connected automated driving.eu

EUROPEAN COMMISSION

HORIZON 2020 PROGRAMME

CALL TOPIC DT-ART-02-2018: SUPPORT FOR NETWORKING ACTIVITIES AND IMPACT ASSESSMENT FOR ROAD AUTOMATION

ARCADE

Aligning Research & Innovation for Connected and Automated Driving in Europe

Proceedings of workshop on Data sharing

Deliverable no.	Contribution to D2.4
Dissemination level	Public
Work Package no.	WP2, WP4
Main author(s)	Sytze Kalisvaart, Jo-Ann Pattinson, Erik Svanberg, Yvonne Barnard, Johannes Hiller
Co-author(s)	
Version number	1.0
Status (F: final, D: draft)	F
Keywords	Data sharing, workshop, proceedings, ARCADE, CCAM
Project Start Date and Duration	1-10-2018 / 30-9-2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Document Control Sheet

Main author(s) or editor(s):

Work area: WP2

Document title: D2.4 Joint CCAM Networking Report

Version history:

Version	Date	Author	Summary of changes
0.1	4-3-21	Sytze Kalisvaart	General structure
0.2	9-3-21	Jo-Ann Pattinson, Erik Svanberg, Yvonne Barnard, Johannes Hiller	Description of break-outs and webinar
0.3	21-3-21	Jo-Ann Pattinson	Final language revision.
0.4	23-3-21	Erik Svanberg	Final editing
1.0	29-3-21	Jo-Ann Pattinson, Sytze Kalisvaart	Release

Approval:

	Name	Date
Prepared	Sytze Kalisvaart	29-3-2021
Reviewed		

Circulation:

Recipient	Date of submission
EC	
Project consortium	
Public	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Executive Summary

On 25 February 2021, ARCADE arranged an online workshop with between 70 and 80 participants on data sharing as defined in the CCAM Strategic Research and Innovation Agenda.

In a webinar and three break-out sessions, a large number of bottlenecks and directions for data sharing were identified and discussed. New developments like home working, the European Data Strategy, sharing approaches in GAIA-X, International Dataspaces Association, the AI community and safety related traffic information (SRTI) provide fresh input to data sharing. Also, the large European projects like L3Pilot and Autopilot push data sharing in a positive direction.

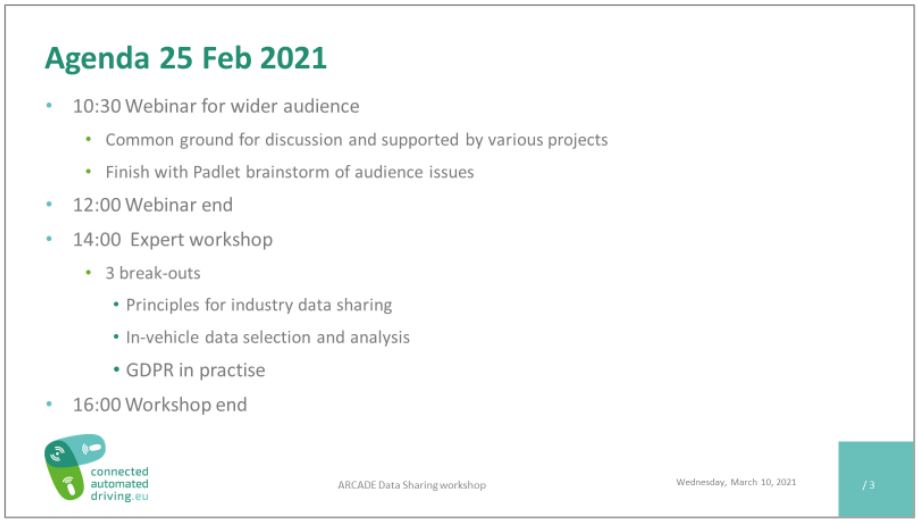
We must continue to find the balance between core intellectual property and sharing (usually) aggregated data. We must continue to collaborate on shared and harmonized data formats and processes. There is also a strong need for harmonized legal processes, creating frameworks or test beds for a more efficient and rational legal process, both from test approval and the General Data Protection Regulations (GDPR).

The results will be taken up by the ARCADE task on data sharing and some of the content will be published in the Connected and Automated Driving knowledgebase. The topic of data sharing will be further elaborated under coming European projects such as Hi-DRIVE, in the CCAM partnership, and within the data sharing initiatives heavily supported by public bodies, academia and private companies.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

1. Introduction

On 25 November 2021, an online workshop on Data sharing in the CCAM area was held with between 70 and 80 participants. The agenda was as follows:



The online workshop began with a morning webinar (see Chapter 2). After an introduction of the CCAM Working Group 2 and its targets for data sharing, three presentations on legal and technical topics were given, followed by three connected and automated driving project presentations on experiences with data sharing.

The focus was on approaches that facilitate data sharing, as the bottlenecks are well known. The morning webinar was concluded with audience input on data sharing to be addressed.

The webinar continued into an afternoon Workshop (see Chapter 3). Three expert break-out sessions were held to further explore previously selected topics of interest. The meeting ended with a plenary wrap-up.

The results have been made available on the EU CAD Knowledge Base event library (1). The Knowledge Base also has a dedicated section on Data Sharing (2).

1.1. Purpose of the document

This document is meant to capture the results of the workshop on 25 February 2021.

1.2. Intended audience

Experts, stakeholders and those interested in the field of data sharing evaluation in the domain of Connected and Automated Driving.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

1.3. Acronyms

Acronym	Full text
AI	Artificial Intelligence
CAD	Connected & Automated Driving
CCAM	Connected Cooperative & Automated Mobility
CVE	Common Vulnerabilities and Exposures
DG	Directorate General, a part of the EC organisation
EU CAD	European website and knowledge base for Connected and Automated Driving
FOT	Field Operational Test
GDPR	General Data Protection Regulation
GPS	Global Positioning System
NAP	National Access Point
NDS	Naturalistic Driving Study
SRIA	Strategic Research and Innovation Agenda
SRTI	Safety-Related Traffic Information
TDSF	Test Data Sharing Framework
WG	Working group

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

2. Webinar

2.1. Welcome

The workshop was opened by the ARCADE project coordinator Stéphane Dreher (ERTICO) with a brief introduction to the ARCADE project and project contribution to Connected, Cooperative and Automated Mobility (CCAM) (3) and Strategic Research and Innovation Agenda (SRIA) (4).

