# i4Driving Data Management Plan

D8.4 | 28.04.2023


i4Driving
integrated 4D driver modelling under uncertainty

# i4Driving Data Management Plan

| Project Acronym | Grant Agreement # | Project Title | Deliverable Reference # | Deliverable Title |
|---|---|---|---|---|
| i4Driving | 101076165 | Integrated 4D Driver Modelling under Uncertainty | 8.4 | Data management Plan |

**AUTHORS**



| | |
|---|---|
| Maria Rodrigues | Panteia |
| Lucia Schlemmer | Panteia |



**DISSEMINATION LEVEL**

| | | |
|---|---|---|
| X | P | PUBLIC |
| | C | CONFIDENTIAL |

Funded by the European Union

This project has received funding from the European Union's Horizon Europe programme, under grant agreement No 101076165.

**Disclaimer**

# Version History

| Revision | Date | Authors | Organisaton | Description |
|---|---|---|---|---|
| V0.1 | 23.03.2023 | Lucia Schlemmer | Panteia | First draft |
| V0.2 | 14.04.2023 | Vincenzo Punzo | UNINA | Feedback on first draft |
| V0.3 | 20.04.2023 | Lucia Schlemmer | Panteia | Revision – second draft |
| V0.4 | 26.04.2023 | Vincenzo Punzo | Panteia | Feedback on second draft |
| V1.0 | 28.04.2023 | Maria Rodrigues | Panteia | Final revision |

## Executive Summary

This deliverable presents the first version of the i4Driving Data Management Plan (DMP) (deliverable 8.4). The purpose of this DMP is to provide an overview of all datasets collected and generated by the project and to define the i4Driving consortium's data management policy used regarding these datasets and describes:

- The guiding principles for data management in the project;
- An overview of what data will be gathered and processed in the project;
- How data will be stored and processed according to the H2020 FAIR Data Management principles, making data: finable, accessible, interoperable, and reusable;
- Resource allocation: The costs of making data FAIR in this project; and
- Data security, storage and backup: How we intend to keep the data secure and store the data.

The purpose of the DMP is also to give instructions on naming conventions, metadata structure, repositories and how to make public data available. The DMP will evolve during the lifespan of the project. The next versions will go into more detail regarding the datasets collected and produced by the i4Driving project and include updated instructions for how to access open data.

# Contents



**List of Figures**



**List of Tables**

# 1 Introduction

The vision of i4Driving is to lay the **foundation for a new industry-standard methodology** to establish a **credible and realistic human road safety baseline** for virtual assessment of CCAM systems. The two central ideas we propose are (1) a multi-level, modular and extendable simulation library that combines existing and new models for human driving behaviour; in combination with (2) an innovative cross-disciplinary methodology to account for the huge uncertainty in both human behaviours and use case circumstances. This rigorous treatment of the uncertainty is crucial to assess how much of our confidence in model inputs, parameters, and structure is justified. It also makes explicit how experts from different disciplines judge the outcomes and how justified the underlying assumptions really are.

## 1.1 Deliverable scope and structure

This Data Management Plan (DMP) provides an overview of the approach to data management in i4Driving. It describes the types of data that will be generated or gathered during the project, the standards that will be used (findable, accessible, interoperable and re-usable (FAIR), in accordance with the "Guidelines on FAIR Data Management in Horizon 2020"), the ways how the data will be exploited and shared and how the data will be stored and preserved during and after the project lifecycle. The Project Data term refers to all proprietary data generated within the project operations and transactions, e.g., documents, reports and related information, and, more importantly, user data obtained, possessed, and processed in the context of the provided services.

This document is the first version delivered in month 6 of the project and will be updated during the lifecycle of the project, discussed in section 1.2 below. As such, it is considered a living document that will be reviewed throughout the project lifespan. This document constitutes the first version of the i4Driving DMP. The data presented in this document are drafted during the initial phase of the project (completed by the 30th of March 2023). At this point, the documentation is constrained in reflecting the intentions of the project partners regarding developing the overall project datasets. Therefore, updated versions will elaborate on such information in greater detail.

