



D7.2 Data Management Plan

Project Number:	690705
Classification	Public
Deliverable No.:	D7.2
Work Package(s):	WP7
Document Version:	Vs. 1.0
Issue Date:	22.12.2016
Document Timescale:	Project Start Date: September 1, 2016
Start of the Document:	Month 3
Final version due:	Month 4
Compiled by:	Jürgen Pichen, ULM
Authors:	Jürgen Pichen, ULM Nicola Fricke, ULM Martin Baumann, ULM
Technical Approval:	Fabio Tango, CRF
Issue Authorisation:	Andreas Lüdtke, OFF

© All rights reserved by AutoMate consortium

This document is supplied by the specific AutoMate work package quoted above on the express condition that it is treated as confidential to those specifically mentioned on the distribution list. No use may be made thereof other than expressly authorised by the AutoMate Project Board.



Table of Contents

1	Introduction.....	5
2	Handling of research data	5
2.1	During the project.....	5
2.2	After the project	6
3	What data will be collected	6
4	Which methodology and standards	7
4.1	Methodology	7
4.2	Standards	8
4.2.1	Data standards.....	8
4.2.2	Variable standards (Annotation template)	9
5	Data preservation	9
6	Appendix.....	10
6.1	Data set template	10
6.2	Annotation template.....	10



1 Introduction

This Data-Management-Plan (DMP) has been created following the template of the "Guidelines on Data Management in Horizon 2020" (Version 3.0, 26.07.16). Furthermore, the consortium of AutoMate will comply with the data protection requirements of the Directive 95/46/EC of the European Parliament (http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf). This data management document is a first version, which might be updated throughout the duration of the project. Parts of the management plan are two templates explained in more detail below and attached to this document. The first template is the "data set template" (see table 1 and Appendix 6.1), in which the partners collecting the data describe each data set. The second template describes every variable of a dataset in a variable list named "annotation-template" (see table 3 and Appendix 6.2). Additionally there is a list in which all the data types for data collection are defined (see table 2).

2 Handling of research data

2.1 During the project

Partners in AutoMate who will collect data are located in work packages (WP) 2, 4 and 6. Specifically, in WP2 the partners will validate the driver and situational models, in WP4 the partners will validate the driver-automation interaction concept with user studies, in WP6 partners will evaluate the overall TeamMate technologies using simulators and instrumented vehicles (in real road traffic if possible otherwise using a test track).

All partners will create a repository of their collected data. Additionally, copies of the generated data, which is not confidential, will be made accessible via a research repository through the collaboration platform AjaXplorer, which is used in the AutoMate project. The data will be archived as raw, de-identified data avoiding proprietary data formats whenever possible. Additionally, annotations following the annotation template (see table 3) will allow sharing of data amongst all project partners. This will ensure that all partners can access data from other partners to ease collaboration.

Besides storing the raw and annotated data on the research repository (AjaXplorer), selected and further processed and de-identified data will be made publicly available on the project-website according to the AutoMate Grant Agreement 29.3.a.



All data which is collected from participants during the studies in WP2, WP4 and WP6 will be treated according to the ethics principles and processes defined in WP9.

2.2 After the project

In order for the results of AutoMate to be preserved, re-used by other projects and validated, selected data will be archived and stored on publicly available databases. The specific database will be determined throughout the project duration and might be linked to a platform such as the OpenAire platform (<https://www.openaire.eu>).

3 What data will be collected

In the AutoMate project several types of data will be collected within various WP. First of all, experimental and observational data will be collected within WP2 and WP 6, which will be available in multiple formats and, as raw and derived data. Secondly, qualitative data for the validation of the human-machine-interface aspects in WP4 and overall implemented TeamMate concepts in WP 6 will be collected via behavioural data, as well as questionnaires and interview data. Thirdly, in WP6 additional data will be recorded through the TeamMate sensor platform.

There is a template for the collected data, which each partner needs to fill out whenever they are collecting data (see table 1). This template will develop throughout the project and reflects the current status within the consortium about the data that will be produced.

Table 1: Data set template

Name/Identifier	
Partners/Department	Which partner will collect/analyse the data?
Deliverable (data for concrete deliverable?)	Is there a concrete deliverable connected to the data?
Format (.csv, .mp4 etc.)	Which format does the data have. Is it a video, excel file etc.?
Description (observational data recorded during experiment, sensor data etc.)	A short description about what kind of data was collected.
Comments (if the structure of the data is not self-explaining)	Further comments to understand the collected data (variable names etc.).