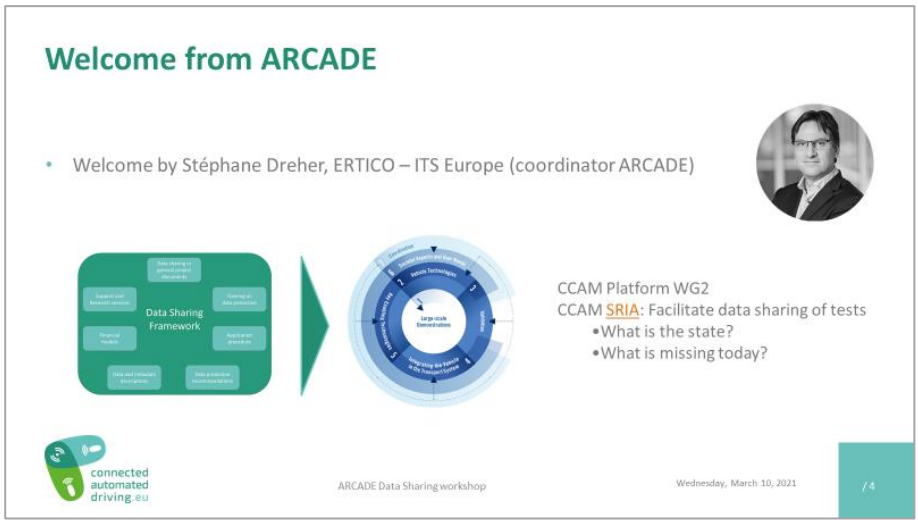


Figure 1: Stakeholder Contribution cycle presented by Stéphane Dreher

Stéphane Dreher also introduced the EU CAD Data Sharing Framework (2). The aim of the framework is to support projects and organisations by giving hands-on recommendations in different topics, all important to enable data re-use or sharing. The framework is focused on field operational tests or naturalistic driving studies, but has proven to be useful (although adapted) to other domains. Some parts are related to early phases of data collection giving examples of pitfalls to avoid, that later can make sharing or re-using data impossible.

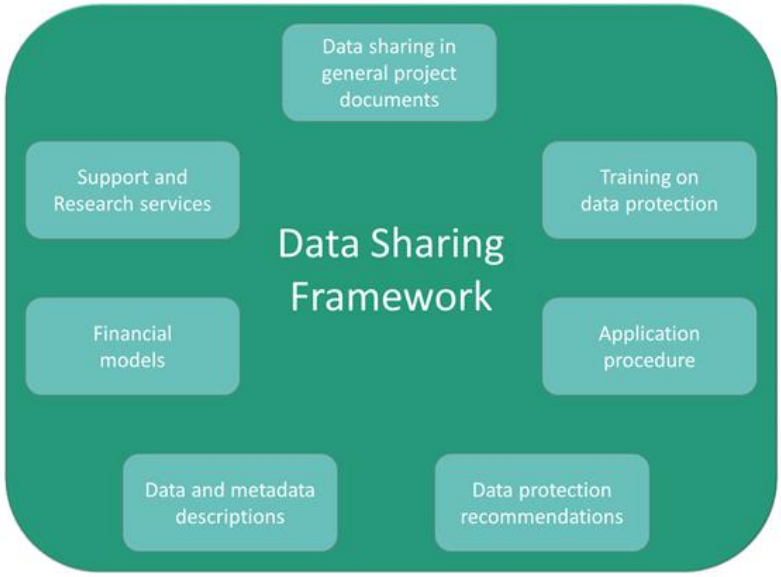


Figure 2: EU CAD Data Sharing Framework

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

There are five administrative topics: 1) agreements and contracts, 2) training, 3) application procedures, 4) support and research services and 5) financial models.

Data and metadata description and data protection recommendations are more practical and data centric.

The data and metadata description outlines a model that considers many different aspects of describing data. Not only the actual raw measures must be described, but also derived measures, file formats, policies, and the context in which the data was collected.

The section on data protection will not detail the exact measures to put in place (which depends on the use case) but supports organisations in asking relevant questions internally, and also to partners wanting to re-use data.

The framework is to be reviewed this year and comments or proposals of changes to the content are welcomed. On the EU CAD website there is a Feedback Form that can be used to get in contact.

2.2. Introduction of Data sharing in CCAM context by Tom Alkim

Tom Alkim of DG Research & Innovation introduced the CCAM Partnership and explained its objectives. Tom Alkim and Guus van de Schouw lead Working Group 2 (WG2) on Coordination and cooperation of research and innovation.

Next, he introduced the CCAM Platform and the European Common Test Data Sharing Framework for CCAM. One of the goals of WG2 is to 'Facilitate data sharing of tests'.

A conclusion document from WG2 (Q1 2021) will include challenges, priorities and requirements for the Test Data Sharing Framework (EU-TDSF, see Figure 2)

2.3. Sharing personal data in Legal issues on data sharing, GDPR and research data

Jo-Ann Pattinson began by set out the context for why GDPR (5) was enacted. Research in the digital age has supercharged our ability to collect, communicate and analyse data, however this has impacted our privacy. Privacy is a fundamental human right protected under international law (such as the Universal Declaration of Human Rights, the International Convention of Civil and Political Rights and the European Convention on Human Rights) however these international conventions are implemented domestically by national law. The wide array of national laws prompted the development of the GDPR to; harmonise data protection law, increase protection for personal data in the face of the digital economy and otherwise set an international standard for protecting personal data online.

Automated vehicle research often involves the collection of personal data. Instruments such as radar, lidar, cameras and GPS tend to collect data which may pinpoint the date and time an individual was at a location, which may reveal their identity, and may also allow us to draw inferences about their personal lives. Those involved in such research may be concerned about how to properly comply with GDPR requirements, and avoid the potentially

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

high penalties for a breach. However, the point was made, that high fines approaching the maximum (€20 million or 4% of total global turnover) are very rare.

