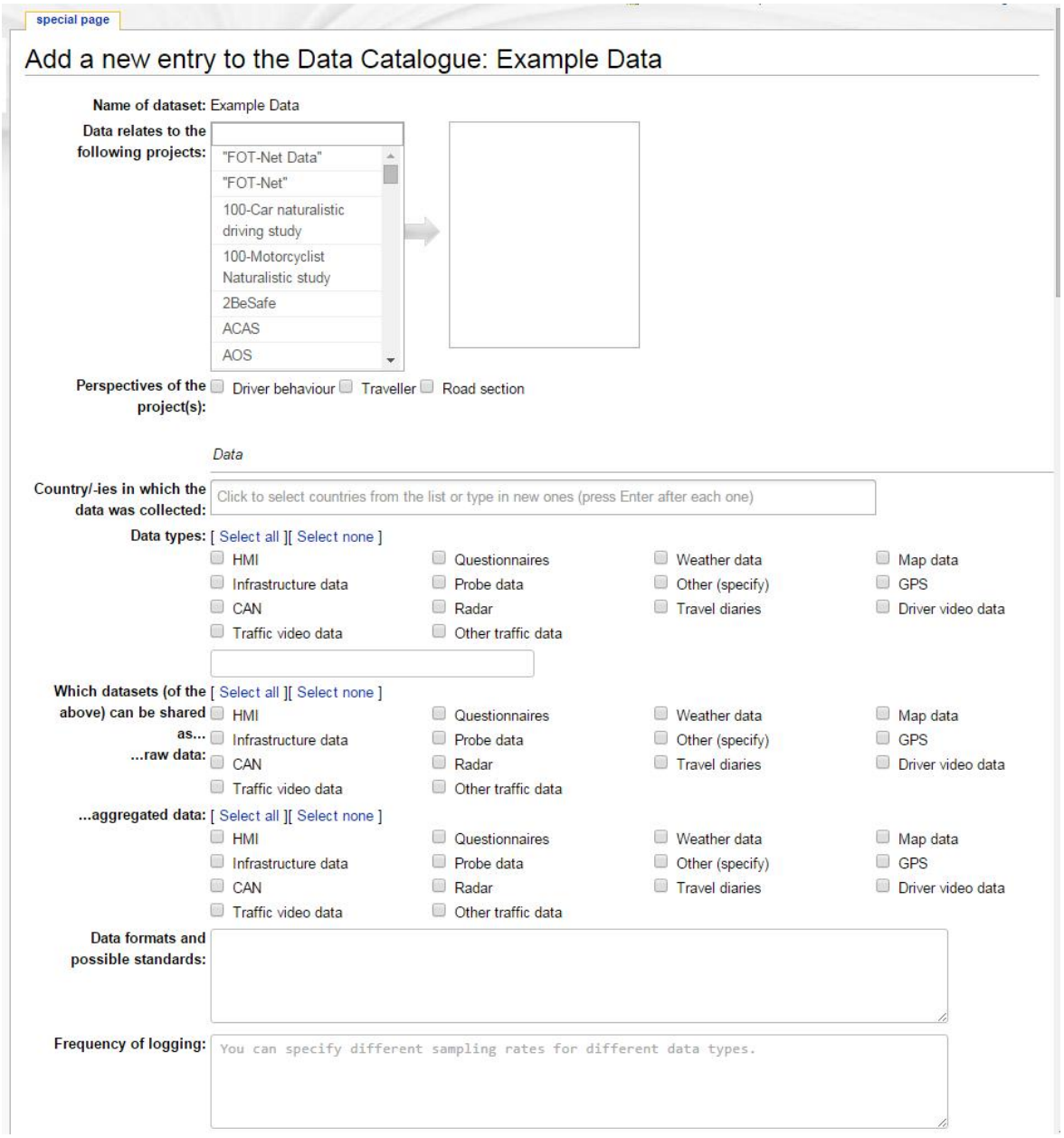


Figure 7. FOT Data Catalogue development process

3.3 Catalogue Content

The main principle in defining the catalogue content has been to ensure that it is in line with the terminology and content of the metadata and data descriptions, and with the data model suggested by FOT-Net's Data Sharing Framework. The Data Catalogue is targeting information on a high level and can be seen as a summary needed to get an idea on the available data and provide input for smart browse and search functionalities (summary of metadata). The following screen shots show the content of the Data Catalogue fields when adding a new entry using a form.