## 1.2 Data management plan update schedule

As mentioned above, the i4Driving DMP is considered a living document and will be updated in month 18 and 36. These updates correspond with the two large R&D cycles of the project and are also in line with the two periodic reviews. In the **first cycle** (M1 - M18) R&D will be based on **existing evidence** (WP1: gathering evidence from existing data) and on multi-disciplinary knowledge about human driver behavioral modelling available among partners. In the **second cycle** (M19 - M36), R&D will be based on **experimental data collected** in the project (WP3: driving simulator experiments; WP5: field lab experiments).

The first DMP (the document at hand), submitted in month 6, marks the completion of several deliverables, such as D1.1 'Methods to harmonise data on human driving performance from different datasets', D1.3 'Methods to extract statistically significant relationships between human/external factors and driver behavioral mechanisms, in uncritical and critical situations', D2.1 'i4Driving Framework design & modeling and coding design principles', D3.1 'Validated ethics assessment plans and approval from ethics committees' and D5.1 'Validated ethics assessment plan and approval from ethics committees'. Since a significant part of the initial project phase will have been completed by that time, the first version of the DMP (v1.0) has been scheduled for delivery in M6 of the project.

The second DMP is scheduled to be delivered in Month 18 of the project, halfway to completion. At this point, the first R&D cycle will have been completed and the second cycle set to begin. At this point in time, various tasks and deliverables will have been completed, including D1.6 'Methodology and results: relevant use cases and safety-critical scenarios', D4.1 'Critical review of state-of-the-art techniques to model drivers' heterogeneity', D3.2 'Experimental setup for the driving simulator experiments', D5.2 'Evaluation criteria

and detailed description of field experiments', D1.2 'Harmonised, annotated, and processed data in usable format', D1.4 'Open-source library of data mining techniques', D1.5 'Casual relationships between human/external factors and human driving behaviors: modelling requirements & framework of testable hypotheses', D3.3 'Software for automatically increasing the criticality of scenarios', D2.4 'Integrated LC with social interactions and HF models – theory and principles', D3.4 'Map of the heterogeneity of human/external factors into driving behaviour performance', D5.3 'Experimental data relevant to model development', D6.1 'Implementation framework of the scenario-based evaluation workflow' and D6.3 'Open-source GitHub evaluation software toolchain'.

The third version of the DMP is anticipated at the end of the project (M36), along with the due dates of several deliverables and the final reporting period. At that time, the final results will be summarised and presented in a final DMP.

Figure 1 highlights a graphical representation of the DMP update evolvement along the project timeline with an indication of the main milestones. The projected plan is subject to adjustments depending on the progress of the actual work.