The factors relevant in the levying of fines under GDPR were explained, using the seminal case of the €50 million fine issued by National Data Protection Commission (France) to Google, to illustrate one end of the spectrum. The relevant facts being that Google has access to an unprecedented amount of sensitive data about millions of people, and an ability to combine that data to create tertiary data, and explained how the case concerned a lack of transparent communication to users about the data collected and its potential use. However, at the other end of the spectrum, it was explained that depending on the circumstances, sometimes fines are not issued at all even in the case of a breach, and that fines issued are reflecting the extent of the detriment caused to individuals' privacy, and the efforts made by the relevant organisations to comply with GDPR. Examples were provided from Spain, France and the Czech Republic of €7,880, €75,000 and €588 respectively.

The main message being that the highest fines are for severe cases where the data protection rules have been blatantly flouted, and the risk or detriment to individuals is high. Organisations sharing personal data can significantly reduce their risk by implementing organisational and technical measures, such as;

- A data sharing agreement
- Utilising data minimisation techniques
- Using encryption or secure data sharing platforms
- Ensuring their IT meets industry standards and ISOs (International Organisation for Standardisation Certifications)
- If transferring data to another party to verify the state of their IT security
- Ensure the partner's obligations are reflected in the data sharing agreement.

When such organisational and technical measures are in place, personal data can be shared for research purposes at low risk.

2.4. Technical examples on facilitating data exchange

Erik Svanberg introduced the topic by quoting the European data strategy (6) where it is mentioned that data sharing has not taken off at sufficient scale due to lack of economic incentives including; fear of losing competitive edge, lack of trust in agreements being followed, fear of misappropriation of the data by third parties, and a lack of legal clarity on who can do what with the data. This is not news to the automotive sector, and we as industry must work to match different stakeholder interests building mutual trust.

Erik Svanberg mentioned a few interesting initiatives: L3Pilot (7) sharing aggregated data among 12 vehicle owners, GAIA-X (8), International Data Spaces Association (9), KRAKEN project (10), and the recently established national access points. Schrems II (11) and GDPR was also mentioned as legal requirements needed to be taken into account, but also something that we should now be starting to be more efficient in adopting.

Two examples of facilitating or enabling data was presented by Joost Vantomme, ACEA and SRTI Ecosystem, presenting "SRTI Ecosystem - Data sharing for road safety", and Mats Nordlund, Zenseact and AI Sweden, presenting "Edge lab within AI Sweden".

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Joost Vantomme set the scene as to why safety related traffic information (SRTI) is one component addressing road fatalities and accidents in the EU. The initiative was started by European Commission in regulation in 2013. Information on eight different traffic incidents should be shared free of charge (to end-users). This led to the establishment of a data task force in 2017, and a successful proof-of-concept of implementing SRTI for ITS Europe in Eindhoven 2019. The data task force was transformed to the SRTI Ecosystem, where a multi part agreement was signed in October 2020.

The SRTI Ecosystem builds on a cooperative approach, where you can access messages if you also contribute. Data is being governed under different categories: L1 data (raw data that will not leave the vehicle), L2 data (SRTI messages), L2 data (aggregated messages), and L3 data (service creation). The L3 data is broadcasted using national access points (NAP).

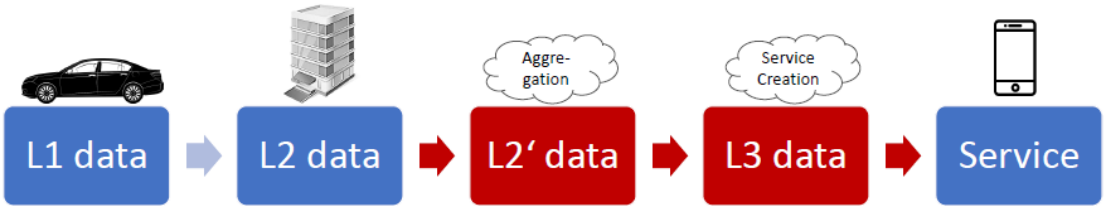


Figure 3: Levels of data in SRTI ecosystem

The entire system follows a decentralized approach. Any participant in the system takes on a defined role with obligations formulated. The Data taskforce report final report is well-worth reading, and the multi-party agreement also gives insights and is available Data for road safety website (12).

Mats Nordlund presented the basic concept of edge computing. Many domains can benefit from the concept, most notably edge computing is currently used within healthcare. In automotive, and especially in developing CAD functions, there is a constant need for data and the current levels of collecting data is not sustainable due to the cost for storage and transfer.

Features are required to be developed to cope with the increase in data, and to cover rare (edge) cases. The data is generated in the edges; why not keep it there? Two main principles were presented: 1) aggregation- enriched models shared with a central server and 2) swarm learning- where the improved models are shared among nodes (vehicles).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

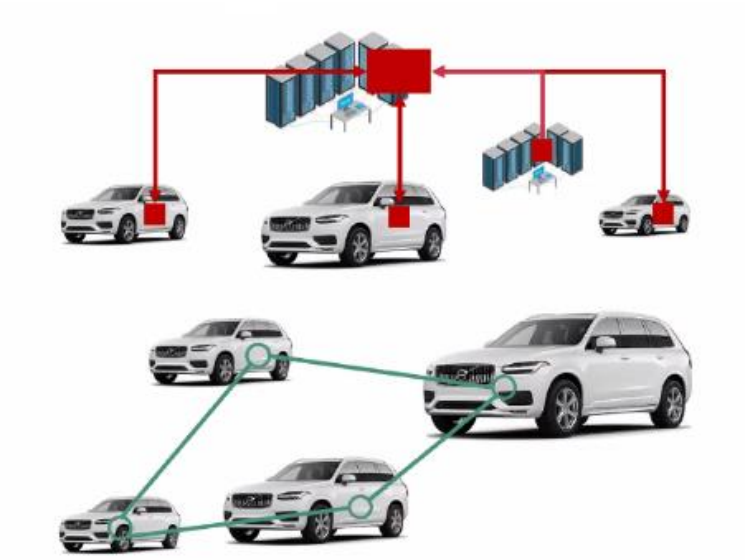


Figure 4: Edge computing using central server or swarm learning

Mats Nordlund then presented the Edge lab, inaugurated in February 2021 (13), which is maintained under AI Sweden, a Swedish national initiative on AI. The lab has the necessary features to learn how to implement edge computing in vehicles in an efficient manner including server hardware, emulators, and data loggers.