Quantity of the data

Kilometres driven / travelled: km

Number of journeys:

Number of events:

Hours of logging: hours

Number of detected vehicles / travellers:

Main indicators

Measurements/recordings in the raw data: Position Speed Headway Travel time / Traffic flow Vehicle control related Fuel consumption HMI related V

Calculated indicators / derived variables: Speed related Position related Headway related Traffic flow related Indicators annotated from video

Enriched information (map matching, etc.): Speed limit Weather condition Traffic volume Road related information

Quality of the data

Applied checks:

Limitations known:

Test subjects

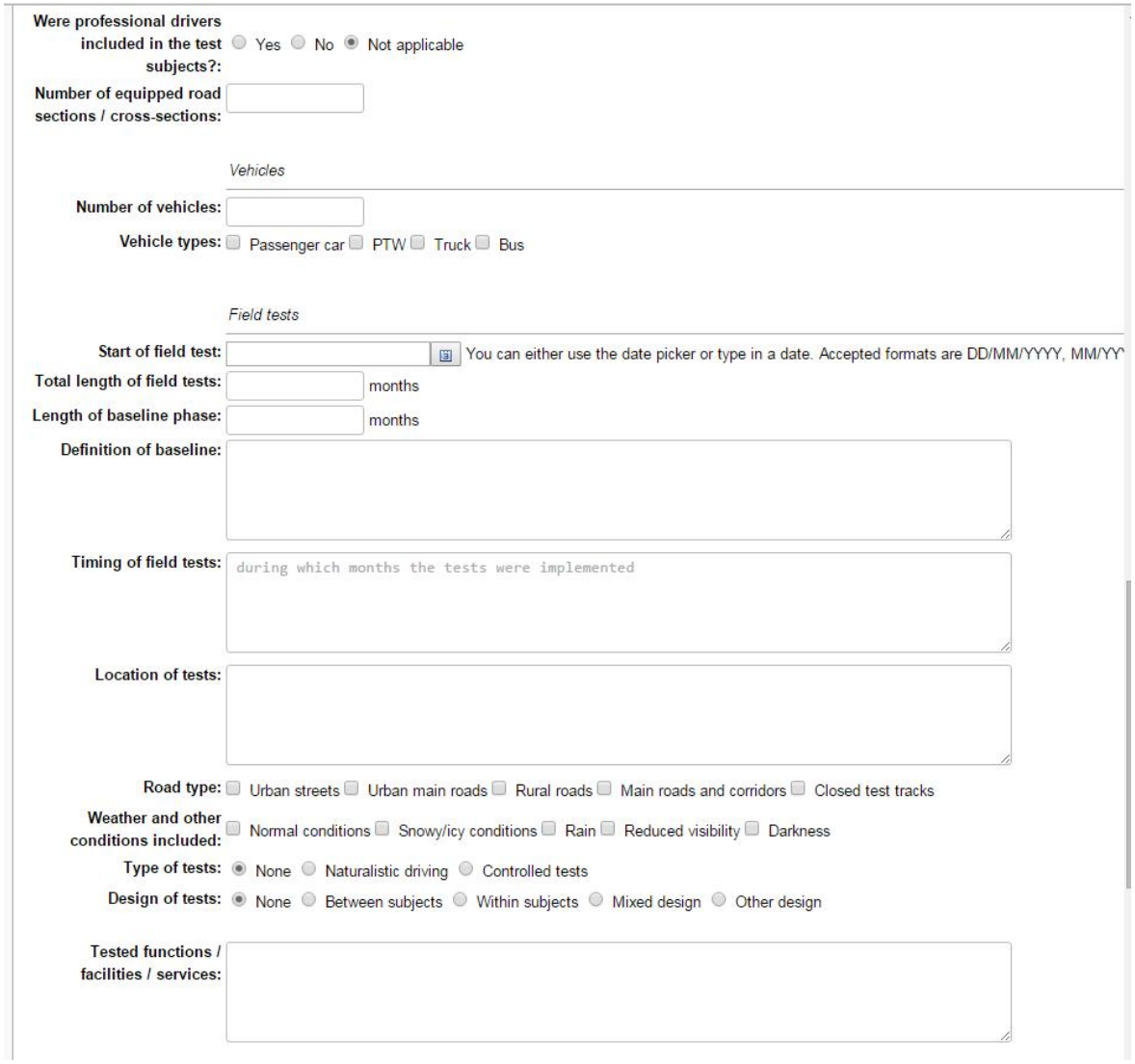
Number of subjects:

Number of male subjects:

Number of female subjects:

Age distribution:

Were professional drivers included in the test subjects?: Yes No Not applicable



Were professional drivers included in the test subjects?: ○ Yes ○ No ◉ Not applicable

Number of equipped road sections / cross-sections:

Vehicles

Number of vehicles:

Vehicle types: ☐ Passenger car ☐ PTW ☐ Truck ☐ Bus

Field tests

Start of field test: You can either use the date picker or type in a date. Accepted formats are DD/MM/YYYY, MM/YY

Total length of field tests: months

Length of baseline phase: months

Definition of baseline:

Timing of field tests: during which months the tests were implemented

Location of tests:

