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

Interim Report on FOT Data Catalogue

Deliverable no.	Interim version of 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	2.0
Date	27 January 2016
Project Start Date and Duration	January 2014, 36 months

www.fot-net.eu

Document Control Sheet

Main author(s): Satu Innamaa and Sami Koskinen (VTT)

Work area: WP4

Document title: Interim Report on FOT 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	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 in 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

Review:

Name	Date
Helena Gellerman	24.11.2014

Circulation:

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

Table of Contents

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 is compiling 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 will be available for further research work and described in reasonable detail for potential re-users.

The work is based on the FOT-Net Wiki, which contains a FOT Catalogue section describing projects across the world. The list of FOTs has been complemented with a *FOT Data Catalogue*.

This second version of the interim report summarises the progress made on the Data Catalogue during the first and second project year of FOT-Net Data. The content of this report is also a draft on upcoming deliverable D4.1 Data Catalogue.

The project has identified first potential high-priority and otherwise suitable field trial and naturalistic datasets for re-use. We have been building templates and experimenting with helpful wiki tools such as forms and search capabilities for the FOT Data Catalogue. The catalogue was launched with first datasets described – and the project continues building it. At first, the catalogue entries were submitted by project partners and the operation of the catalogue was piloted. After a testing period we invited external organisations to produce entries. Eventually the catalogue should be easily maintainable by the FOT community, since it is built as a wiki. All functionalities included aim at ease-of-use. A dataset entry can be inputted simply by filling in a form on a web site.

This report covers state of the art of existing catalogue services. It also presents the current structure and content fields of the data catalogue.

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

FOT-Net Data is a Coordination and Support Action in the EU 7th Framework Programme for Research, submitted for the call FP7-ICT-2013-10. It stands for Field Operational Test Networking and Data Sharing Support. FOT-Net Data is a continuation of FOT-Net's activities. In external communication the activities will be referred as FOT-Net in order to show continuity.

The main objectives of FOT-Net Data are to:

- Support efficient sharing and re-use of FOT datasets
- Develop and promote a framework for sharing data
- Build a detailed catalogue of available data and tools
- Operate an international networking platform for FOT activities.

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

The project partners are VTT Technical Research Centre of Finland Ltd., ERTICO – ITS Europe, SAFER Vehicle and Traffic Safety Centre at Chalmers University of Technology, Institut für Kraftfahrzeuge (ika) at RWTH Aachen University, Galician Automotive Technology Centre CTAG, University of Leeds, the European centre of studies on safety and risk analysis CEESAR and the automotive company Daimler. The project coordinator is Dr. Sami Koskinen, VTT.

1.2 Purpose of the FOT Data Catalogue

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

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

The first thing that organizations looking to analyse FOT data need, is a thorough description of available datasets. The Data Catalogue features datasets in reasonable detail from the

perspectives of different impact assessment areas. A catalogue entry would, for example, introduce the study design and describe tested functionalities, groups of test subjects, measured variables and processed summary data. The catalogue includes contact details and information on conditions for data access. It is also possible to attach sample data.

FOT-Net Data has contacted and will continue to contact data owners and tool providers inviting them to provide information for the catalogue (in the autumn 2015, after internal piloting phases). The Data Catalogue was launched to pilot group in January 2015 and to the public in June 2015. The catalogue is an extension of the FOT-Net wiki, including descriptions of datasets as well as methods for inputting and searching them. FOT-Net Data continues building the catalogues throughout the project.

FOT-Net Data has also reserved budget for improving the availability of high-priority datasets, to support for example documentation or legal work. The project will select a few, prioritized data sets, where the partners work in collaboration with data owners to make the datasets available in accordance with the FOT-Net data sharing framework. The datasets will be selected based on specific criteria, where interest for re-use of the dataset in the FOT/NDS community is essential. Selected datasets and the potential re-use of the data help to prove the benefits of opening FOT data for a wider community. Real cases build the knowledge on required steps to re-use data, such as data provision, support and training, and bring out potential hurdles.

The purpose of this interim report is to describe the work done on the catalogue. The report includes a review on the state of the art on catalogue services, and description of the content and functionalities of the first version of the data catalogue. It also lists datasets identified as high value so far.