2.5. Experiences with data sharing in large projects

UDRIVE

Erik Svanberg of SAFER, presented UDRIVE (14), a large-scale naturalistic driving study, running from 2012-2017. The project had 120 Renault cars driving in France, Germany, Poland and the UK. Also, more than 40 Volvo distribution trucks collected data in the Netherlands and 40 scooters in Spain. The vehicles often collected more than a year of normal driving data from 2015 to 2017.

The UDRIVE dataflow, in its basics, were encrypted data disks sent to one of three "local data centres", pre-processing the data. The pre-processed data was then sent to SAFER acting as the central data centre in the project.

All the equipment and records were registered in an online monitoring tool, keeping track on everything from vehicles, data loggers, encrypted disks, down to individual records of data.

The UDRIVE dataset consists of 271 000 trips, 62 000 hours of driving data, and representing 2.6 million km of day to day driving. The data analysis was performed by 120 users from 14 analysis sites all over Europe. The process was supported with a data viewer and annotation tool, including the capacity of data processing and data versioning, developed by CEESAR. There were actually 140 iterations of re-processing the complete individual databases.

To ensure proper usage and management of data, a data protection concept was developed. The concept put requirements on each actor in the chain, from data collection

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

sites, to data centres and analysis sites. Each actor type implied different requirements, and every UDRIVE partner managing data had to respond to the requirements and passed a certification step.

The data protection concept was developed in cooperation with the FOT-net data project (running 2014-2016), and is included and further developed in the EU CAD Data sharing framework.

Lessons learned:

- The implementation and use of the online monitoring tool enabled the project to have full control over data to ensure it ended up at the right place.
- The analysis processed followed an iterative approach, which was appreciated by experienced analysts. It should be mentioned that it was challenging for those not as fluent in software development concepts.
- The remote workstation environment is not a local workstation. The performance was not as good as expected even though the backbone was quite powerful. The selected platform, based on Matlab, is not ideal for such environment.
- The UDRIVE analysis continued after the project in the UDRIVE USER Group. The user group finished towards the end of 2020.

The UDRIVE dataset is available from former UDRIVE partners in France, Germany, the Netherlands, UK, and at SAFER in Sweden. If personal data is to be analysed, this must be performed at the partner premises (as stated in the consents with the participants).

AUTOPILOT

Jordi Pont (IDIADA) presented the way in which data was handled in the AUTOPILOT project (15). AUTOPILOT (AUTOmated driving Progressed by Internet Of Things) was a European Innovation Action, running from January 2017 till December 2019, with 44 partners and a budget of 25M €.

The goal of the project was to use Internet of Things technologies to accelerate, enhance and enable automated driving functions.

Data coming from different pilot sites and different systems was shared amongst partners to be analysed. The quality of the data was checked to ensure homogeneity. The data was stored on the Pilot Site Test Server and Central Test Server. All data was anonymised, and a set of common metadata was used to share useful technical data after the project.

The following picture shows the data acquisition, storage and use process:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

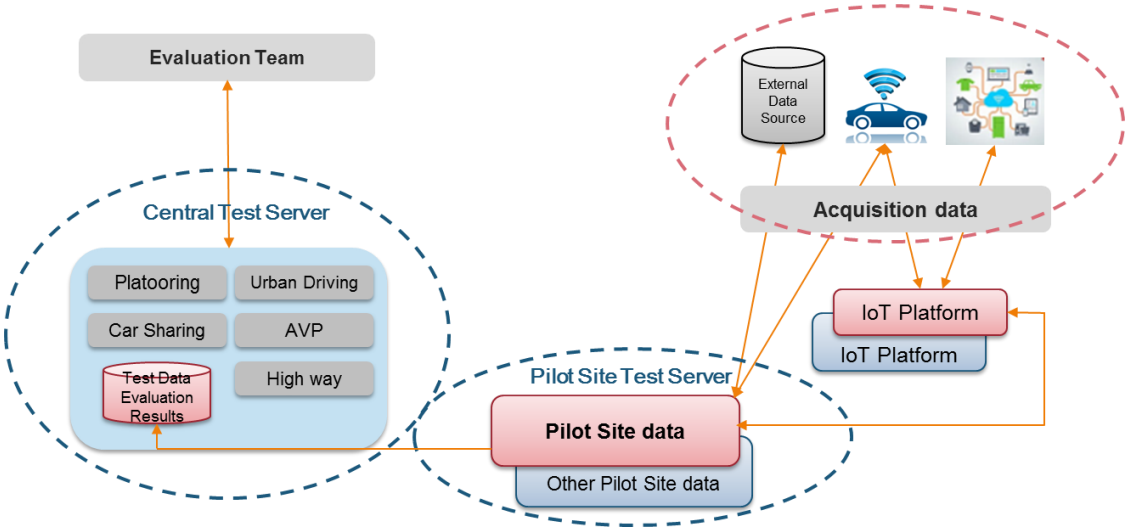


Figure 5: AUTOPILOT data flow

Jordi emphasised the lessons learned from this large and complex project: ensure good communication between Evaluators and Pilot Sites. To be able to work together efficiently it is important to define (and use) logging guidelines (Common Data Format, File format and naming). Finally, to ensure good quality of the dataset: automize the data validation process.

L3Pilot

Yves Page, Renault, presented the L3Pilot project (7), a European flagship project for piloting automated driving. It is a large project with 12 vehicle manufacturers participating, with the project coordinated by VW. Data is collected by vehicle manufacturers in sites all over Europe, performing different types of pilot activities in highway or urban conditions. The project is ending this year with the final event at ITS World congress in Hamburg.

A method for data flow was designed before the project commenced. Raw data was collected by each respective vehicle manufacturer. The data then passed a first filter, reducing the number of signals from thousands to a hundred, and this step included conversion from the vehicle manufacturer's internal data format, to a, by the project developed, common data format. The data format has been published as open source (16).

The data is shared between the vehicle manufacturer and selected analysis partner. This partner runs the common data processing framework, which was also developed within the project. Aggregated scenario data is created and anonymized (the second filter) and then uploaded to a L3Pilot "consolidated database", containing the aggregated scenario data from all pilot sites. Also, subjective data is collected and shared in the consolidated database.