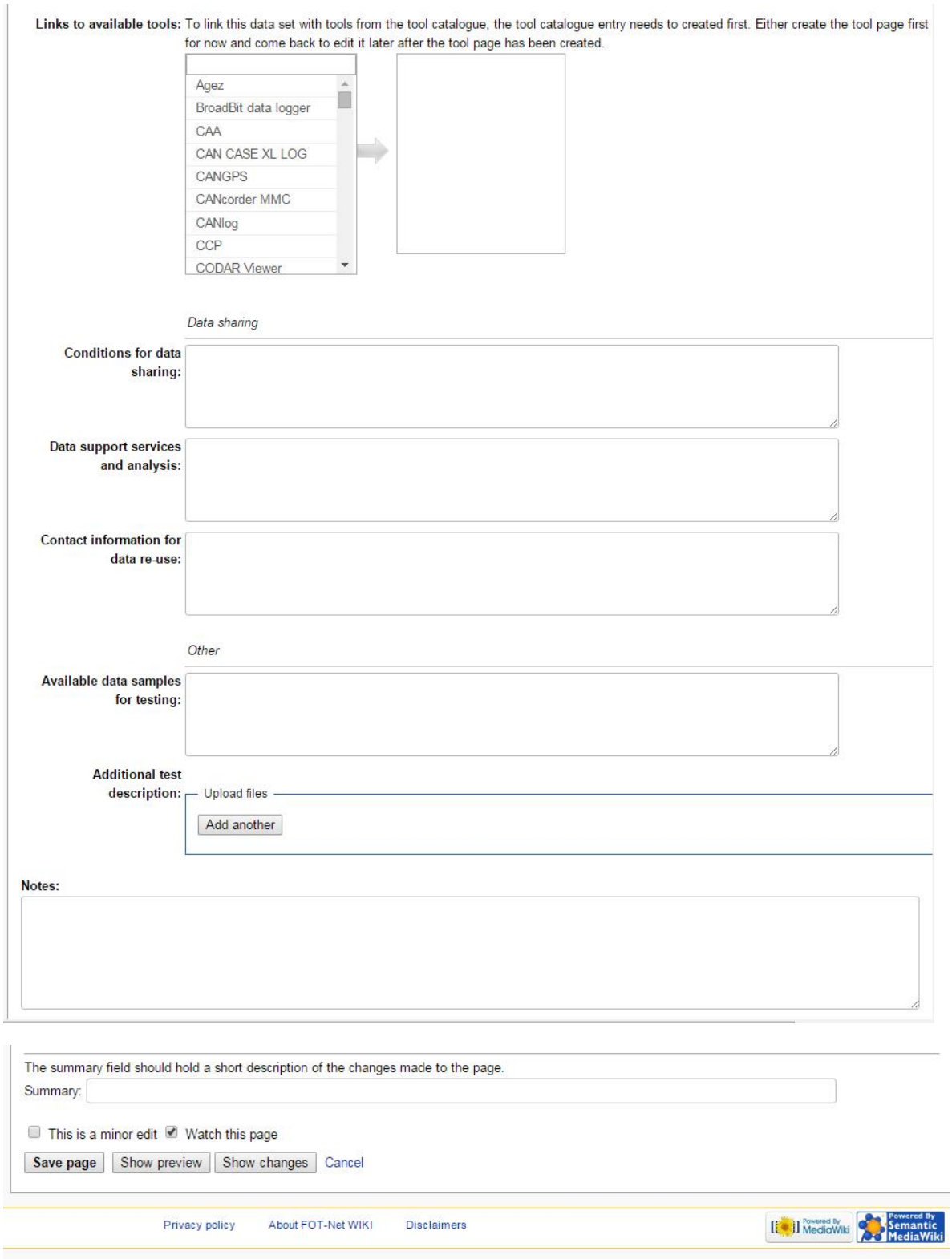
Road type: ☐ Urban streets ☐ Urban main roads ☐ Rural roads ☐ Main roads and corridors ☐ Closed test tracks

Weather and other conditions included: ☐ Normal conditions ☐ Snowy/icy conditions ☐ Rain ☐ Reduced visibility ☐ Darkness

Type of tests: ◉ None ○ Naturalistic driving ○ Controlled tests

Design of tests: ◉ None ○ Between subjects ○ Within subjects ○ Mixed design ○ Other design

Tested functions / facilities / services:



3.4 Future of the Data Catalogue

The FOT-Net community has been invited to provide information on available datasets. The functionalities to provide entries and update previously provided information have been made so easy that FOT/NDS data owners can now keep the catalogue updated even after the FOT-Net Data project has ended. An FAQ section has been provided with the catalogue to assist catalogue users.

4 First Datasets for Re-Use

During the first project year, FOT-Net identified several datasets that are of high value for re-use purposes. They were known for their quality and were both large and rich enough for various research purposes. The main datasets are introduced briefly in Chapter 4.1. FOT-Net held discussions with the owners of these datasets during the second project year, and with all other data owners listed in the FOT Catalogue during the third year. Other available datasets are detailed in Chapter 4.1.8. In total, there were 20 datasets[1] in addition to an entry about RDE (a Data Catalogue in the US) in the Data Catalogue at the writing this report:

- DRIVE C2X
    - Finnish Test Site
    - French Test Site
    - Swedish Test Site
- euroFOT
    - French Test Site
    - Swedish Test Site
- INTERACTION - questionnaire data from nine countries
- SeMiFOT
- 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 Center
    - Dutch Site (Trucks and cars)
    - French Site (Cars)
    - German Site (Cars)
    - Polish Site (Cars)
    - UK Sites (Cars)
    - Spanish site (PTWs)

[1] counting large-scale integrating projects (DRIVE C2X, euroFOT, TeleFOT, UDRIVE) as one dataset per test site, some of them including with several sub-studies; other projects were counted as one dataset per project

- ITS Platform – Denmark
- Trafisafe – Finland
- Research Data Exchange – US.