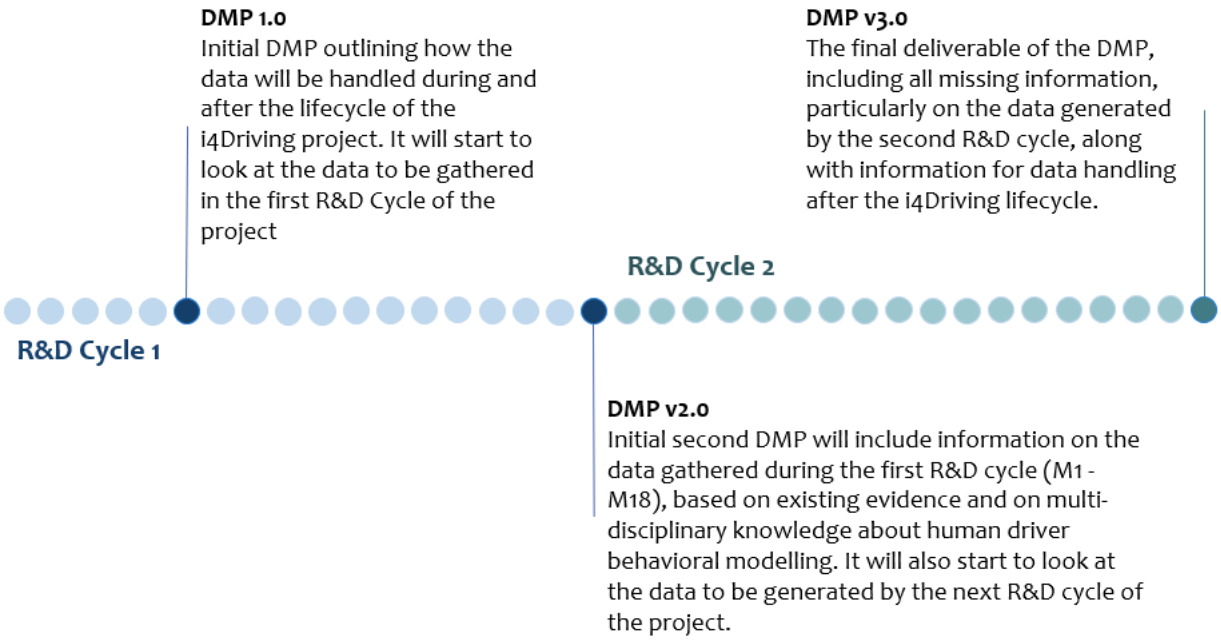


Figure 1. DMP Updates

*Source: Author's own*

## 1.3 Deliverable structure

The deliverable consists of the following chapters:

- Chapter 1 is the introductory section which presents the purpose and structure of the document;
- Chapter 2 refers to the data used/ generated during the project;
- Chapter 3 presents the i4Driving data management framework and the way that the 'FAIR' principles will be implemented and followed during and after the project lifecycle; and
- Chapter 4 provides some concluding remarks regarding the data management plan.

## 1.4 Abbreviation list

Table 1 presents the main abbreviations used in this document.

Table 1. Main acronyms

| Acronym | Full Name |
|---|---|
| CINEA | European Climate, Infrastructure and Environment Executive Agency |
| EC | European Commission |
| GA | Grant Agreement |
| PC | Project Coordinator |
| WP | Work Package |
| WPL | Work Package Lead |
| TL | Task Leader |
| SC | Steering Committee |
| AB | Advisory Board |
| H2020 | Horizon 2020 |
| DMP | Data Management Plan |
| FAIR | Findable, Accessible, Interoperable and Re-usable |
| GDPR | General Data Protection Regulation EU regulation (regulation (EU) 2016/679) on the protection of natural persons regarding the processing of personal data and on the free movement of such data. |
| Personal data | Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information which collected together can lead to the identification of a particular person, also constitute personal data. Personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the law. Personal data that has been rendered anonymous in such a way that the individual is not or no longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible |

# 2 Data summary

Data management constitutes an integral part of the work conducted throughout the i4Driving project lifecycle and refers to the collection, storage, and use of data in a secure, efficient, and cost-effective manner. The purpose of the current Data Management Plan (DMP) deliverable is to provide relevant information concerning the data that will be collected and used by the project partners and, eventually, be openly accessible. In this document, the term 'data' refers to five main categories of information that will be used and/or produced during the project, including:

1. Datasets from the data collection from WP1 and during the simulations;
2. Project deliverables that will be openly accessible to people interested in the project insights;
3. Dissemination materials, including workshops, videos, blogs, presentations, posters and flyers;
4. Software artefacts, such as open-source code, produced by the various components that make up the i4Driving system and accompanied by the appropriate documentation; and
5. Scientific publications referring to the procedures and/or the results of the research conducted.

While the focus of the first version of the DMP is mainly on data collected, the next version will also report on data produced in the context of the project and non-sensitive data that can be made publicly available in open data repositories and registered at relevant catalogues.

## 2.1 Overview of the data

This section provides an overview of the datasets currently expected to be used and generated in i4Driving and their planned accessibility. As mentioned, the expected datasets gathered/produced will develop and grow as the project evolves. Thus, some information concerning the datasets remain unknown at this time, e.g. size of the datasets. An updated version of the datasets will be provided at the end of the project.

There are two main model development cycles in the project:

- The **inner model development process** (WP2 - Robust 4D human driver models under uncertainty), where **existing datasets** from different sources (L3PILOT, HI-DRIVE, UDRIVE, levelXdata, high/in/roun/exi-D, NGSIM, SafetyPool™, SHRP2 NDS, NADS ADS for Rural America, University of Tongji's data, SwissRE data) will be harmonised (by WP1) and used for model development. This inner model development process will move from conceptualisation to calibration and validation, through computer code implementation.
- The **outer model development process** (WP1, WP3, WP4, WP5) is fed by **experimental data collected at a later project stage** using the seven driving simulator facilities in our consortium (at VTI, CTAG, UNINA, TUM, NADS, TJU and UQ).

### 2.1.1 Existing datasets

The table below provides an overview of the datasets that i4Driving expects to use for the project.