The L3Pilot consolidated data will pass a final filter for a final deliverable from the project in the form of a publicly available database (a proposal is currently reviewed).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

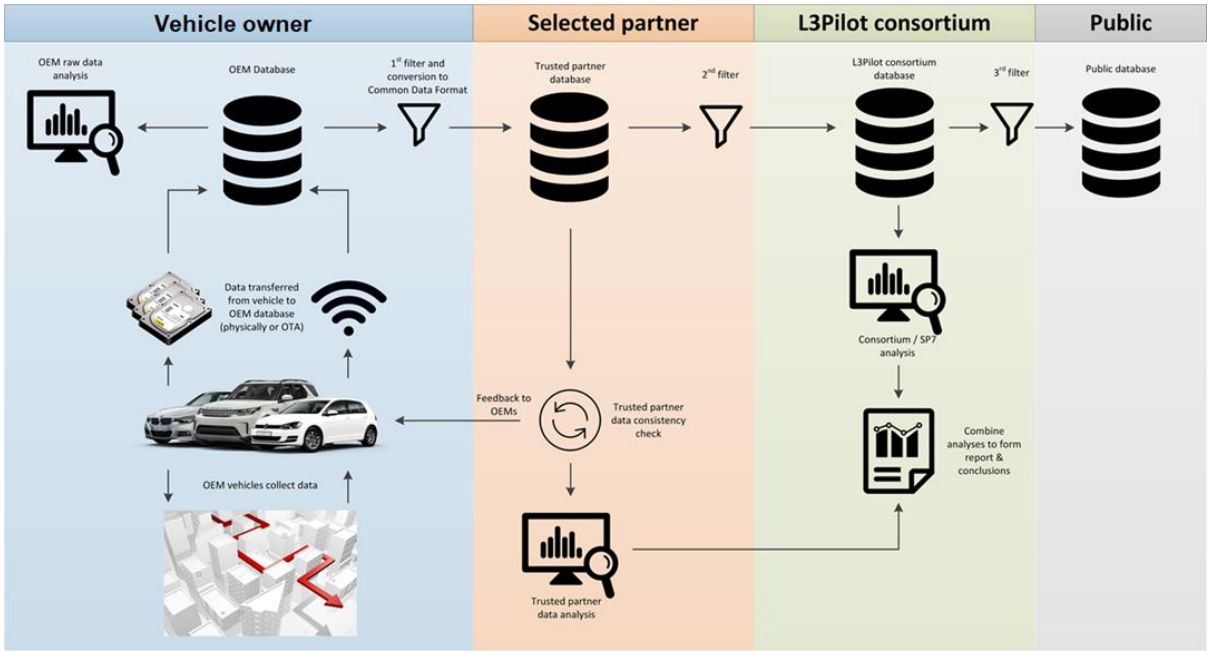


Figure 6: L3Pilot data flow

The data processing toolchain creates derived measures, scenarios and performance indicators on common format data. In some case this includes an iterative approach where data is created in different steps, e.g. manual or automatic video annotation. In-depth technical analysis can be made at the selected analysis partner.

Looking into the consolidated database the data originates from the automated driving and baseline data for comparison. All the vehicle owners have uploaded aggregated and pseudonymised data, which researchers can download for the use in impact assessment and cost benefit analysis. For the latter, also additional external data sources will be used.

Examples of the scenarios created and shared include;

- uninfluenced driving,
- car following,
- driving in traffic jam,
- cut-ins,
- lane changes,
- incidents with rear, front or side objects,
- additional scenarios for urban environments.

To conclude, the amount of data shared in L3Pilot extends beyond, previous European projects within this domain. We hope to build on these principles for future field operational tests in automated driving.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

2.6. Inventory of data sharing topics to be addressed

The webinar audience was asked which topics they thought should be addressed in data sharing using a Padlet (17).

This resulted in a rich list of topics. This was used as input for the afternoon expert workshops. You can find the readable form of the results in Annex 1 Inventory of data sharing topics to be addressed.

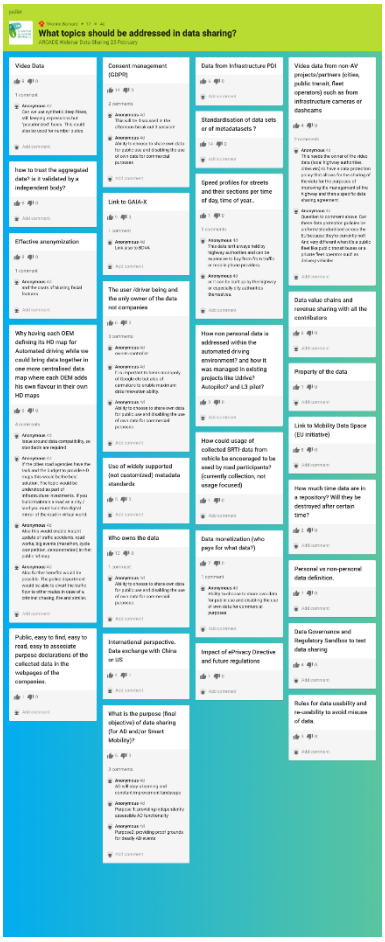


Figure 7: Webinar Padlet

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

3. Break-out sessions

3.1. Principles for industry data sharing

In this break-out we discussed principles that enable sharing of industry data as well as identified bottlenecks. Best-practices such as data protection and privacy-by-design, the use of trusted third parties, the explicit right of use for specific topics only and data coding were addressed. A number of well-known bottlenecks were sketched in the following figure:

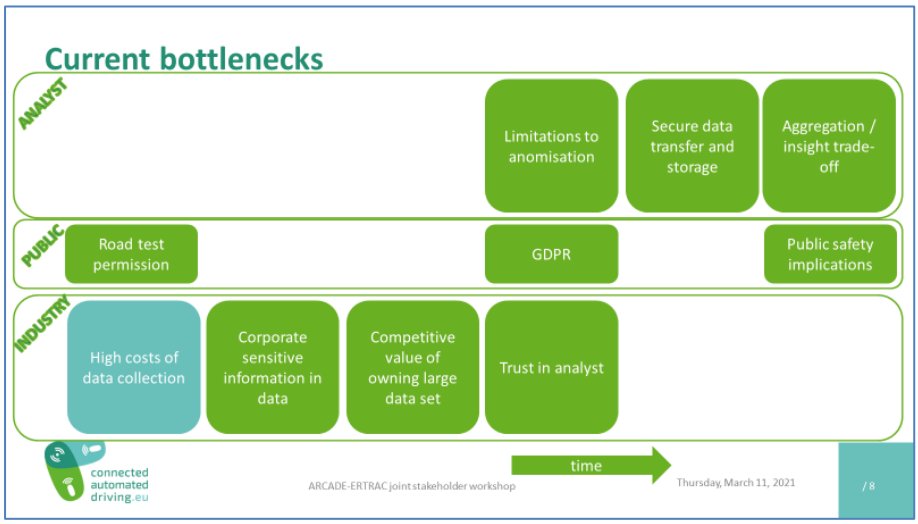


Figure 8: BO1 - bottlenecks in data sharing

Carlo van Driesten (BMW) presented the Enhanced Validation & Certification of Highly Automated Vehicles processes within the ENVITED Data Ecosystem. The ENVITED research cluster bundles stakeholders and their requirements in the field of virtual design, validation and certification of automated vehicles and connected mobility systems (18, 19). The goal of this initiative is to develop an Open Distributed Data Ecosystem.