4.1 Datasets Identified as High Value

4.1.1 DRIVE C2X

The European Integrated Project DRIVE C2X, the follow-up to preparation project PRE-DRIVE C2X, took place in 2011–2014. The project built strongly on previously developed cooperative systems, which were considered to be mature enough for large-scale field operational tests. A consistent Europe-wide testing environment for cooperative systems was created, including seven test sites in Finland, France, Germany, Italy, the Netherlands, Spain and Sweden. The project assessed the impacts of cooperative systems on users, the environment and society. More than 750 drivers were able to test the cooperative functions.

Data from DRIVE C2X test sites was documented in detail to allow several analysts from different organisations in different countries to evaluate the data from all the test sites. This data sharing exercise within the project formed a good basis also for sharing the data with third parties. All drivers had been requested to sign an agreement releasing the data for scientific research, also for other organisations than the one that collected the data.

Finnish Test Site

The Finnish test site data from DRIVE C2X, owned by VTT, can be shared under an NDA. Since the dataset is reasonably well documented, VTT does not foresee a significant need to support third party analysts. An entry was made to the Data Catalogue.

The Finnish test data includes data related to the following cooperative functions in addition to the baseline data:

- Broken vehicle warning
- Warning on slippery road conditions
- Roadworks warning
- Speed limit information and speeding warning
- In-vehicle signage
    - Child sign
    - Yield sign
    - Stop sign
    - Pedestrian crossing ahead sign

Note that many of the cooperative functions on the Finnish test site were not based on vehicle-to-vehicle communication, but rather on communication between vehicles and infrastructure stations.

French Test Site

The subjective data from the French test site of DRIVE C2X (SCORE@F), owned by PSA, can be shared. This subjective data includes questionnaires. Data is related to the following cooperative functions:

- Cooperative in-vehicle signage / speed limit
- Roadworks warning
- Car breakdown warning

An entry was made to the Data Catalogue.

Swedish Test Site

The Swedish test site collected naturalistic data during one year from 20 cars, resulting in 10TB of data. The data set consists of the same data as euroFOT complemented with data from the cooperative systems; that is CAN data, video, map attributes, subjective data and cooperative data. The data is co-owned in part by Volvo Cars and SAFER/Chalmers. All data analysis need to be approved by the owners and also performed in cooperation with the owners. The analysis takes place on the premises of the project partners for privacy reasons and only project results can be extracted via a specific procedure.

An entry has been made to the Data Catalogue.

4.1.2 euroFOT

euroFOT developed and performed the field testing of new Intelligent Vehicle Systems having the potential to improve the quality of European road traffic. This permitted assessing the systems' effectiveness on actual roads, while determining how they perform towards the intended objectives. The field test offered early publicity of the technologies and enabled the analysis of user acceptance and its subsequent potential for market penetration.

Instead of publishing the data, the granting of access was based on a research proposal that was reviewed by owners of the data. Research proposals were addressed to the management board of euroFOT. The accessible content, after successful review of the proposal, represents the large body of data that has been analysed to derive the results obtained in euroFOT. An exception has been made not to provide personal data, such as video and GPS position, as this would require the acceptance of all FOT participants. The site of access is at the site of the owner of the database. Stored results were extracted after their content has been reviewed and checked against the research proposal. All costs for the access – like training session, operator, and consumables – were agreed upon before the analysis were started. The euroFOT project agreed that access to the data would end one year after the end of project, which concluded on June 30, 2013.

French Test site

No data sharing is foreseen for the French test site data of euroFOT. However, on research subjects in relation to euroFOT objectives (e.g. road safety analysis), CEESAR can perform analysis on the French euroFOT dataset. A relevant entry was made to the Data Catalogue.

Swedish Test Site

The Swedish test site collected naturalistic data from 100 cars and 80 trucks during one year and consists of 50TB of data. The data set consists of CAN data, video, eye-tracker data, map attributes and subjective data. The car data is co-owned in part by Volvo Cars and SAFER/Chalmers and the truck data is co-owned in part by Volvo Trucks and SAFER/Chalmers. All data analysis need to be approved by the owners and also performed in cooperation with the owners. The analysis takes place on the premises of the project partners for privacy reasons and only project results can be extracted via a specific procedure.

An entry has been made to the Data Catalogue.

Other datasets

The German test site dataset of euroFOT was available for re-use for one year after the end of the project and was therefore not added to the Data Catalogue.