Table 2. List of existing datasets to be used in i4Driving

| Type of data | Description | Origin | Formats | Expected size | Utility[1] |
|---|---|---|---|---|---|
| L3PILOT | Naturalistic and FOT data from EU projects. The L3Pilot Open Data contains processed data collected during the Piloting of pre-series automated prototype vehicles on public European roads. The dataset contains driving data in the form of performance indicators derived for all instances of certain driving scenarios such as Car Following or lane changes. Furthermore, it contains data from the questionnaires handed to both professional and ordinary driver piloting the vehicles. | https://l3pilot.eu/data | CSV | TBC[2] | WP2, WP4 |
| HI-DRIVE | Naturalistic and FOT data from EU projects. HI-DRIVE project participant and autonomous vehicle evaluation data resulting from track testing. Still pending definition by WP leaders and confirmation of the possibility of using them in the i4Driving project. | Hi-DRIVE | TBC | TBC | WP2, WP4 |
| U-DRIVE | Naturalistic and FOT data from EU projects. U-DRIVE Data is a comprehensive naturalistic driving data collected in a project sponsored by the European Commission. The dataset includes 38,157 hours of passenger car data collected from 192 drivers across five countries (Germany, France, the Netherlands, Poland, and the United Kingdom), 14,503 hours of truck data collected from 46 drivers in the Netherlands, and 497 hours of powered two-wheeler (scooter) data collected from 39 drivers in Spain. All data is analysable and has been meticulously collected and curated to provide valuable insights into driving behaviour and patterns across Europe. TUM intends to use this data to calibrate and validate the i4Driving human driver model in WP2. | Available at UDRIVE CDC (Chalmers) and PDC's (SWOV/ CEESAR/ Loughborough)[3] | SQL database | TBC | WP2, WP4 |
| levelXdata | Urban and freeway trajectory data sets. levelXdata provides high-quality naturalistic traffic and scenario data for a wide range of applications. levelXdata is a product portfolio of fka GmbH and includes:<br>• Precise trajectory data collected from an aerial perspective with a high level of completeness | https://levelxdata.com/ | TBC | TBC | WP2, WP4 |

[1] Who will use this data
[2] TBC stands for "To Be Confirmed"
[3] Data can be accessed by i) Contracts with participants, ii) EU / National legislation (GDPR), iii) UDRIVE Data Protection Concepts, iv) SWOV protocols related to ISO27001 & NEN7510 certification, v) PDC SWOV in a protected environment in SWOV premise (no remote access).

| Type of data | Description | Origin | Formats | Expected size | Utility[1] |
|---|---|---|---|---|---|
| | • Application-specific datasets from large, diverse trajectory database<br>• Derived, enriched and statistical data<br>• Scenario detection and scenario datasets<br>• Precise digital maps for context and re-simulation<br>• 3D representations for simulation<br>• Data as a service<br>• Drone-based tracking as reference for testing, e.g. vehicle-based sensor sets, scenario-based testing, etc.<br>• Analysis services and project support for your use case | | | | |
| high/ in/ roun/ exi-D | Urban and freeway trajectory data sets. The exiD dataset is a new dataset of naturalistic road user trajectories recorded at exits and entries of highways in Germany. Using a drone, typical limitations of established traffic data collection methods like occlusions are overcome. The trajectory for each road user and its type is extracted. | https://www.exid-dataset.com/ | TBC | TBC | WP2, WP4 |
| Next Generation Simulation (NGSIM) Open Data | Urban and freeway trajectory data sets. Researchers for NGSIM program collected detailed vehicle trajectory data on southbound US 101 and Lankershim Boulevard in Los Angeles, CA, eastbound I-80 in Emeryville, CA and Peachtree Street in Atlanta, Georgia. Data was collected through a network of synchronised digital video cameras. NGVIDEO, a customised software application developed for the NGSIM program, transcribed the vehicle trajectory data from the video. This vehicle trajectory data provided the precise location of each vehicle within the study area every one-tenth of a second, resulting in detailed lane positions and locations relative to other vehicles. | https://datahub.transportation.gov/Automobiles/Next-Generation-Simulation-NGSIM-Vehicle-Trajector/8ect-6jqj | TBC | TBC | WP2, WP4 |
| CommonRoad | CommonRoad is a collection of composable benchmarks for motion planning on roads, which provides researchers with a means of evaluating and comparing their motion planners. | https://commonroad.in.tum.de/<br>TUM | TBC | TBC | WP2, WP4 |
| SafetyPool™ | The SafetyPool™ scenario database is developed and hosted at the WMG, it is the world's largest database with over 250,000 scenarios for a diverse set of ODDs. The database incorporates an ODD and behaviour-based approach for sorting and searching of scenarios. It also offers mapping between real world route and scenarios to find the suitable execution | https://www.safetypool.ai/database<br>WMG | TBC | TBC | WP2, WP4 |