Sytze Kalisvaart (TNO) presented the TNO PUMAS project (20) in which scenarios are shared between four partners, but not raw data. He explained the sharing principle such as need-to-know access to data sets for staff at trusted third party, data collecting party remains owner, aggregation before distribution, joint processing agreement. Data and scenarios are anonymized: camera data is blurred before transfer (facial features, license plates), and scenarios are placed on typical road geometry, not on a unique GPS location. The application for using data beyond the project is under separate agreement.

The group continued with a Padlet discussion addressing the following topics:

- How can confidential data be anonymised, changed, or aggregated, to allow for more open access?
- Are there special conditions for sharing and re-using the data after the project?
- What should be regarded as confidential information and what can be shared?
- What next steps should have priority?
- What if you find public safety or other ethical issues in the data analysis?
- What best practices do you propose to share data?
- What should be done about standardization of datasets and metadata sets?

A large set of topics and recommendations were gathered. Some of the favourite topics were:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

- The privacy of data in accident videos is critical however this data is arguably more important data for safety research.
- For anonymization, it is good to translate raw data into scenarios and labelled driving sections, however for a full system assessment more information/data is needed.
- Conditions for sharing should be clarified at the start of a partnership.
- In principle, data should be public if funded by public funds, but it may be more complicated with mixed funding, and for confidentiality reasons, data from ego vehicles could not be analyzed by external parties.
- To deal with findings impacting public safety, an approach similar to software vulnerabilities exposure could be used (CVE, see 21).
- It is a best practice to publicly disclose what data is or will be collected in a project.
- Best practice: inform the (prospective) data user about why the data was gathered and what the data can be used for (restricted by participant consent or other data agreements).

Interesting links provided by the participants for further information (see also 5 References: NHTSA Automated Vehicles for Safety (22), Safety Pool (23), and ASAM VCD - Video Content Description (24).

3.2. In-vehicle data selection and analysis

The breakout-session was led by Adrian Zlocki (fka) and Johannes Hiller (ika).

To kick off the discussion on the topic of data selection, there were two impulse presentations, offering two insights on the topic. The first presentation by Johannes Hiller was out of the context of L3Pilot. It offered insight into the way of handling data in one of the largest EU piloting projects. Within the presentation, the L3Pilot Common Data Format was presented which enables pilot sites and researchers to work together more efficiently and to commonly work on one toolchain. However, it was also highlighted, that the process of getting to that point is not easy and that many iterations between various partners might be necessary, before the toolchain can be used effectively.

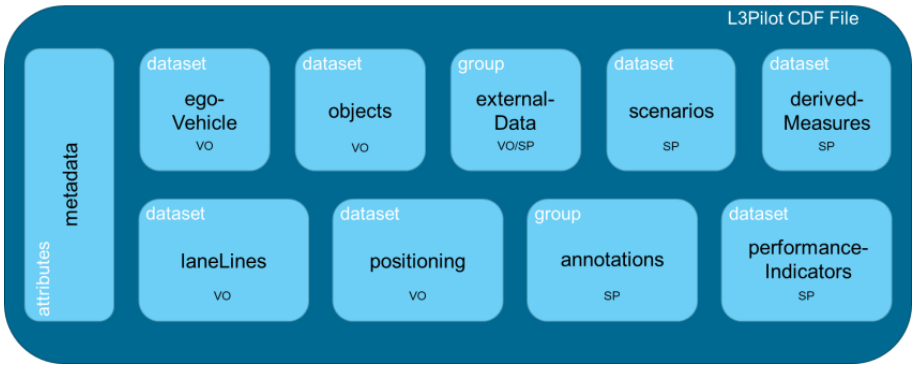


Figure 9: L3Pilot common data format

In the second presentation, which was given by Clement Val (CEESAR), an overview of the activities towards more in-vehicle data processing was given. The presentation summed up the experiences that CEESAR made with national and EU-wide projects with data of vehicles. One of the take-aways was, that data is often re-used in unexpected ways and it is

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

therefore good to have it. A big focus was on the question, what data can actually be merged (e.g. from different sources) and the problems arising from that. The solution from the experience is to not just share the data, but also share data collection equipment and processing frameworks to simplify integration.

The further discussion revolved around the topic of data and the formats needed. Some time was spent on discussing the re-use of the L3Pilot CDF as well as the influences towards the ASAM activities (25). It was also highlighted in a discussion, that often the engineering side and the legal side of data sharing are handled separately, whereas a closer combination of the two could lead to a better understanding of what is possible.

The fact that raw data is an issue when wanting to work with it was also highlighted. The current solutions here seem to be, to do a pre-selection of signals and data within the car. However, you have to keep in mind that this might restrict the re-use of the collected data at a later stage. It was agreed, that there is a trade-off to be made here and generally accepted that re-use of data is probably only possible within limited use-cases (or also “families of use-cases”).

Another topic discussed was the availability of HD maps and since roads are publicly owned in many European countries, if it would be a role for the authorities to provide these.

In the end it was agreed, that data has many aspects and that it will always be difficult to cover all those aspects in one go. However, more common use-cases and formats could help ease the transition between different use-cases and applications.

3.3. GDPR in practise

This breakout utilised a collaborative Padlet with discussion topics which were distributed to the Breakout participants prior to the Webinar, to allow participants to reflect and contribute. During this breakout, participants were particularly engaged in discussing; the potential of regulatory sandboxes, issues with consent and consent chains, issues with legitimate interest and academic/research exceptions to GDPR, and data minimising techniques including the use of ‘deep fakes’ to generate synthetic data which is realistic enough for machine learning.