4.1.3 FOTsis

FOTsis (European Field Operational Test on Safe, Intelligent and Sustainable Road Operation) was a large-scale field test of road infrastructure management systems. It concentrated on seven close-to-market cooperative I2V, V2I & I2I technologies (the FOTsis Services), in order to assess in detail both 1) their effectiveness and 2) their potential for full-scale deployment on European roads. Specifically, FOTsis tested the road infrastructure's capability to incorporate the new cooperative systems technology at nine test sites in four European test-communities (Spain, Portugal, Germany and Greece).

FOTsis reviewed the road infrastructure and communication networks required to secure proper connectivity from traffic control centres with users/vehicles. Relying on the common European and open ITS architecture guideline proposal, able to incorporate available and future ITS services and systems, FOTsis aimed to contribute to the safety, mobility and sustainability challenges faced nowadays by the European road transport system.

Negotiations were carried out on the possibilities of data re-use. FOT-Net proposed co-operation regarding data-sharing aspects for FOTsis near the end of their project. Unfortunately, no entries resulted from this discussion by the end of the FOT-Net Data project.

4.1.4 TeleFOT

TeleFOT was an integrated EU project covering the period 2008–2012. It carried out several FOTs for assessing the impacts of driver support functions provided by smartphones, navigators and other in-vehicle aftermarket devices on the driving task and driver behaviour. (Mononen et al. 2013)

Trials were carried out in eight European countries: Sweden, Finland, France, UK, Germany, Spain, Italy and Greece. In total, the project partners recruited more than 2,800 test users and collected data from more than 10 million kilometres. (Mononen et al. 2013)

Finnish LFOT (OuluFOT)

OuluFOT was a large-scale trial in Finland that tested a smartphone (Nokia Series 60) traffic information application and a green-driving adviser. The dataset was added to the Data Catalogue.

The data owner, VTT, considers that this dataset can be shared under a basic NDA but is no longer high-priority due to the complex experimental procedure and outdated phone applications tested. Understanding the experimental procedure requires extensive knowledge of the test, even though the steps were documented in a lot of detail. Furthermore, the tested functions (smartphone apps) age quickly, even within the same project if it is long lasting. Nonetheless, the baseline data can be useful and illustrates the general driving behaviour across Finland.

According to VTT, despite the extensive documentation, analysts would require support, which would mean contracting at least several days of work from VTT. On the positive side, various post-processed summary tables and analysis results are already available, giving new analysts

a quick start. Similar summary tables are available from all the TeleFOT test sites, enabling comparisons of e.g. driving style in different countries.

UK LFOT, UK DFOT1, UK DFOT2, UK DFOT3

Loughborough University, which owns TeleFOT's UK datasets, is willing to consider sharing these datasets to third parties depending on the intended use of the data. Entries were made to the Data Catalogue.

The large-scale FOT (LFOT) includes GPS data, questionnaire and travel diary datasets. The tested functionalities included navigation, speed information & speeding alert provided with nomadic device (BLOM). The detailed FOTs (DFOT) supplemented the datasets collected during the LFOT with e.g. rich video material.

The LFOT dataset covers 12 months of driving for 80 test users resulting in 300 000 logged kilometres.

Greek LFOT

ICCS, the owner of the Greek LFOT data, has agreed that the type of agreement for data provision will be assessed on a case-by-case basis between ICCS and the interested party. An entry was made to the Data Catalogue.

The tested functions included navigation, speed limit information and speed alert, and traffic information, all provided with a nomadic device. The data includes GPS data, questionnaire and travel diary datasets of 148 participants who drove more than 800 000 km during the 11-month test period.

Italian LFOT

UNIMORE, the owner of the Italian LFOT data, has agreed to provide their available data for research purposes only. An entry was made to the Data Catalogue.

The tested functionalities include navigation, speed limit information and speed alert. The data includes GPS data, questionnaire and travel diary datasets of 168 participants who drove more than 4 million km during the 12-month test period.

4.1.5 UDRIVE

UDRIVE is the first large-scale European Naturalistic Driving Study (NDS) on cars, trucks and powered two-wheelers. The acronym stands for "eUropean naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment". The UDRIVE project builds on the experiences of the PROLOGUE feasibility study and various FOTs (e.g. euroFOT), and aims to contribute to developing this in-depth knowledge on drivers´ normal driving behaviour by conducting the first large-scale European NDS.

Over a period of 18 months, UDRIVE collects naturalistic (meaning that the behaviour of road users is observed unobtrusively in a natural setting) data on passenger cars, trucks, and powered two-wheelers. All data – including video data showing the forward view of the vehicle and a view of the driver, as well as geographic information system (GIS) data – are collected continuously to bring knowledge in the various research areas well beyond the current state of the art.

Description of UDRIVE data was included in the Data Catalogue. The Catalogue includes entries regarding the data collected in France, the Netherlands, Germany, UK, Spain and

Poland. The data includes variables related to position, speed, headway, and fuel consumption as well as video. In addition, it includes information such as the speed limit, weather condition, and road attributes. The total duration of data collection is 18 months.

All collected data is stored in a common data storage, the central data centre at SAFER. The sharing of data after the project was a pre-requisite in the GA, if funding was available, and all agreements and procedures were developed with data sharing as a requirement. The project analysis is performed remotely on the dataset at SAFER.