| Type of data | Description | Origin | Formats | Expected size | Utility[1] |
|---|---|---|---|---|---|
| | locations for the scenario. A flexible API is available for external communication with the database. Within the i4Driving project, SafetyPool™ scenario database will be used to host scenarios. | | | | |
| SHRP2 NDS | Naturalistic and FOT data from international initiatives. SHRP2 safety data consists of two large databases; the naturalistic driving study (NDS) database and the roadway information database (RID). Through video and other recording devices, the NDS compiled an unprecedented amount of data about actual driver behaviour during every trip taken by 3,147 volunteer drivers (ages 16-90+) over a 1- or 2-year period. The data includes detailed video of the driver and the roadway, as well as data on the vehicles' speed, acceleration, braking, and other manoeuvres. Information such as seatbelt use and the presence of alcohol is also available. | https://insight.shrp2nds.us/ | TBC | TBC | WP2, WP4 |
| NADS ADS for Rural America | Naturalistic and FOT data from international initiatives. | TBC | TBC | TBC | WP2, WP4 |
| University of Tongji's data | Naturalistic and FOT data from international initiatives. Test drivers' free driving behaviour in sunny days, foggy days with 80m-visibility, foggy days with 40m-visibility, and foggy days with 15m-visibility. The experiment does not set front cars. The test driver will go through three continuous processes: the starting acceleration, smooth driving, and deceleration to stop, on a two-way and four-lane highway with a speed limit of 120km/h. The experiment utilizes a driving simulator with 8 degrees of freedom | TJU | TBC | TBC | WP2, WP4 |
| SwissRE data | Insights from data analysis of near-critical events. We will be providing risk insights and risk-based recommendations based on our global knowledge of motor market | SWRE | TBC | TBC | WP2, WP4 |
| VTI Awareness and attention dataset | Data set on awareness and attention | VTI, can only be used by VTI | TBC | TBC | WP1, WP2, WP4 |

### 2.1.2 Generated datasets

The table below provides an overview of the datasets that i4Driving expects to generate.

Table 3. List of datasets to be generated in i4Driving

| Type of data | Description | Origin | Formats | Expected size | Utility[4] |
|---|---|---|---|---|---|
| CTAG data (still to be defined by WP leaders) | Participant evaluation data resulting from simulator testing (WP3) and vehicle data resulting from track testing (WP5), according to the project requirements. | Simulator test results | TBC | TBC | TBD |
| AIMSUN Data | Model outputs: we are looking to define new parameters and parameter values from data provided to us by driving simulators/ on-road instrumented vehicles. In addition, we will generate data on vehicle space-time trajectories as part of the output from our simulation software. this can be generated at specified time intervals and can also be supplemented with parameters such as TTC, braking levels etc., as desired. | Model outputs | TBC | TBC | TBD |
| TH-AB Data | **D**ataset calibration and **validation** for Lane Change under weather conditions (datasets as provided) | | TBC | TBC | WP2, WP4, WP6 |
| VTI data | Data collected during field experiments and simulator experiments | Field and simulator experiments | Depends on scenarios and measuring parameter defined in WP1 | Depends on scenarios and measuring parameter defined in WP1 | WP2, WP4 |
| UNINA Data | Experimental data collected at a later project stage using the driving simulator facilities | Simulator test results | TBC | TBC | TBD |
| TUM Data | Experimental data collected at a later project stage using the driving simulator facilities | Simulator test results | TBC | TBC | TBD |
| UQ Data | Experimental data collected at a later project stage using the driving simulator facilities | Simulator test results | TBC | TBC | TBD |
| SWRE Data | We expect to generate a dataset that encompasses all the insights and data sources used throughout the project. | TBC | TBC | TBC | TBC |
| TUD Data | TUD does not intend to generate any data (except the simulation outputs) within the context of the project, but may collect data (trajectories, driving simulator experiments, etc.) within other (related) projects. These will—if possible—be made available. | TBC | TBC | TBC | TBC |

[4] Who will use this data

# 3 i4Driving Data Management Framework

## 3.1 Legal Framework

As of 2018, the General Data Protection Regulation (GDPR) is applicable in all Member States. GDPR updates and covers the protection of natural persons regarding the processing of personal data and on the free movement of such data. GDPR grants individuals a set of rights that must be protected by any actor who processes personal data.