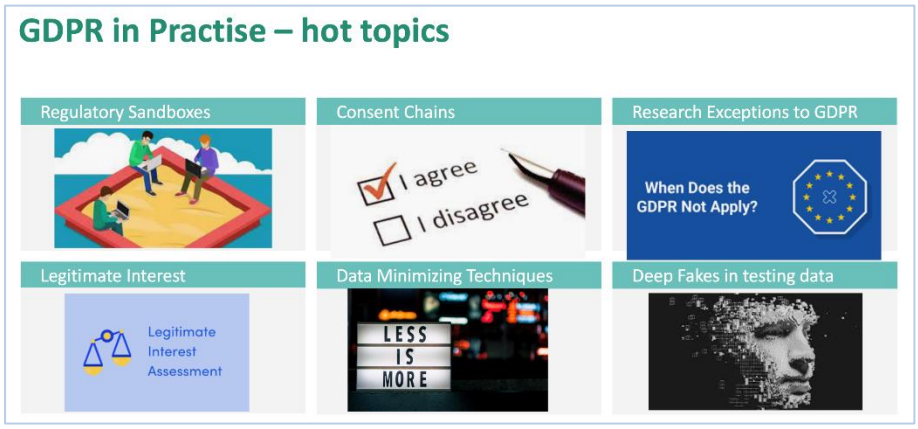


Figure 10: Hot topics in GDPR

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

What are the most difficult challenges in sharing personal data?

- Some participants described scenarios where other parties to research were reluctant or refused to share data, citing GDPR as the reason. The purpose of GDPR is not to stop personal data being shared, but to share such data in a manner which aligns with the principles of data protection, and protects the privacy of individuals as much as possible
- We need data sharing / re-use to motivate the significant cost in collecting research datasets.
- Proportionality was as a mean to justify the level of the measures taken into place.
- Regulatory sandbox for CCAM tests – including GDPR but also support in the process of applying for test on public road.
- Data as a commodity (ownership) cannot be used. There are rights to process and store data, thus something that can relate to GDPR concepts of data processor and data controller. This is where we need to put the focus, establishing Personal data processing agreements, which are clear, easy to read, and comply with consents from persons that are included in the dataset.
- International data sharing has been made more difficult since Schrems II. Also, standard contractual clauses should be reviewed when working on the large US controlled platforms. Sharing data with Chinese organizations should be handled cautiously and similar to US, where the state has the authority to access data in a greater extent than in Europe.

Consent Chains/management

- The data subject should be in control over where, who, when, and for what purpose her data is being processed.
- Introduce a process for communicating updates with regards to processing personal data.
- There is an opportunity for consent management where the individual should be put in charge. It would be preferable to collect many different services under one umbrella where consent could be managed. There are challenges associated to this, but if handled correctly this could support data sharing of personal data.
- A model of dynamic consent used in healthcare research is presented in this YouTube video (26).

Academic and Research Exceptions to GDPR

- It emerged this is a topic of much interest yet one which is not well understood, and under-utilised by academics and researchers.
- Exceptions differ country to country
- Little practical guidance as to how these exceptions can be used
- Case study of the UK: Most obligations under GDPR can be disregarded where complying would render the research task impossible or nearly impossible.
- Case by case basis, use of exceptions must be well documented and justified.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

3.4. Wrap-up and next steps

The workshop was concluded with the wrap up of the breakout sessions (see the following chapter) and next steps. The participant's input regarding the next steps was collected using Padlet.

The organisation team extracted the following possible next steps from the workshop:

a) Definition of a regulatory sandbox in which data collection experiments can take place;
b) A list of potential research exemptions to GDPR;
c) Further explanation of practical use of GDPR to enable data sharing;
d) Formulation of an easy and understandable approach for organising participant consent and consent chains;
e) Developing effective and efficient means for anonymization;
f) Standardisation of metadata, e.g. for operational design domains, scenarios, functionality and services;
g) Common Vulnerabilities and Exposures (CVE) as a model for managing public safety relevant events;
h) Definition of approaches for balancing between public interest and industrial confidentiality;
i) Define requirements for data sharing in real-time and for in-vehicle selection of data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

4. Conclusion

On 25 February 2021, ARCADE arranged a webinar followed by a workshop on Data Sharing with between 70 and 80 participants.

In the one-and-a-half-hour morning webinar, new perspectives to data sharing were presented, including “Data for road safety” sharing safety-related traffic information (SRTI) in real-time within the partnership, and “Edge computing” within the lab hosted at AI Sweden. Also, three large European projects presented their take on data sharing.

The workshop had three breakout sessions on specific themes, where insights and experiences from the participants were shared and relevant data sharing topics discussed.

There are many positive initiatives on data sharing in clusters, partnerships and platforms. There are still many hurdles and pitfalls to avoid, but many promising proposals for actions where discussed.

We must continue to find the balance between core intellectual property (IP) and sharing (usually) aggregated data. We must continue to collaborate on shared and harmonized data formats and processes. There is also a strong need for harmonized legal processes, creating frameworks or test beds for a more efficient and rational legal process, both from test approval and GDPR.

The results will be taken up by the ARCADE task on Data sharing and some of the content will be published in the Connected and Automated Driving Knowledge Base under the data sharing section.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