Data sharing is possible if funding is provided to maintain and access the data. The data sharing set-up and procedures used in the project will be used also after the project. External remote access to non-personal data is possible for third parties. Personal data is only available through the UDRIVE partners (on their premises). It is not possible to acquire the dataset to be stored elsewhere. The data will be available in 2018, if funding is provided.

4.1.6 SHRP2

In 2005, the United States Congress created the second Strategic Highway Research Program (SHRP 2) to address the challenges of moving people and goods efficiently and safely on the nation's highways. SHRP2 was administered by the Transportation Research Board of the National Academies, under a Memorandum of Understanding with the Federal Highway Administration (US Department of Transportation) and the American Association of State Highway and Transportation Officials.

SHRP2 was a targeted, short-term research programme carried out through competitively awarded contracts to qualified researchers in the academic, private, and public sectors. SHRP 2 addressed four strategic focus areas: the role of human behaviour in highway safety; rapid renewal of aging highway infrastructure; congestion reduction through improved travel time reliability; and transportation planning that better integrates community, economic, and environmental considerations into new highway capacity.

A naturalistic driving study investigates ordinary driving under real-world conditions in order to make the driving experience safer. In the SHRP2 study, about 3000 volunteer drivers will agreed to have their cars fitted with cameras, radar, and other sensors to capture data as they went about their usual driving tasks.

SHRP2 data has been re-used widely thanks to active promotion and research contracts. An entry to the Data Catalogue was under discussion at the end of FOT-Net Data project and is likely to come.

4.1.7 Safety Pilot and Research Data Exchange

Connected Vehicle Safety Pilot was a research programme that demonstrated the readiness of DSRC-based connected vehicle safety applications for nationwide deployment. The vision of the Connected Vehicle Safety Pilot programme was to test connected vehicle safety applications in real-world driving scenarios in order to determine their effectiveness at reducing crashes and to ensure that the devices were safe and did not unnecessarily distract motorists or cause unintended consequences.

The Research Data Exchange (RDE) is a core element of the USDOT's Connected Data Systems Program. The RDE is intended to support research, analysis, application development, and testing. It hosts several datasets available for re-use. Data Catalogue includes an entry for RDE which hosts e.g. Safety Pilot data.

4.1.8 Australian Naturalistic Driving Study

The Australian Naturalistic Driving Study (ANDS) is an on-going project, involving 360 drivers in the age-span of 20–70 years old. Each vehicle is owned by the participating driver and is instrumented for four months. Around half of the drivers/vehicles are recruited from the Sydney region in New South Wales and half from the Melbourne region in Victoria. The project consortium consists of universities, road safety authorities, road user association and commercial partners.

The ANDS project uses the same data loggers as the SHRP2 project and therefore collects the same kind of vehicle and video data. It also collects data from MobilEye systems and eye-trackers from Seeing Machines.

No data catalogue entry has been made yet, as the project is still on-going.

4.2 Other Available Datasets

In addition to the datasets identified as having the highest value, contact was sought also for other datasets included in the FOT Catalogue, unless they represented also a test site (no dataset), demo (very small dataset) or data related to specific technical testing (very limited dataset). In total, these other datasets for which an attempt to contact was made totalled 158. Contact was initiated with the contact person mentioned in the FOT Catalogue entry. Some persons were linked to several datasets; thus emails were sent to 67 persons in all. Some contact information was outdated, and not everyone responded. Some datasets no longer existed or could not be shared for e.g. contract reasons. Some of those contacted were not interested in providing an entry. Overall, the success rate was low. Nonetheless, attempting contact was considered worth the effort as it involves personal dissemination and is a form of marketing to potential future data owners. Assistance and information were offered to all those contacted.

4.2.1 National FOTs

Trafisafe

Trafisafe was a Finnish field trial project in which a new driving style feedback system was tested with 75 novice drivers and their parents. The idea of providing information about the driving style of both the novice driver and their parents to each party was to help the novice discuss their driving style with a more experienced driver. Feedback was based on data from OBD (on-board diagnostics), GPS, and internal acceleration sensors of the data logger. The project was carried out in 2012–2014.

The project was funded by the Finnish Transport Safety Agency, and it has been agreed by the partners of the Trafisafe project that the data is suitable for re-use. The data is co-owned by the organisations that collected it. Private data was deleted upon completion of the project. However, the owners maintain an interest in controlling the dataset and any public reporting based on it to ensure against potential misuse or breach of good scientific practice.

VTT has created an NDA (non-disclosure agreement) under which the dataset can be accessed also for other scientific research. The data is included in the catalogue and has so far been shared with the University of Leeds.

ITS Platform

Over the last few years, extensive resources have been allocated to the ITS field in Denmark. One of the larger ITS initiatives was conducted in the North Denmark Region, aimed to be the Danish ITS region where new technologies are developed and tested before they are implemented throughout the country. The ITS Platform was the first large-scale project of its kind and the first step in realising the vision of an ITS Platform that was open and available for new applications. The ITS Platform had a total budget of €5 million and was implemented in 2010–2013.