**Permissions for collecting and handling personal data**

All data collected and generated in the project will be done in accordance with the GDPR and applicable ethical standards and requirements in the respective countries of the data collection, as well processed and handled securely and in line with applicable rules and regulations on privacy and data protection. This is covered in more detail in the following deliverables:

- 3.1. Validated ethics assessment plans and approval from Ethics committee
- 5.1. Validated ethics assessment plan and approval from Ethics committee

## 3.2 FAIR principles in i4Driving

This chapter sets out the **findability, accessibility, interoperability and reuse** (FAIR principles) of the project data. Thus, anyone will be free to access and reuse – share (i.e., copy and redistribute the material in any medium or format) and adapt (i.e., remix, transform and build) for any purpose, even commercially – as long as the attribution terms are complied with. An overview of each of the FAIR principles is provided below.

### 3.2.1 Making data findable

The first principle is concerned with making i4Driving project data 'findable'.

#### *3.2.1.1 Repository*

In line with Open Science practices, we guarantee that our data and processes are transparent and accessible by: (1) providing open access to our data analysis and methods; (2) sharing and making public our approaches, training material and guidelines, (3) involving an open peer review with various stakeholders throughout the course of the project at different levels of maturity, and (4) encouraging re-use of data, models and software produced in project.

All data/software outputs from the i4Driving project will be **published in and found on GitHub** in the i4Driving community. This will ensure that access to research data, project results, and any scientific publications is maximised and re-used. In some cases, datasets and/or publications will also be published on Zenodo, arXiv, Open Research Europe, ResearchGate and the i4Driving project website. Specific details regarding the type of each type of deliverable are provided in Table 4 below.

Table 4. Deliverables and open science practices

| Type of deliverable | Open science function | Repository/platform |
|---|---|---|
| Working paper/ draft publication | Feedbacks/ comments/ discussions | Website, ResearchGate |
| Draft/ final publication | Invited open reviews | Open Research Europe |
| Pre-print publications | Open access to paywall articles | arXiv, Zenodo, Website |
| Open access journal articles | Open access | Website, ResearchGate |

| | | |
|---|---|---|
| Presentations/ recordings | Communication with general audience | YouTube, GitHub, website |
| Software and datasets | Reproducibility of research, re-use | GitHub, Zenodo, website |
| Miscellaneous supplementary information | Reproducibility of research | GitHub, website |
| Other research outputs | Reproducibility of research, re-use | GitHub, Zenodo, website |

Further, other general actions will be carried out in the project:

- All deliverable/research outputs, including datasets, that are placed in a repository will be promoted via various dissemination channels.
- Consortium repositories are high-profile trusted repositories, usually based in their universities. Links to aggregators, such as openAIRE, will be made to increase findability. This in addition to publishing in GitHub, which supports versioning and is more suited to encouraging collaboration.
- i4Driving will appoint an Open Science advocate (WP7) to encourage, advice and track each partner's compliance and performance of Open Science activity. Monitoring will include Altmetrics.

#### 3.2.1.2 *Metadata*

All deliverable/research outputs, including datasets, that are placed in a repository will be accompanied with sufficient CC0 metadata. In addition, DOIs will be issued for each output type published in the GitHub repository, as well as other repositories where necessary, and published in deliverables and journals.