2 State of the Art on Catalogue Services

2.1 Open and Big Data

Open data and big data are hot topics in ITS research today as well as in other scientific disciplines. This is visible for example in international congresses like the ITS World. Cities are opening their data storages such as maps, public transport routes and timetables, as well as 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, etc.

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, i.e. movement of individuals, opening ITS datasets is restricted by laws on personal data. 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 doesn't therefore tell what type of data one is dealing with. Many transport datasets like the ones collected in large-scale field tests, can be regarded as big data. However, it's usually clearer to specify for an audience that the discussion or presentation is about e.g. FOT or certain specific transport datasets instead of '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, intelligent transport systems (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 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 in Horizon 2020, with just the first call for proposals in 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 current views, that all valuable datasets collected with public money, should be made available for a wider community than only to those, who collected the data.

ITS projects have not yet, as of 2015, faced hard 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 the data is also an aspect in evaluation of project proposals and the importance is growing in H2020.

FOT-Net Data is among the first ITS projects to specifically target data sharing, but more Horizon 2020 projects are likely to join the work.

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

The e-IRG (e-Infrastructure Reflection Group, e-IRG 2013) considers in their recent white paper, that agencies running large research facilities are encouraged to provide a data archive and to open it for a community, after a proprietary period allowing the original partners to benefit the most. They continue that data publication and sharing have to be fully recognized as part of scientific activities and criteria for evaluating this work (quality, impact, etc.) have to be established.

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

2.2 Catalogue Services in Other Disciplines

2.2.1 General Observations

Catalogue services usually have a beginning in their user communities, instead of being universal data libraries. Various catalogues and community-specific metadata definitions exist. Some interesting examples are reflected in this chapter.

From the perspective of analysing FOTs and being accustomed to read minimally 20 pages of documentation to get an understanding of the field trial, the metadata visible in many catalogues of other disciplines seem to be brief, or rather: well-structured and concise. When the data is of defined type or in standard form, 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 on well-structured metadata, the metadata in 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 type 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 rather well explain ice core data, where there's a practise of how ice drilling is executed, but regarding specific social studies, one would likely have to get in contact with the researchers to learn more. The example social dataset (referenced above) was catalogued but unpublished data. The ice core data, when downloaded, was provided as an Excel table and supported with a descriptive document about definitions and scanned hand-written 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 provide means to define also 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 National Science Foundation and NASA.

In a DataFed publication (Husar & Poirot 2005), it is noted that traditionally air quality analysis was 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 (UK's) Economic and Social Research Council. It's a large collection of research data in the social sciences and humanities. They have catalogue service 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 Data Documentation Initiative's (DDI, www.ddialliance.org) schema. The DDI makes an effort 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's 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.

10.5255/UKDA-SN-6392-1

Citation:
Police Service of Northern Ireland. Central Statistics Unit, *Northern Ireland Road Traffic Collision Data, 2006* [computer file]. Colchester, Essex: UK Data Archive [distributor], February 2010. SN: 6392, http://dx.doi.org/10.5255/UKDA-SN-6392-1

Change log:
11 October 2011: DOI created

Figure 1. Example DOI from 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 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

EUDAT project (European Data Infrastructure) 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. about personal data and financial models.

2.3 ITS Catalogues

ITS data catalogues are not yet as well established than their counterparts in social sciences, for example. However, there are prominent examples – and 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 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 practises used in different scientific communities offer us 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.

The large open data catalogues like what EUDAT project is building, contain also 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 ITS datasets.

The process for integrating a new discipline in general-purpose catalogues starts by the discipline 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 in the next steps. Also, much work is needed to establish the data sharing practices and requirements and basically the culture for a scientific discipline.

3 FOT Data Catalogue

FOT-Net is compiling a FOT Data Catalogue, which describes FOT and NDS datasets. The datasets included in the catalogue must be available for further research work under defined conditions and they are to be described in reasonable detail for potential re-users. The project has provided the first European data catalogue for driving studies and field trials of ITS, in co-operation with similar international catalogue efforts such as the RDE. FOT-Net has come up with a format and definitions for the data catalogue by considering current best practices for FOT dataset documentation and existing catalogue standards. Feedback and further input has been and will be gathered from the FOT-Net community, especially from dataset owners.

The work is based on the previous version of FOT-Net Wiki, which already contained a FOT Catalogue section describing projects across the world. The catalogues are moderated by FOT-Net, but as wiki pages, they receives input and updates from the FOT community – various organisations carrying out large-scale trials.

3.1 Catalogue Principles

FOT-Net Wiki includes three catalogues: The FOT Catalogue, Data Catalogue and Tools Catalogue. The FOT Catalogue is the main catalogue that aims to include all the FOTs/NDSs made 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 catalogue has three main principles:

1. Extension of the current FOT/NDS Catalogue (wiki)

2. Ease-of-use

3. Data remains with their owners.

The FOT Data Catalogue works as an extension of the current 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.
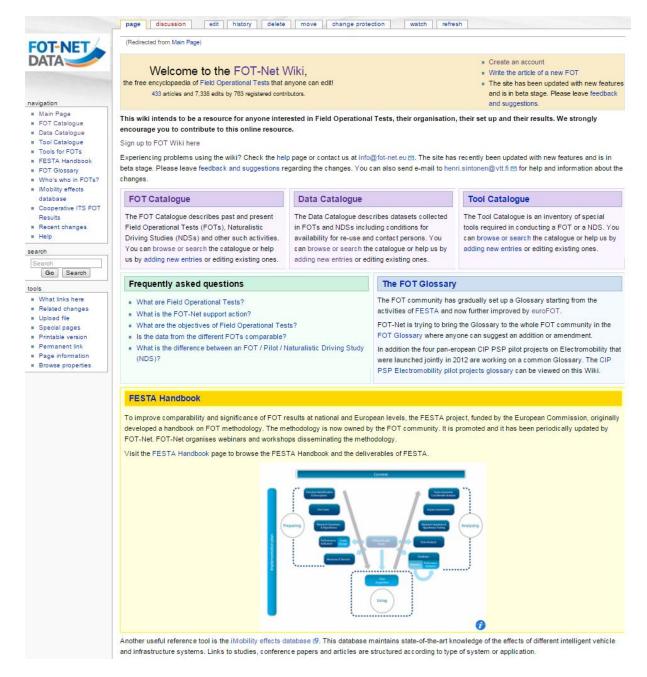
Figure 2. FOT-Net Wiki pain page with links to the three catalogues (with purple background).

Ease-of-use is a key principle for the catalogue. For someone willing to share data this includes easiness 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 includes essential information on datasets and their re-use conditions as well as easiness 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 engine as it is a 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, 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 has experimented with wiki forms to enter new data or edit the current information about a trial (Figure 3).
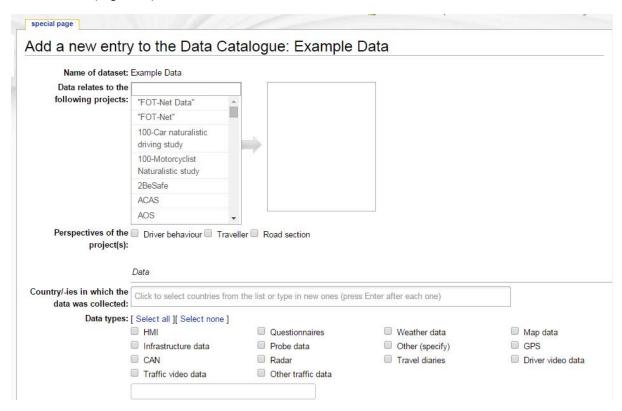

Figure 3. Part of form with which an entry can be made to Data Catalogue.

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.

Search

Enter values below. You can fill as many or as few search fields as you want. You can type partial search values (e.g. "tra" would match both "trace" and "abstract"). This search is case-sensitive.

Name contains: [   ]

Country contains: [Select countries from the list or type in search keywords. Press Enter after each one.]

Type of experiment: ☐ Naturalistic driving ☐ Controlled tests

Design of experiment: ☐ Between subjects ☐ Within subjects ☐ Mixed design ☐ Other design

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

Indicators in the raw data: ☐ Position ☐ Speed ☐ Headway ☐ Travel time / Traffic flow ☐ Vehicle control related ☐ Fuel consumption ☐ HMI related ☐ Video

Environmental conditions: ☐ Normal conditions ☐ Snowy/icy conditions ☐ Rain ☐ Reduced visibility ☐ Darkness

Number of test subjects: [min] - [max] .