If Intelligent Transport is to make the expected contribution to making the traffic of the future safer, more efficient and more eco-friendly, it requires both car-to-car and car-to-infrastructure communication. In order to achieve this objective, close collaboration between public authorities and businesses is necessary. On the one hand, equipment with both positioning and communication systems must be installed in vehicles, while on the other services must be developed so that road users are motivated to take advantage of the possibilities afforded by it.

Currently, two datasets have been identified for re-use: one car with all the data from a two-year period, and 425 cars with the data from a two-year period but as anonymised floating car data. These datasets are related to automatic parking payment-customised traffic information (beta version) and collected traffic statistics from road authorities (beta version). The length of the field tests was 32 months and the data totals 15 million logged kilometres. All road types and weather conditions are included. An entry was added to the Data Catalogue.

SeMiFOT

The SeMiFOT project in Sweden developed the Naturalistic FOT method into a powerful tool in traffic safety research in 2008–2009. The naturalistic method involved collecting data continuously from a suite of vehicle sensors in order to assess safety in the interactions of driver, vehicle and environment. Environment sensing and video were essential for identifying near-collisions and other incidents, and for validating that intelligent vehicle systems perform as expected. SeMiFOT developed own tools for the complete methodology chain (data acquisition-data storage-data analysis) needed to perform a Naturalistic FOT, to gain knowledge concerning how to achieve good quality data and research.

The SeMiFOT project included ca 6-month data collection including GPS, CAN, video and some external sensors, totaling 14 vehicles and 250,000 logged km. Non-partners of SeMiFOT may only work in projects together with SeMiFOT partners, where the SeMiFOT partners have access to the data and can provide analysis results to the new consortium. The project need to be approved by SAFER and the concerned OEM, all according to the original agreements. An entry was added to the Data Catalogue.

4.2.2 Other Datasets

INTERACTION

Understanding driver interactions with In-Vehicle Technologies (IVT) was the main objective of the INTERACTION project, which started in November 2008. The project lasted 42 months and involved 10 European partners from eight countries and two Australian institutes under the EC 7th Framework Programme (FP7).

Amongst all the available IVT, INTERACTION focused on a limited set of mature technologies, already available in a wide range of car models and already adopted by most European car

drivers, such as communication or navigation systems, speed control or distance control systems.

For this selected set of IVT, INTERACTION tackled the following questions: Why, when, where and how do drivers use IVT? What are the patterns of IVT use in everyday driving? What are the individual factors that explain, or do not, the adoption of IVT by drivers? What are the differences or similarities between countries and their reasons? What are the actual supports to the driving task given by the systems? Are there involuntary or voluntary misuses of systems? Can these systems induce unexpected unsafe behaviour and skills?

The questionnaire data of the INTERACTION project can be shared and entries were made to the Data catalogue for all nine countries.

4.3 Lessons Learned when Getting Entries to the Catalogue

Contacting FOT data owners took a lot of time and several contacts per dataset. To begin with, some original contacts for completed FOTs had moved on, requiring time and effort to find someone with whom to discuss data sharing.

Next, the data owners of completed FOTs did not know right away whether their dataset could be shared from e.g. a contractual point of view. For them this meant going back to the original contracts and contacting the relevant legal advisers on whether the data could be shared at all and, if so, under what conditions.

Finally, if there was a positive decision on data sharing, this required yet more effort to locate and examine the documentation before making an entry. Thankfully, as FOT-Net Data partners have been involved in several of the European FOTs, we had inside information regarding these datasets and were able to provide practical support in creating the entries.

The ability to make and edit entries using a form has proven successful. We offered the data owner the possibility to pre-fill the form, after which they then checked and complemented the information. With the form, this has been (technically) easy.

Projects which shared data from different test sites already within the project have an advantage over those where data was analysed locally, in that better metadata already exists.

Another lesson learned is that not all data owners are fully aware of the potential limitations to sharing data openly. In some cases, data owners can be too quick to share data, forgetting that a dataset can contain information that is considered at least indirectly to be personal. In general, test subjects have been assured in agreements that their data will be used for scientific research only. Certain control of access to the data or further anonymisation of the dataset may be required, depending on the case. The FOT-Net or other catalogue operator should ensure, where possible, that data sharing processes are followed correctly in each case.

The Data Sharing Framework’s guidelines should help scientists and provide a lot of information on maintaining test subjects’ privacy. However, there is still a clear need for training sessions for new scientists on data management and laws on personal data.

5 Discussion

Datasets collected in FOTs of vehicle information technology form a valuable resource for further research. An important method for promoting these datasets is cataloguing them. FOT-Net Data has compiled the first European Data Catalogue for driving studies and field trials, in tight cooperation with similar international efforts and the FOT community. The datasets included in the catalogue are available in different ways for further research and described in reasonable detail for potential re-users.