Dimension (How big is the data file)	How big is the collected data more or less (1Tb, 100Mb etc.)?
Access (Open, confidential, partly confidential)	Is all the data open access or is it (partly) confidential?
Location of data (link)	Where is the data saved (all the open data should be saved in one repository)?
Start of data collection	Beginning month of the data collection.
Expected end of data collection	When is the data collection expected to end?
Results/processed data	What are the results or how does the processed data look like?
How will the data be preserved (after project)	Where will the data be saved after the project and who will assure the preservation?
Is the data useable after the project? By whom? (e.g. useable by third parties, useable by project partners, etc.)	Where will the data be saved after the project and who can download it?
How is the data secured?	Which technology/method secures the collected data? How can WP9 be assured?

4 Which methodology and standards

4.1 Methodology

The aim of this DMP is to create a decentralised database where partners are responsible for their collected data. This means that each partner is self-responsible for saving and securing the data locally. Data, which can be shared, should be made available and saved on the AjaXplorer, accessible for all partners.

Every partner involved in collecting data and keeping it tracked in the data set template (see table 1) should check the data for up-to-datedness throughout the project.

In case of data loss or problems concerning the access to data on the central data management server (AjaXplorer), the developer of the database can be contacted (OFFIS). They are also responsible for the security and backup of the data saved on the AjaXplorer, for tracking the free storage space making sure there is enough space throughout the project. When uploading very big data (especially videos) the developer must be contacted to check if there is enough space. Additionally, all big data files should be compressed and/or



split into smaller parts to account for the limitations of the server software (e.g. no resume function for up- or downloads).

When collecting user data, the participant must be informed about the usage of the data according to the data management plan.

4.2 Standards

In this chapter all standards regarding the collected data are defined.

4.2.1 Data standards

There are several standards that should be obeyed (see table 2). These standards prescribe the type of different kinds of data. If it is not possible to save the data in a recommended format or the partner has a reasonable explanation of using another format, it should be discussed in the consortium and an agreement has to be achieved. It will not be possible to use proprietary data types. Data shall be stored as raw as possible. If the further processing of this data needs special software, accessibility for all partners needs to be ensured (e.g. by transferring the raw data into a data format from the data type list such as .csv).

Table 2: Data type list

Kind of data	Standard format
Text	.doc or .txt
dates	European standard (dd-mm-yyyy)
Images	SVG, TIFF or JPEG (in high resolution)
Audio	.wav or .mp3
Video	.avi, .mov or MPEG4
Simulator variables	.csv
Sensor data	tbd
Questionnaire data	.csv
Declarations	Scanned in .pdf format



4.2.2 Variable standards (Annotation template)

For every data set the created variables should be defined and explained. There is a template available on the Ajaxplorer, as well as attached to this document (see Appendix 6.2). This template has six categories (see Table 3).

Table 3: Variable list (Annotation template)

Category	Explanation	Example
Name	The name of the variable, which is meaningful, short and informative.	Latitude
Code name	The name of the variable saved by a program (e.g. simulator).	cor_lat_meas
Format	Data format (with storage space)	CHAR(16)
Range	The possible range of the variable.	0 – 360 degrees
Definition	A detailed description of the variable. Questions to be answered inside this category: <ul style="list-style-type: none"> - What kind of data is this? - How can the data be interpreted? - How is the variable materialised? - Why is the data needed? 	Coordinate Latitude is the angle between the plane of the reference ellipsoid's equator and a normal to the ellipsoid surface. It is formatted by direction, degrees, minutes, decimal seconds (60 24 32.56 N). This item is analogous to the 'Y' value of a rectangular coordinate system.
Confidentiality	The confidentiality of this variable. If it's only partly confidential the person groups permitted to see and use the data should be defined.	Open for everyone

5 Data preservation

To make the collected data reusable for other researchers, the data will be preserved and made available for others. This only includes data that is not



confidential. At this point of the project the estimation of the needed storage capacity for the collected data is not possible. Therefore, the decision on which platform to use for the storage will be made at a later point during the project. The OpenAire platform (<https://www.openaire.eu>) will be considered as a possible option and will be reviewed as soon as the size of the data is definite.

Huge data files (e.g. long videos) should only be stored locally on a server of the partner involved in recording them. All data should be preserved for a period of 10 years following guidelines of good scientific practice (e.g. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/guidelines_research_data.pdf).

6 Appendix

6.1 Data set template



Data set template.xlsx

6.2 Annotation template



Data annotation
template.xlsx