Number of detected vehicles: [min] - [max] .

[Run query]

Figure 4. Search function in the catalogues.
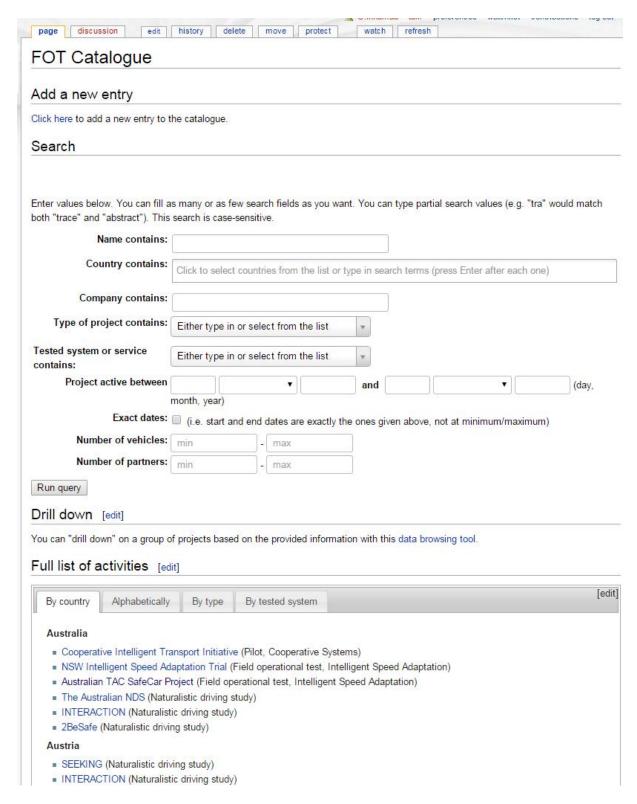
Figure 5. Browse function in the catalogues.

3.2 Catalogue development process

The process in developing the Data Catalogue was the following:

1. Definition of desired catalogue functionalities (Status: completed)

2. Enhancement of (FOT-Net FOT/NDS Catalogue) Wiki tool with additional software packages to enable the desired functionalities (Status: completed)

3. Definition of catalogue contents (Status: completed)

4. Deployment of the catalogue (Version 0.1) (Status: completed)

5. Pilot (mostly) on the catalogue content with two datasets owned by FOT-Net Data partners (First pilot: completed)

6. Update on the catalogue (Version 0.2: completed)

7. Pilot on the catalogue content and functionalities with more datasets owned by FOT-Net Data partners (Second pilot: completed)

8. Update on the catalogue (Version 1.0: completed)

9. Public launch. (Status: completed)

A summary of the process is presented with a timeline in Figure 7. In practice, the public launch of the catalogue means 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 will continue building and promoting the catalogues throughout the project.

Figure 7. FOT Data Catalogue development process

3.3 Catalogue content

The main principle in defining the catalogue content has been to make sure that it is in line with terminology and content of the metadata and data descriptions, and the data model suggested by FOT-Net Data's Data Sharing Framework, being compiled in WP3. 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 filtering/search functionalities

(summary of metadata). The following screen shots show the content of the Data Catalogue metadata fields in December 2015, when adding a new entry using a form:

Quantity of the data

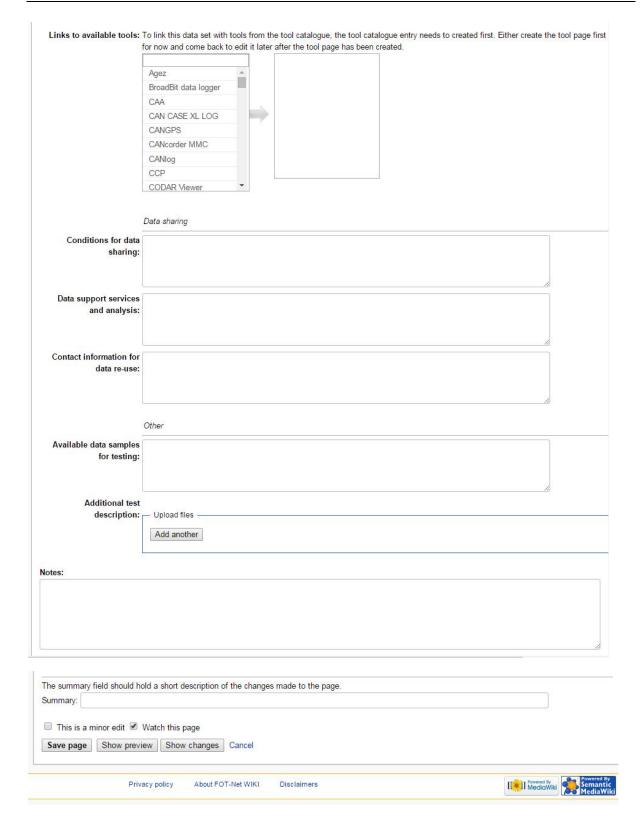|	
Kilometres driven / travelled:	[ ] km
Number of journeys:	[ ]
Number of events:	[ ]
Hours of logging:	[ ] hours
Number of detected vehicles / travellers:	[ ]

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: [   ]

Links to available tools: To link this data set with tools from the tool catalogue, the tool catalogue entry needs to created first. Either create the tool page first for now and come back to edit it later after the tool page has been created.

Agez
BroadBit data logger
CAA
CAN CASE XL LOG
CANGPS
CANcorder MMC
CANlog
CCP
CODAR Viewer

Data sharing

Conditions for data sharing:

Data support services and analysis:

Contact information for data re-use:

Other

Available data samples for testing:

Additional test description: Upload files

Add another

Notes:

The summary field should hold a short description of the changes made to the page.

Summary:

☐ This is a minor edit ☑ Watch this page

Save page   Show preview   Show changes   Cancel

Privacy policy   About FOT-Net WIKI   Disclaimers

Powered By MediaWiki   Powered By Semantic MediaWiki

3.4 Future of the Data Catalogue

The FOT-Net Community is invited to provide information on available datasets. The functionalities to provide input and update previously provided information are made so easy that FOT/NDS data owners are able to keep the catalogue up to date after FOT-Net Data project has ended. An FAQ section is provided together with the catalogue to provide help for 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. The main ones are briefly introduced in Chapter 4.1. FOT-Net has entered into discussions with the owners of these datasets during the second project year. Most of the currently available datasets are detailed later in Chapter 4.2.

4.1 Datasets Identified as High Value

4.1.1 DRIVE C2X

The European Integrated Project DRIVE C2X, the follow-up of preparation project PRE-DRIVE C2X, took place in 2011–2014. The project strongly built 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, Netherlands, Spain and Sweden. The project assessed the impacts of cooperative systems on users, 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 3rd parties. All drivers had been requested to sign an agreement releasing the data for scientific research, also for other organisations than the one who collected the data.

4.1.2 EuroFOT

EuroFOT identified and coordinated an in-the-field testing of new Intelligent Vehicle Systems with the potential for improving the quality of European road traffic. This permitted assessing their effectiveness on actual roads, while determining how they perform towards the intended objectives. The field testing offered an early publicity of the technologies, and enabled the analysis of the user acceptance and its subsequent potential for market penetration.

Instead of publishing the data, the granting of access is based on a research proposal that will be reviewed by owners of the data. Research proposals are addressed to the Management 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 can be extracted after their content has been reviewed and checked against the research proposal. All costs for the access – like training session, operator, and consumables – will be agreed upon before the analysis is started. EuroFOT project agreed the access to the data ends one year after end of project that is June 31, 2013.

4.1.3 FOTsis

FOTsis (European Field Operational Test on Safe, Intelligent and Sustainable Road Operation) was a large-scale field testing of the 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 a

full-scale deployment in European roads. Specifically, FOTsis tested the road infrastructure's capability to incorporate the new cooperative systems technology at 9 Test Sites in four European Test-Communities (Spain, Portugal, Germany and Greece).

FOTsis reviewed the road infrastructure and communication networks required to secure a proper connectivity from the traffic control centres with the 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 on the possibilities of data re-use are ongoing. FOT-Net proposed co-operation regarding data sharing aspects to FOTsis near the end of their project and discussions were started.

4.1.4 TeleFOT

TeleFOT was an integrated EU project in 2008–2012. It carried out several Field Operational Tests 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)

Some TeleFOT datasets are already available for re-use.

4.1.5 UDRIVE

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

Over a period of two years, UDRIVE will collect 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 - will be collected continuously to bring knowledge in the various research areas well beyond the current state-of-the-art.

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 is administered by the Transportation Research Board of The National Academies, under a Memorandum of Understanding with the Federal Highway Administration (U.S. Department of Transportation) and the American Association of State Highway and Transportation Officials.

SHRP2 is a targeted, short-term research program carried out through competitively awarded contracts to qualified researchers in the academic, private, and public sectors. SHRP 2 addresses 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 agree to have their cars fitted with cameras, radar, and other sensors to capture data as they go about their usual driving tasks.

Experience with earlier naturalistic driving studies demonstrates that drivers quickly forget the presence of cameras and sensors, which are as inconspicuous as possible. This allows researchers to study driving behaviour that is as close to "natural" as possible: thus a "naturalistic driving study." This kind of study is needed because driver behaviour contributes to more than 90% of crashes and is the primary factor in more than 60% of crashes.

4.1.7 Safety Pilot

Connected Vehicle Safety Pilot is a research program that demonstrates the readiness of DSRC-based connected vehicle safety applications for nationwide deployment. The vision of the Connected Vehicle Safety Pilot program is 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 are safe and do not unnecessarily distract motorists or cause unintended consequences.

4.2 Currently Available Datasets

4.2.1 DRIVE C2X

Finnish Test Site

The Finnish test site data of DRIVE C2X, owned by VTT, can be shared under an NDA. Since the dataset is reasonably well documented, VTT doesn't expect large need to support 3rd party analysts.

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

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

It must be noted that not many of the cooperative functions on the Finnish test site were based on vehicle to vehicle communication, but rather 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 following cooperative functions:

- Cooperative in-vehicle signage / speed limit
- Road works warning
- Car breakdown warning

Other datasets

FOT-Net Data will continue discussions with the data owners of the German (simTD), Italian, Spanish (SISCOGA) and Swedish test sites of DRIVE C2X.

4.2.2 EuroFOT

French Test site

No data sharing is foreseen for the French test site data of euroFOT. However, on research subjects in relation with EUROFOT objectives (e.g. road safety analysis), CEESAR can perform analysis on the French EUROFOT dataset.

Other datasets

German Test site dataset of euroFOT was available for re-use for one year after the project. It does not exist anymore and, therefore, won't be added to the catalogue.

FOT-Net Data will continue discussions with the data owners of the other test sites of euroFOT.

4.2.3 TeleFOT

Finnish LFOT (OuluFOT)

OuluFOT was a large-scale trial in Finland. It tested a smartphone (Nokia Series 60, at the time) traffic information application and a green driving advisor.

The data owner VTT sees that this dataset can be shared under a basic NDA, but it's not a high-priority dataset anymore due to complicated experimental procedure and old mobile phone applications. Understanding the experimental procedure requires much knowledge of the test, even though the steps have been documented to a rather high detail. Furthermore, the tested functions (smartphone apps) become old even during a long-lasting project. Nevertheless, the baseline data can be usable and shows general driving behaviour across Finland.

VTT sees that despite the documentation, analysts would require support and arranging that would mean that the analysts would have to contract at least a few days of work from VTT. On the positive side, there are various post-processed summary tables and analysis results already available, giving new analysts a fast start. Many analyses seem likely even based on

the existing summary tables. Similar summary tables are available from all TeleFOT test sites, enabling comparisons e.g. on driving style in different countries.

UK LFOT, UK DFOT1, UK DFOT2, UK DFOT3

Loughborough University, the owner of UK datasets in TeleFOT, is willing to consider sharing of their datasets to third parties. They see that data-sharing may be possible depending on intended use of the data.

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) supplement with e.g. rich video material the datasets collected in the LFOT:

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

Greek LFOT

ICCS, the data 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. The tested functions included navigation, speed limit information and speeding alert, traffic information; all provided with 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 tests.

Italian LFOT

UNIMORE, the data owner of the Italian LFOT data, has agreed to provide their data available for research purpose (only). 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 tests.

Other datasets

FOT-Net Data will continue discussions with the other LFOT data owners of TeleFOT, namely the Spanish and Swedish test sites.

4.2.4 National FOTs

In addition to EU-funded FOTs, there are various national FOTs that we target to contact in a later phase of the project. One case has been lifted to piloting phase, as was already explained:

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 Finnish Transport Safety Agency and the data has been agreed by the partners of the Trafisafe project to be suitable for re-use. It is co-owned by the organisations that collected it. Private data has been deleted after the project has ended – however, the owners have interest in controlling the dataset and public reporting based on it, to ensure that there cannot be even 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.

simTD

Discussions on possibility to share simTD data (or part of it) have been initiated.

4.3 Other Discussions with Data Owners

INTERACTION

Understanding driver interactions with In-Vehicle Technologies was the main objective of INTERACTION project, which started in November 2008 for 42 months and gathers 10 European partners from 8 countries and 2 Australian institutes under the EC 7th Framework Programme (FP7).

Amongst all the available In-Vehicle Technologies, INTERACTION focused on a limited set of mature technologies, already available on a wide range of car models and already adopted by most of the European car drivers, such as communication or navigation systems, speed control or distance control systems (More information on the system selection in the Results section).

For this selected set of In-Vehicle Technologies (IVT), INTERACTION aims to tackle the following questions:

- Why, when, where and how drivers use IVT?
- What are the patterns of IVT use in everyday driving?
- What are the individual factors that explain or 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 INTERACTION project can be shared and entries were made to the Data catalogue.

FKFS Electric Vehicle Fleet

The Research Institute of Automotive Engineering and Vehicle Engines Stuttgart (FKFS) is an independent institute that provides research and development services for the international automotive industry. This contact regarding their electric vehicle fleet data was initiated in Tokyo ITS World Congress. Since FOT-Net Data was at that time piloting the Data Catalogue, this discussion is yet to be continued with named contact persons.

Further information regarding their projects: http://www.fkfs.de/english/automotive-mechatronics/leistungen/electromobility/electric-vehicle-fleet/.

Other contacted datasets

Safety pilot data is being shared through RDE, which we are in contact with. The data sharing happens through RDE but information can be linked.

UDRIVE project has been contacted. The discussions will continue as the project proceeds to collect data.

4.4 Lessons learned when getting entries to catalogue

Contacting of the FOT data owners has taken a lot of time and several contacts per dataset. First of all, the original contact persons for the completed FOTs may have changed job and it has required an effort to find someone to discuss the data sharing with.

Secondly, the data owners of these already completed FOTs do not know right away whether their dataset can be shared or not from e.g. contract point of view. Thus, it has required for them going back to the original contracts and contacting legal department etc. to get a view on whether the data can be shared at all and if so, under which conditions.

Finally, if the decision on data sharing has been positive, it has required yet effort to find and going back to documentation of the data to make the entry. As FOT-Net Data partners have been involved in several of the European FOTs, we have had inside information of the set and therefore we have been able to provide practical support in creation of the entries.

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

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

Another lesson learned has been that not all the data owners are fully aware of the potential limitations to share data openly. In single cases, data owners can be even too quick to share data, forgetting that a dataset can contain some information which can at least indirectly be considered as personal data. Generally the test subjects have been promised in agreements that their data is used for scientific research purposes, only. Certain access control to the data or further anonymisation of the dataset may be required, depending on the case. FOT-Net or other catalogue operator should check, if 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 regarding maintaining test subjects' privacy. However, organising also training sessions for new scientists regarding data management and laws on personal data is something that there's currently a clear need for.

5 Conclusions

This document is an interim report that describes the work done on the FOT Data Catalogue by the end of December 2015. The report includes a review on state of the art on catalogue services and description of the content and functionalities of the current version of the Data Catalogue. It also lists first datasets that have been entered into the catalogue and those that have been contacted but about which the discussion is still ongoing.

During the second project year, we have maintained the established connections with USDOT and e.g. the EUDAT (European Data Infrastructure) project. Additionally we continued to monitor e-infrastructure work aiming at long-term preservation of data and harmonisation of metadata.

The work is strongly on-going but the focus has shifted from creation of the catalogue to promotion of available datasets and invitation of entries.

This interim report will be updated with the 2016 work and published in the finalised form as FOT-Net Data Deliverable D4.1.

List of Figures

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_2 0.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/