5. References

1. EU CAD website event, https://knowledge-base.connectedautomateddriving.eu/5th-joint-stakeholder-network-on-data-sharing/, accessed on March 23rd 2021.
2. EU CAD data sharing framework, https://knowledge-base.connectedautomateddriving.eu/data-sharing/methodology-data-sharing-framework/, accessed on March 23rd 2021.
3. EU CCAM, https://ec.europa.eu/transport/themes/its/c-its_en, accessed on March 23rd 2021.
4. Strategic Research and Innovation Agenda (SRIA) V1.0. Proposed European Partnership under Horizon Europe, February 11, 2020, https://www.ertrac.org/uploads/images/CCAM Partnership SRIA v1.0 02-11-2020.pdf, accessed on March 23rd 2021.
5. GDPR, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434, accessed on March 23rd 2021.
6. A European strategy for data, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066, accessed on March 23rd 2021.
7. L3Pilot project, https://www.l3pilot.eu/, accessed on March 23rd 2021.
8. Gaia-X, https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html, accessed on March 23rd 2021.
9. International data spaces association, https://www.internationaldataspaces.org, accessed on March 23rd 2021.
10. Kraken project, https://www.krakenh2020.eu/, accessed on March 23rd 2021.
11. Schrems II, https://edpb.europa.eu/news/news/2020/statement-court-justice-european-union-judgment-case-c-31118-data-protection_en, accessed on March 23rd 2021.
12. Data for road safety, https://www.dataforroadsafety.eu/documents, accessed on March 23rd 2021.
13. AI Sweden, https://www.ai.se/en/news/revisiting-data-factory-day-and-grand-opening-edge-lab, accessed on March 23rd 2021.
14. UDRIVE project, https://www.swov.nl/en/project/udrive, accessed on March 23rd 2021.
15. Autopilot project, https://autopilot-project.eu/, accessed on March 23rd 2021.
16. L3Pilot common data format, https://github.com/l3pilot/l3pilot-cdf, accessed on March 23rd 2021.
17. Padlet dashboard, https://padlet.com/dashboard, accessed on March 23rd 2021.
18. Envited marketplace, https://www.envited.market, accessed on March 23rd 2021.
19. Automotive Solution Center for Simulation e.V., https://asc-s.de/, accessed on March 23rd 2021.
20. PUMAS project, Scenario-based safety validation for Connected and Automated Driving TNO, https://www.tno.nl/streetwise, accessed on March 23rd 2021.
21. Common Vulnerabilities and Exposures, https://cve.mitre.org/about/index.html, accessed on March 23rd 2021.
22. NHTSA Automated Vehicles for Safety: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety, accessed on March 23rd 2021.
23. Safety Pool: The Global Incentive-based brokerage of shared Driving Scenarios and Safety Data for Autonomous Driving Systems https://www.safetypool.ai/, accessed on March 23rd 2021.
24. VCD - Video Content Description: https://www.asam.net/members/product-directory/detail/vcd-video-content-description/, accessed on March 23rd 2021.
25. ASAM, https://www.asam.net/, accessed on March 23rd 2021.
26. Dynamic consents, Jane Kaye, 2017, https://youtu.be/1NgZVuhY8w8, accessed on March 23rd 2021.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

6. Annex 1 Inventory of data sharing topics to be addressed

padlet

universityofleeds.padlet.org/ybarnard/3q49rx9w1zo4j99l

What topics should be addressed in data sharing?

ARCADE Webinar Data Sharing 25 February

YVONNE BARNARD FEB 22, 2021 11:17AM

Video Data

Can we use synthetic deep fakes, still keeping expressions but "pseudonized" faces. This could also be used for number plates. — ANONYMOUS

Consent management (GDPR)

This will be discussed in the afternoon break out 3 session! — ANONYMOUS

Ability to choose to share own data for public use and disabling the use of own data for commercial purposes — ANONYMOUS

Data from Infrastructure PDI

Video data from non-AV projects/partners (cities, public transit, fleet operators) such as from infrastructure cameras or dashcams

This needs the owner of the video data (local highway authorities, cities etc) to have a data protection policy that allows for the sharing of the data for the purposes of improving the management of the highway and then a specific data sharing agreement. — ANONYMOUS

Question to comment above: Can these data protection policies be uniform/standardised across the EU because they're currently not! And very different when it's a public fleet like public transit buses or a private fleet operator such as delivery vehicles — ANONYMOUS

Standardisation of data sets or of metadatasets ?

how to trust the aggregated data? is it validated by a independent body?

Speed profiles for streets and their sections per time of day, time of year..

This data isn't always held by highway authorities and can be expensive to buy from from traffic or mobile phone providers. — ANONYMOUS

or it can be built up by the highway or especially city authorities themselves. — ANONYMOUS

Link to GAIA-X

Link also to BDVA — ANONYMOUS

Effective anonymization

and the costs of blurring facial features — ANONYMOUS

The user /driver being and the only owner of the data not companies

owner=controller — ANONYMOUS

It is important to break monopoly of Google etc but also of carmakers to enable maximum data innovation ability. — ANONYMOUS

Ability to choose to share own data for public use and disabling the use of own data for commercial purposes — ANONYMOUS

Data value chains and revenue sharing with all the contributors

Why having each OEM defining its HD map for Automated driving while we could bring data together in one more centralised data map where each OEM adds his own flavour in their own HD maps

Issue around data compatibility, so standards are required. — ANONYMOUS

If the cities road agencies have the task and the budget to provide HD maps this would be the best solution. The topic would be understood as part of infrastructure investments. If you build/maintain a road as a city / land you must build the digital mirror of the road in virtual world. — ANONYMOUS

Also this would enable instant update of traffic accidents, road works, big events (marathon, cycle competition, demonstration) in that public hd map. — ANONYMOUS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251

Also further benefits would be possible. The police department would be able to divert the traffic flow to other routes in case of a criminal chasing, fire and similar. — ANONYMOUS

How non personal data is addressed within the automated driving environment? and how it was managed in existing projects like Udrive? Autopilot? and L3 pilot?

Property of the data

Link to Mobility Data Space (EU initiative)

How could usage of collected SRTI-data from vehicle be encouraged to be used by road participants? (currently collection, not usage focused)

Use of widely supported (not customized) metadata standards

How much time data are in a repository? Will they be destroyed after certain time?

Who owns the data

Ability to choose to share own data for public use and disabling the use of own data for commercial purposes — ANONYMOUS

Data monetization (who pays for what data?)

Ability to choose to share own data for public use and disabling the use of own data for commercial purposes — ANONYMOUS

Personal vs non-personal data definition.

Data Governance and Regulatory Sandbox to test data sharing

Public, easy to find, easy to read, easy to associate purpose declarations of the collected data in the webpages of the companies.

International perspective. Data exchange with China or US

Impact of ePrivacy Directive and future regulations

Rules for data usability and re-usability to avoid misuse of data.

What is the purpose (final objective) of data sharing (for AD and/or Smart Mobility)?

AD will stay a learning and constant improvement landscape — ANONYMOUS

Purpose 1: providing independently assessible AD functionality — ANONYMOUS

Purpose2: providing proof grounds for deadly AD events — ANONYMOUS

******

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824251