The list below describes the metadata that will be provided for each dataset in SharePoint:

- File name
- Date
- Version
- File type
- Description
- WP (Work Package) number
- Responsible person
- Lead partner / Data Controller
- Dissemination level

#### 3.2.1.3 *Approach to search keywords*

The i4Driving project will use search keywords to optimise the data findability and possibilities for re-use. The Data Controllers will be responsible for uploading public datasets that they have generated and to assign specific keywords relevant to these datasets. Dataset specific keywords must be descriptive to the content of the dataset. For example, a dataset containing information on naturalistic driver behaviour should be tagged with corresponding keywords such as, "naturalistic driver behaviour". In addition, the project has defined a set of general keywords that should apply to all public datasets, scientific publications and public deliverables. These are as follow:

- Autonomous driving
- Human driving behaviour
- Automated vehicles
- Automated driving systems
- Human driver models

- Human road safety baseline

#### 3.2.1.4 *Naming conventions*

Datasets will be named using the following naming conventions:

DS_PilotCode_DataCategoryNr_DataController_Description_H2020_Acronym_DataNr

Explanation of the naming convention:

- "DS" stands for dataset
- The test site identification codes are as follows:
  - Spain: ES
  - Germany: DE
  - Sweden: SE
  - Italy: IT
  - America: US
  - China: CN
  - Australia: AU
- "DataController" refers to the short name of the partner responsible for the dataset
- "Description" refers to a short description of the content of the dataset (see the example below)
- H2020_Acronym refers to the project acronym
- "DataNr" is the version number generated by the research metadata list in SharePoint

Example of dataset name: DS_ES_CTAG_Participant-Evaluation-Data_H2020_i4Driving_0001

#### 3.2.1.5 *Versioning*

GitHub and Zenodo provides versioning of all datasets uploaded to their communities, which will allow us to edit and update the uploaded datasets after they have been published. This also allows us to cite specific versions of an upload and cite all versions of an upload.

### 3.2.2 Making data openly accessible

The H2020 Open Access Mandate aims to make research data generated by H2020 projects accessible with as few restrictions as possible, but also accept protection of personal or sensitive data due to privacy concerns and/or commercial or security reasons.

As mentioned above, all data/software outputs from the i4Driving project will be **published in and found on GitHub** in the i4Driving repository and made openly available, free of charge. In some cases, datasets and/or publications will also be published on Zenodo, arXiv, Open Research Europe, ResearchGate and the i4Driving project website. Publications and underlaying data sets will be linked through use of persistent identifiers (DOI versioning).

Datasets with dissemination level "confidential" (non-anonymous datasets) will not be shared due to privacy concerns. In addition, there may be some datasets that are restricted due to protection for commercial exploitation. If such cases arise during the project, this will be specified in the final version of the DMP.

### 3.2.3 Making data interoperable

The data produced in the i4Driving project will be made interoperable, allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. Interoperability for each of the relevant project deliverables is summarised in Table 5 below.

Table 5. Interoperability of i4Driving datasets

| Deliverable | Type | Interoperability | Responsible |
|---|---|---|---|
| D1.2. Harmonised, annotated, and processed data in usable format | Experimental data (large data sets) | Use of non-proprietary data formats with complete and fully annotated metadata | TUM |
| D1.4. Open-source library of data mining techniques | Source code | Open-source programming language (e.g., GPL/BSD-like license) interoperable with other simulation software | CNR |
| D2.2. Suite of unit tests & simulation models (cases) for model development & validation. | Software & experimental data | same as D1.2 and D1.4 (integration in OTS) | TUD |
| D2.3. Incremental versions of i4Driving/software library in opentrafficsim.org | Source code | Same as D1.4 (integration/interfacing with CommonRoad, OTS, CARLA, Aimsun) | TUD |
| D3.3. Software for automatically increasing the criticality of scenarios | Source code. | Same as D1.4 (integration/interfacing with CommonRoad, OTS, CARLA, Aimsun) | TUM |
| D4.4. Open-source library of techniques to encode drivers' heterogeneity from data into models | Source code | same as D2.3 | UNINA |
| D4.5. Open-source library of validated probabilistic human driver behavioural models | Source code. | Same as D2.3. | UNINA. |
| D5.3. Experimental data relevant to model development. | Experimental data | Same as D1.2. | TUM |
| D5.4. Experimental results relevant to model validation. | Experimental data | Same as D1.2. | TUM |
| D6.3. Open-source evaluation software toolchain | Source code | D1.4 (integration in CARLA). | WMG |

### 3.2.4 Increase data re-use

i4Driving will enable third parties to access, mine, exploit, reproduce and disseminate public data sets free of charge and will regulate this by using Creative Commons Licences.

i4Driving will make use of the Creative Commons Attribution 4.0 International (CC-BY-4.0). This gives recipients maximum freedom to use the datasets of the licensor. Recipients redistributing the work must give credit to the original author of the work and state any changes made, including a URL or link to the original work, this CC-BY licence and a copyright notice (European Commission, n/d).