This deliverable was designed to describe the work done for the FOT-Net Data Catalogue. The deliverable includes a review of the state of the art on catalogue services and a description of the content and functionalities of the Data Catalogue. It also lists and briefly describes the first datasets entered into the catalogue, and some lessons learned during the contacting process.

The main purpose of the Data Catalogue is to support potential data re-users in identifying suitable datasets for their purposes and to facilitate data sharing. The development work for the catalogue was based on three main principles: (1) It is an extension of the previous FOT/NDS Catalogue (wiki); (2) Ease of use; and (3) Datasets remains with their owners.

The FOT Data Catalogue works as an extension to the previous FOT-Net Wiki on FOTs/NDSs. The new Data Catalogue was added to the wiki's main page in line with the FOT and Tool Catalogues. The wiki functionalities were enhanced for all three catalogues to be fully in line with each other. These functionalities include a form with which it is easy to create or update an entry in the catalogue, automatically updating lists of entries, and intelligent search and browse functionalities.

The project first identified potential high-priority and otherwise suitable field trial and naturalistic datasets for re-use, and contacted the relevant persons to discuss data sharing possibilities and for an entry to the Data Catalogue. Initially the catalogue entries were submitted by project partners and the operation of the catalogue was piloted. After a testing period, external organisations were invited to produce entries, starting with those we knew personally followed by all those listed in the FOT Catalogue. From now on, the catalogue should be easily maintainable by the FOT community, since it is built as a wiki. All functionalities included aimed at ease of use. A dataset entry can be inserted simply by filling in a form on a web site.

Lessons learned when contact with FOT data owners highlighted that unless data sharing is considered at the very start of a project (preparation), it is hard to achieve. There is also a clear demand for a FOT-Net Data Sharing Framework with regards to that not all data owners are fully aware of the limitations of data re-use and sharing.

List of Figures

Figure 1. Example DOI from the UK Data Service. 12

Figure 2. FOT-Net Wiki main page with links to the three catalogues (purple background). 16

Figure 3. Part of the form with which an entry can be made to the Data Catalogue. 17

Figure 4. Search function in the catalogues. 18

Figure 5. Browse function in the catalogues. 19

Figure 6. Data Catalogue front page with automatically updating full list of activities at the bottom. 20

Figure 7. FOT Data Catalogue development process 21

References

APARSEN (2014). Exemplar Good Governance Structures and Data Policies. APARSEN Project Deliverable. D35.1

ASTM (2014). http://www.astm.org/Standards/E2468.htm

EC 2012. Towards better access to scientific information: Boosting the benefits of public investments in research. European Commission. 2012. http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf

EUDAT (2014). Project website www.eudat.eu, including e.g. eudat.eu/eudat-communities, entry October 30 2014.

Husar, R. and Poirot, R. (2005). DataFed and FastNet: Tools for agile air quality analysis. http://datafedwiki.wustl.edu/images/c/c4/EM_DataFed_FASTNET_050720.pdf

ISO (2014). Geographic information, Metadata, Part 1: Fundamentals. http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=53798. Entry 9.10.2014

Mononen, P., Franzen, S., Pagle, K., Morris, A., Innamaa, S., Karlsson, M., Touliou, K., Montanari, R., Fruttaldo, S. (2013). TeleFOT Final Report. D1.15, TeleFOT, European Commission.

Oxford English Dictionary (2014). Entry 9.10.2014.

PolarData (2014a). Polardata Catalogue, v.1.0.2, Metadata Record, Data: Ice core data from ice patches in the Mackenzie Mountains, Northwest Territories, 2007-2009. https://www.polardata.ca/pdclite/viewmetadata.ccin?id=2985, Entry 9.10.2014

PolarData (2014a). Polardata Catalogue, v.1.0.2, Metadata Record, Data: Negotiating Change: Community-based Mental Health and Addictions Practice in the Northwest Territories. https://www.polardata.ca/pdclite/viewmetadata.ccin?id=5082, Entry 9.10.2014

Suomen Akatemia (2014). Suomen tutkimusinfrastruktuurien strategia ja tiekartta 2014-2020. Tutkimusinfrastruktuurien asiantuntijaryhmä, Suomen Akatemia, Helsinki. 80 p. http://www.aka.fi/Tiedostot/Tiekartta/Tutkimusinfrastruktuurien_strategia_ja_tiekartta_2014_20.pdf

e-IRG (2013). e-IRG White Paper 2013, e-Infrastructure Reflection Group. http://media.icordi.eu/Repository/document/Publications%20and%20Reports/e-irg_white_paper_2013_-_final_version.pdf]

MDM (2015). http://www.mdm-portal.de/fileadmin/user_upload/Dokumente/MDM-brochure-en.pdf

DLR Data Repository (2015). http://daten.clearingstelle-verkehr.de

DLR Clearing House (2015). http://www.dlr.de/cs/en/desktopdefault.aspx/1177_read-2160/

NEARCTIS D5 (2009). NEARCTIS_Deliverable5_final.pdf at http://www.nearctis.org/home/resources-desk/deliverables/

NDW (2015). http://www.ndw.nu/en/