The data produced and/or used in the project will be useable by third parties after the end of the project. None of the data is foreseen to be restricted at this point in time.

## 3.3 Allocation of resources

### 3.3.1 Costs

i4Driving uses standard tools and a free of charge research data repository. The costs of data management activities are limited to project management costs and will be covered by allocated resources in the project budget.

Long-term preservation of the public data is ensured through GitHub, Zenodo, the Project website, ResearchGate, Open Research Europe, arXiv and YouTube.

i4Driving has budget allocated for open access scientific publications for UNINA, THAB, CNR.

Other resources needed to support reuse of data after the project ends will be solved on a case-by-case basis.

### 3.3.2 Data Management

The overall responsibility for data management lies with the project coordinator, Ms Maria Rodrigues from PANTEIA.

Supporting the coordinator is a data management team consisting of the Data Controllers for each pilot site (CTAG, VTI, UNINA, TUM), the WP Leaders of WP3 and WP5 on ethics (VTI and TUM), the i4Driving Open Science advocate (WP7 - WMG) and the task leader of Task 8.4 Research Data Management (PAN).

Each Task Leader responsible for producing the individual deliverables will be the directly responsible for ensuring compliance with FAIR principles and the Data Management Plan, as well quality assurance.

## 3.4 Data security, storage and backup

In this chapter, the security features of the research data infrastructure used to store and handle data in the i4Driving project are described.

### 3.4.1 Data security as specified for Panteia SharePoint

Panteia's SharePoint is the online collaboration platform used the i4Driving project. A dedicated project site has been established on this platform, accessible only by the partner representatives in the consortium. Furthermore, a dedicated folder for research datasets has been set up, allowing for stricter access control than the main project site. Only anonymous datasets will be uploaded to this SharePoint folder.

The i4Driving Sharepoint site has the following security settings:

- Access level: Restricted to persons (project members only). Further access restrictions on specific folders is enabled.
- Encryption with SSL/TLS protects data transfer between partners and the SINTEF SharePoint site.
- Threat management, security monitoring, and file-/data integrity prevents and/or registers possible manipulation of data.

Documents and elements in the Panteia SharePoint are stored in Microsoft's cloud solutions.

### 3.4.2 Data security as specified for the data repository platforms

Open results deposited in the GitHub repository are covered by GitHub's privacy protection. For more information, see: https://github.com/security.

Open results deposited in the Zenodo repository are stored in CERN's EOS service (http://eos.web.cern.ch/content/about-eos). The servers are managed according to the CERN Security Baseline for Servers. For more information see http://about.zenodo.org/infrastructure/.

### 3.4.3 Storage and back-up

Daily back-ups are handled by PANTEIA's IT operations contractor. As a baseline, all project data will be stored for 5 years according to PANTEIA's ICT policy, unless otherwise agreed in contracts and data processing agreements.

# 4 Concluding remarks and next steps

The approval and release of this deliverable within the consortium constitutes a formal commitment by partners to adhere to the data management strategy and the procedures it defines. When the deliverable is formally approved by the European Commission, this constitutes confirmation that the procedures are considered by the European Commission to be adequate. As coordinator of the i4Driving project, Panteia will ensure that any data management issues which may arise during the project will be handled appropriately and in a transparent and fair manner.

The DMP is a living document that will expand as the project evolves and new information on data collection, generation and handling arise. Day to day data management will happen through the online tools described in this document, and through continuous collaboration between the coordinator, the Data Controllers, the relevant WP and task leaders. A revised and extended version of this DMP will be prepared towards the middle and end of the project to reflect the current status of data management in the project.

# 5 References

European Commission. H2020 Programme. Guidance for the classification of information in research projects. Version 2.1. 26 October 2016

European Commission. H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. Version 3.0, 26 July 2016

European Commission. Horizon 2020 Programme. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. Version 3.2, 21 March 2017

European Commission. Creative Commons Attribution 4.0 International (CC-BY-4.0). n/d. https://joinup.ec.europa.eu/licence/creative-commons-attribution-40-international-cc-40

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC