



HORIZON 2020 - ICT-14-2016-1

## AEGIS

Advanced Big Data Value Chains for Public Safety and Personal Security

### WP6 - Dissemination and Communication



## D6.3 – Data Management Handling Plan

Version 1.0

**Due date:** 30.06.2017

**Delivery Date:** 30.06.2017

**Author(s):** Konstantinos Tsatsakis (HYPERTECH)

**Editor:** Konstantinos Tsatsakis (HYPERTECH)

**Lead Beneficiary of Deliverable:** HYPERTECH

**Dissemination level:** Public

**Nature of the Deliverable:** Report

**Internal Reviewers:** Alex Stocker (VIF), Silke Cuno, Dustin Stadtkewitz (Fraunhofer)

**AEGIS KEY FACTS**

<b>Topic:</b>	ICT-14-2016 - Big Data PPP: cross-sectorial and cross-lingual data integration and experimentation
<b>Type of Action:</b>	Innovation Action
<b>Project start:</b>	1 January 2017
<b>Duration:</b>	30 months from <b>01.01.2017</b> to <b>30.06.2019</b> (Article 3 GA)
<b>Project Coordinator:</b>	Fraunhofer
<b>Consortium:</b>	10 organizations from 8 EU member states

**AEGIS PARTNERS**

<b>Fraunhofer</b>	Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.
<b>GFT</b>	GFT Italia SRL
<b>KTH</b>	Kungliga Tekniska högskolan
<b>UBITECH</b>	UBITECH Limited
<b>VIF</b>	Kompetenzzentrum - Das virtuelle Fahrzeug, Forschungsgesellschaft-GmbH
<b>NTUA</b>	National Technical University of Athens – NTUA
<b>EPFL</b>	École polytechnique fédérale de Lausanne
<b>SUITE5</b>	SUITE5 Limited
<b>HYPERTECH</b>	HYPERTECH (CHAIPEKTEK) ANONYMOS VIOMICCHANIKI EMPORIKI ETAIREIA PLIROFORIKIS KAI NEON TECHNOLOGION
<b>HDIA</b>	HDI Assicurazioni S.P.A

**Disclaimer:** AEGIS is a project co-funded by the European Commission under the Horizon 2020 Programme (H2020-ICT-2016) under Grant Agreement No. 732189 and is contributing to the BDV-PPP of the European Commission.

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

© Copyright in this document remains vested with the AEGIS Partners

## **EXECUTIVE SUMMARY**

This deliverable presents the first version of the (open) Data Management Handling Plan for the AEGIS project in month 6 of the project. This Data Management Handling Plan investigates the appropriate methodologies and open repositories for data management and dissemination, and tries to offer through open access as much as information generated by the AEGIS project. Such information would be the scientific publications issued by the project consortium, white papers published, Open Source code generated, mock-up datasets used for supporting the development process etc. The list of research data expected during the project consists of Open-Sourced, trust-related software components; original research data and anonymous user statistics. These datasets are expected to be collected during the implementation and evaluation phase of the project and are therefore subject to change, considering also the definition of the AEGIS business models.

The publishing platforms used are the project website, the BSCW platform, Zenodo for long-term archiving (as proposed by E.C.), and GitHub for open-sourced code. All these platforms can be accessed openly.

## Table of Contents

<b>AEGIS KEY FACTS</b> .....	<b>2</b>
<b>AEGIS PARTNERS</b> .....	<b>2</b>
<b>EXECUTIVE SUMMARY</b> .....	<b>3</b>
<b>LIST OF FIGURES</b> .....	<b>5</b>
<b>LIST OF TABLES</b> .....	<b>5</b>
<b>ABBREVIATIONS</b> .....	<b>6</b>
<b>1. INTRODUCTION</b> .....	<b>7</b>
1.1. PURPOSE OF THE DOCUMENT .....	7
1.2. STRUCTURE OF THE DOCUMENT .....	7
<b>2. METHODOLOGICAL FRAMEWORK FOR DATA MANAGEMENT HANDLING PLAN</b> .....	<b>9</b>
2.1. DMP MANAGEMENT PROCESS .....	12
2.2. RESPONSIBILITIES AND DECISION MAKING .....	14
<b>3. DATA ARCHIVING AND PRESERVING INFRASTRUCTURE</b> .....	<b>15</b>
3.1. BSCW.....	15
3.2. AEGIS PROJECT PORTAL .....	15
3.3. RESEARCHGATE.....	16
3.4. ZENODO PROJECT REPOSITORY.....	16
3.5. GITHUB .....	16
<b>4. DATASETS AND PUBLICATIONS FOR DMP</b> .....	<b>18</b>
4.1. AEGIS PROJECT PUBLIC DELIVERABLES.....	18
4.1.1. <i>D1.1 Domain Landscape Review and Data Value Chain Definition</i> .....	19
4.1.2. <i>D6.1 Plan for Dissemination, Communication and Stakeholder Engagement</i> .....	20
4.1.3. <i>D6.2 Project Website and Web 2.0 Channels</i> .....	20
4.1.4. <i>D1.2 The AEGIS Methodology and High Level Usage Scenarios</i> .....	21
4.2. RESEARCH DATASETS AND SOFTWARE COMPONENTS .....	22
4.2.1. <i>Open Source Software Components</i> .....	22
4.2.2. <i>AEGIS Project Datasets</i> .....	24
4.3. AEGIS SCIENTIFIC PUBLICATIONS.....	27
4.3.1. <i>Quantified Vehicles: Novel Services for Vehicle Lifecycle Data</i> .....	27
4.3.2. <i>Quantified Cars: Novel Digital Services and Business Models Based On Vehicle Data</i> .....	28
<b>5. SUMMARY AND CONCLUSIONS</b> .....	<b>30</b>
<b>APPENDIX A: LITERATURE</b> .....	<b>31</b>

**LIST OF FIGURES**

Figure 1: Research Data Management according to the FAIR principles (*Source: Elsevier*) . 10

Figure 2: Open Access strategy for publications and research data..... 12

**LIST OF TABLES**

Table 1: Timetable for updates of the DMP..... 11

Table 2: List of AEGIS Public Deliverables ..... 19

Table 3: List of Open Source software components ..... 22

**ABBREVIATIONS**

AEGIS	Advanced Big Data Value Chain for Public Safety and Personal Security
D	Deliverable
DoW	Description of Work
H2020	Horizon 2020 Programme
DMP	Data Management Plan
EC	European Commission
FAIR	Findable, accessible, interoperable and reusable
FOKUS	Fraunhofer Institute for Open Communication Systems
IPR	Intellectual Property Rights
PDF	Portable Document Format
XML	eXtensible Markup Language
CC	Creative Commons
DCP	Dissemination and Communication Plan
MVP	Most Valuable Product
HDFS	Hadoop Distributed File System
JDBC	Java Database Connectivity
WP	Work Package
UDF	User-Defined Functions
AAL	Ambient Assisted Living
BISE	Business & Information Systems Engineering
DOI	Digital Object Identifier

## 1. INTRODUCTION

Giving particular attention to research data management, the consortium follows a series of dedicated activities toward disseminating the project outcomes in mass audience as described in the AEGIS Data Management (Handling) Plan.

The analysis (and the respective plan) clearly addresses issues such as the overall methodology for handling the scientific outcomes of the project, the specification of data types that the project generates and/or collects, the standards that will be used, the process of how this data will be exploited and/or shared/made accessible for verification and re-use, the data preservation and maintenance processes etc.

In particular, the Data Management Handling Plan is formulated in accordance with the H2020 guidelines regarding Open Research Data, as AEGIS aims to actively take part in the Pilot of Open Research Data. In alignment with the EC Guidelines for Open Access [1], we clearly define how the scientific publications issued by the project consortium, white papers published, open source code generated, and mock-up datasets generated will be further disseminated to a wider audience.

The overall analysis towards the definition of the AEGIS Data Management Plan is provided in the following sections.

### 1.1. PURPOSE OF THE DOCUMENT

Each project in the EC's Horizon 2020 program has to define what kind of results are generated or collected during the project's runtime and when and how they are published openly. This document describes initially the methodological framework taking into account the guidelines from E.C.

Then, the initial selection of the results that are going to be published or expected to be published in the AEGIS project after the first 6 months is provided. For all the results generated or collected during the AEGIS project a description is provided including the source of the data, the standards and metadata used for data handling, etc.. In addition, the process of how this data will be further exploited and/or shared/made accessible for verification and re-use, the data preservation and maintenance processes are defined.

It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6. Rather, the DMP is intended to be a *living document* in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. This is the 1<sup>st</sup> version of the document which is going to be updated on a regular basis in the context of D6.4 [M18] and D6.5 [M30].

### 1.2. STRUCTURE OF THE DOCUMENT

The document is divided into the following sections

- The first section defines the purpose of the document, its structure, and terms that are necessary to understand it.
- In section 2, we define the methodological framework towards handling the results collected or generated during the project. The overall analysis is inline (actually a summarized version) of the methodology proposed by E.C.

- In section 3, we define the tools to be used in order to ensure that the data will be exploited and/or shared/made accessible for verification and re-use along with the data preservation and maintenance processes
- In section 4, we list all publications and related data that is already or may be generated or collected during the project. For each result we provide - in accordance to the Data Management Guideline (European Commission, 2013) - a short description, the chosen way of open access, and a long-term storage solution.

A summary of project activities towards the definition of the AEGIS Data Management Handling plan is provided as a last section.



## 2. METHODOLOGICAL FRAMEWORK FOR DATA MANAGEMENT HANDLING PLAN

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable, a DMP should include information about the handling of research data during and after the end of the project [1]:

- what data will be collected, processed and/or generated; What kind of data will the project collect or generate, and to whom might they be useful later on?
- which methodology and standards will be applied?; what metadata required to enable data to be found and understood, ideally according to the particular standards of scientific discipline?
- whether data will be shared/made open access? and
- how data will be preserved (including after the end of the project)?; How to archive and preserve the open datasets of the project? Funding bodies ensure that publicly funded research outputs can have a positive impact on future research, for policy development, and for societal change.

More specifically, for Horizon 2020 projects a FAIR DMP [2] template has been designed to be applicable to any project that produces, collects or processes research data. The respective activities defined as part of the methodology, adopted also in AEGIS project are [3]:

- Data Summary
- FAIR Data Principles
  - Making data findable, including provisions for metadata
  - Making data openly accessible
  - Making data interoperable
  - Increase data re-use (through clarifying licenses as defined during project period)
- Allocation of resources
  - Explain the allocation of resources
- Data Security
  - Address data recovery as well as secure storage and transfer of sensitive data
- Ethical Aspects
  - In the context of the ethics management plan of the project
- Other Issues
  - Refer to other national/funder/sectorial/departmental procedures for data management if any

Figure 1 presents the FAIR data principles towards promptly disseminating the data outcomes of a research project.

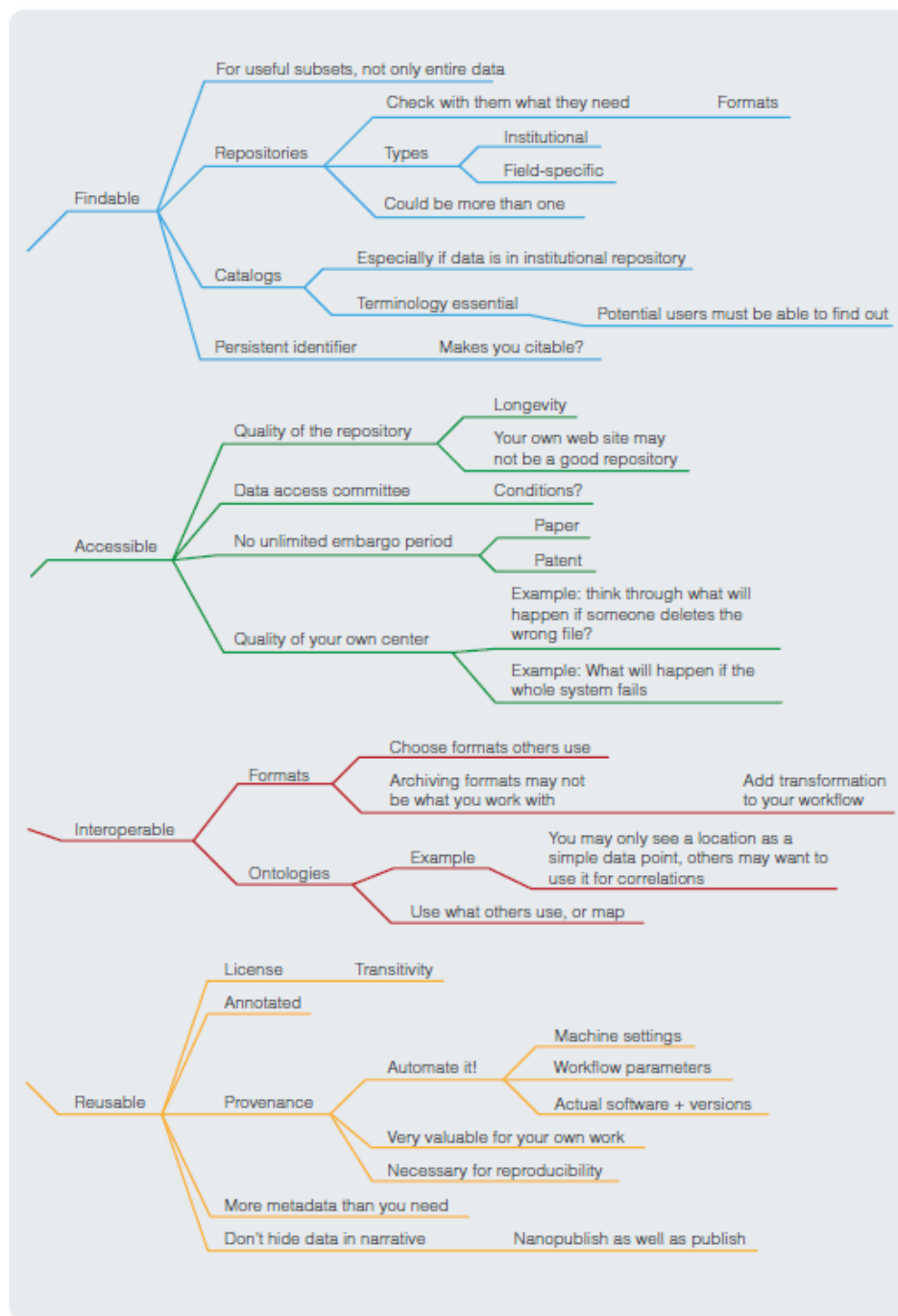


Figure 1: Research Data Management according to the FAIR principles (Source: Elsevier)

A DMP is required for all projects participating in the extended Open Research Data pilot. AEGIS wants to be part of the pilot and thus a preliminary version of the Data Management Plan is provided early in the project. The DMP needs to be updated over the course of the project whenever significant changes arise, such as (but not limited to):

1. new data are generated
2. changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
3. changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

As previously indicated, the Data Management Plan is a living document that should be updated as a minimum in time with the periodic evaluation/assessment of the project. This is the approach followed in the project as the updated versions of the DMP will be available by M18 and M30, according to the following timetable:

DMP Version	Delivery Month	Description
v1	M6	DMP First version
v2	M18	DMP including the definition of the datasets and components developed in the project
v3	M30	DMP Final version

Table 1: Timetable for updates of the DMP

Along with the definition of the datasets, special focus is delivered at the selection of the platform to archive and preserve the datasets. When choosing a repository it is important to consider factors such as whether the repository [4]:

- Gives the submitted dataset a persistent and unique identifier. This is essential for sustainable citations – both for data and publications – and to make sure that research outputs in disparate repositories can be linked back to particular researchers and grants.
- Provides a landing page for each dataset, with metadata that helps others find it, tell what it is, relate it to publications, and cite it. This makes your research more visible and stimulates reuse of the data.
- Helps to track how the data has been used by providing access and download statistics.
- Responds to community needs and is preferably certified as a ‘trustworthy data repository’, with an explicit ambition to keep the data available in the long term.
- Matches particular data needs (e.g. formats accepted; access, back-up and recovery, and sustainability of the service). Most of this information should be contained within the data repository’s policy pages.
- Provides guidance on how to cite the data that has been deposited.

In addition, a main point of the DMP is the definition of the **open access** type over the data. Open Access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean:

- peer-reviewed scientific research articles (published in scholarly journals) or
- research data (data underlying publications, curated data and/or raw data).

Open Access [1] to scientific publications means free online access for any user. The two main routes to Open Access are:

- A. Self-archiving / 'green' Open Access – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository

before, at the same time as, or after publication. Some publishers request that open access be granted only after an embargo period has elapsed.

- B. Open Access publishing / 'gold' open access - an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers. The most common business model is based on one-off payments by authors.

Research data refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form. Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge. The next figure presents the process flow towards defining the open access type in scientific publications and research data.

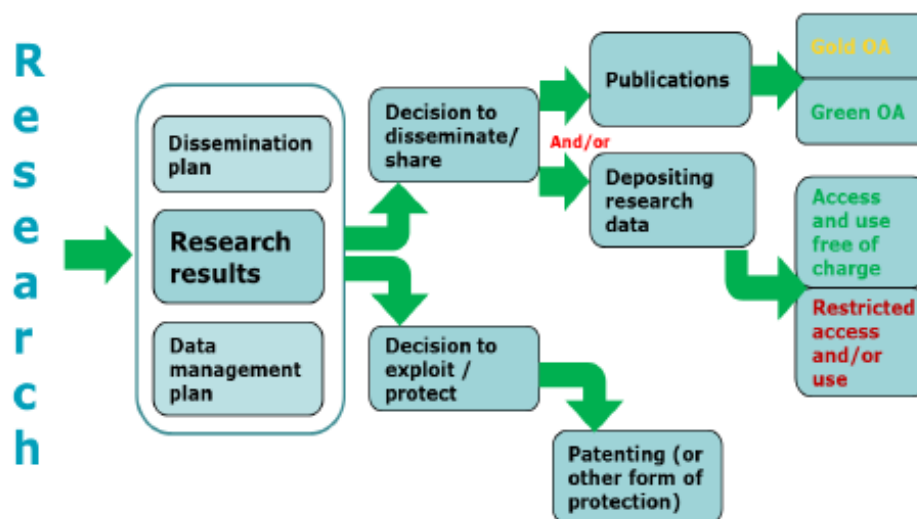


Figure 2: Open Access strategy for publications and research data

The open access mandate comprises two steps:

1. depositing publications in repositories
2. providing open access to them

These steps are explained in the following sections along with the definition of the ways for archiving and preserving the open datasets of the AEGIS project. A methodological process to handle the Data Management Process is provided as part of the framework towards the definition of the Data Management Plan.

## 2.1. DMP Management Process

We highlighted in previous section the main goals and objectives towards the definition of the AEGIS Data Management Plan along with some principles and guidelines.

Taking into account this preliminary analysis, the AEGIS Data Management Process is defined as a step-wise approach for each result generated or collected during the project runtime. The following questions must be answered to classify the different datasets:

1. Does a result provide significant value to others or is it necessary to understand a scientific conclusion?

If this question is answered with yes, then the result is classified as public (granted for open access). If this question is answered with no, the result is classified as non-public. For example, code that is very specific to the AEGIS platform (e.g. a database initialization) is usually of no scientific interest to anyone, nor does it add any significant contribution.

2. Does a result include personal information that is not the author's name?

If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published according to the ethical principles of the project.

3. Does a result allow the identification of individuals even without the name?

This is also a step managed by the ethics management plan of the project as we have committed in AEGIS project to establish anonymization techniques to conceal a single user's identity, e.g. abstraction, dummy users, or non-intersecting features. If this question is answered with yes, the result is classified as non-public.

4. Can a result be abused for a purpose that is undesired by society in general or contradict with societal norms and the project's ethics?

If this question is answered with yes, the result is classified as non-public.

5. Does a result include business or trade secrets of one or more partners of the project?

If this question is answered with yes, the result is classified as non-public. Business or trade secrets need to be removed in accordance to all partners' requirements before it can be published.

6. Does a result name technologies that are part of an ongoing, project-related patent application?

If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after patent has been filed.

7. Does a result break security interests for any project partner?

If this question is answered with yes, the result is classified as non-public.

This is a simple structural approach to determine the different data types defined as part of the DMP. The responsibilities of the AEGIS consortium partners towards disseminating the project outcomes are defined in the following section.

## 2.2. Responsibilities and Decision Making

The Data Management Plan presented in this deliverable aims to identify the project outputs to be massively disseminated as well as to decide on way and means of their Open Access (if applicable). To ensure it, a dedicated time slot will be reserved at each of the project plenary meetings and, if needed, at selected consortium audio conferences. EC and project reviewers will be informed about related work done and publications provided in the project management reports.

Individual responsibilities on data management in the project consortium are:

- Data Management Plan Leader (HYPERTECH) – to prepare and lead related discussions at the relevant project meetings and to maintain the channels for dissemination of project outcomes.
- Scientific and Technical Project Manager – to identify data collected by the project and technical project outcomes eventually suitable for publication
- Dissemination leader – to identify publications suitable for publication in the considered repositories and maintain AEGIS inputs for the Open Access
- Each individual partner – to identify own project results suitable for publication

Moreover, each AEGIS partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation. Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the WP. Metadata constitutes an underlying definition or description of the datasets, and facilitate finding and working with particular instances of data.

Backing up data for sharing through Open Access repositories is the responsibility of the partner possessing the data.

Quality control of these data is the responsibility of the relevant WP leader, supported by the Data Management Plan Leader.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publically available data.

Last but not least, all partners must consult the concerned partner(s) before publishing data in the open domain that can be associated to an exploitable result.

By taking into account the methodological framework, we proceed with the early definition of data sets to be considered as part of the DMP in the following section.

### 3. DATA ARCHIVING AND PRESERVING INFRASTRUCTURE

Before providing the detailed analysis of the datasets/publications to be handled within the context of AEGIS DMP, we provide an overview of the platforms to publish our results openly. The following list presents the platforms selected to present the datasets/publications during the project and describes their concepts for publishing, storage, and backup.

#### 3.1. BSCW

AEGIS Document sharing and collaborative work will be done with the BSCW software. BSCW stands for Basic Support for Cooperative Work is a collaborative workspace software package for collaboration over the Web, developed by the Fraunhofer Society. BSCW is a cloud-based support system, offering a wide range of support for cooperative work. BSCW is developed in Python and features a customized database solution for hosting the user generated content. BSCW is a self-organized system, where the users create and manage their groups without requiring a global platform administrator. BSCW supports document upload, event notification, and group management. Clients require a standard web browser only.

The AEGIS BSCW workspace is hosted by Fraunhofer FOKUS. BSCW is licensed software at the secure Fraunhofer BSCW server where the AEGIS consortium stores all important documents (minutes, deliverables, legal documents, templates etc.). It provides folders for each WP. All directories are accessible for the project participants. The BSCW is accessible through a unique account (username and password).

If a consortium participant needs access to the BSCW the partner beneficiary responsible has to send a request to the Coordinator, who hosts the BSCW. On the basis of the initial structure, the WPLs are responsible for establishment of a well-arranged, self-explaining folder structure on their BSCW folder, so participating partners can easily find the place to upload and install the documents. The AEGIS BSCW will be available during the project runtime, and will still be available for at least one year after the official project end.

Web link: <https://bscw.fokus.fraunhofer.de/bscw/bscw.cgi/2548631>

#### 3.2. AEGIS Project Portal

The partners in the AEGIS consortium decided early to setup its own project-related webpage. This webpage describes the mission and the general approach of the project and its development status. A blog informs about news on a regular basis. A dedicated section for downloads is used to publish reports and white papers. All documents are published using the portable document format (PDF). All downloads are enriched by using simple metadata information like the title and the type of the document. The webpage was designed and developed by the partner of the consortium SUITE5. All webpage-related data is backed on a regular basis. All information on the AEGIS website can be accessed without creating an account, though a private section is also available linked with the BSCW. The webpage is backed manually once per month.

The AEGIS Project Portal will be available during the project runtime, and will still be available for at least two year after the official project end.

Web link: <http://aegis-bigdata.eu/>

### 3.3. ResearchGate

Along with the establishment of the project portal, the AEGIS ResearchGate channel has been established to promote the dissemination of scientific publications of the project. Open Access documents are published using the portable document format (PDF). All downloads are enriched by using simple metadata information like the title, a short description and the type of the document.

The AEGIS ResearchGate channel is managed by SUITE5, a partner of the consortium who periodically updates the material. The link for accessing the AEGIS ResearchGate channel is:

<https://www.researchgate.net/project/AEGIS-Advanced-Big-Data-Value-Chain-for-Public-Safety-and-Personal-Security-2>

The aforementioned tools are defined as the platforms for accessing the project scientific results. Towards the wider dissemination of project outcomes and following the recommendation from E.C., we intend to use Zenodo service for datasets dissemination.

### 3.4. Zenodo Project Repository

Zenodo is an open data repository service maintained by CERN, Geneva and was launched in 2013. Zenodo archives, and makes available, research outputs in all scientific disciplines. Datasets can be located via the Zenodo ElasticSearch engine. Zenodo is compliant with the open data requirements of Horizon 2020, the EU Research and Innovation funding programme and OpenAIRE, the EC-funded initiative in support of the Open Access policies of the European Union.

Zenodo not only supports the publication of scientific papers or white papers, but also the publication of any structured research data (e.g. using XML). Zenodo provides a connector to GitHub that supports open collaboration for source code and versioning for all kinds of data. All uploaded results are structured by using metadata, like for example the contributors' names, keywords, date, location, kind of document, license, and others. Considering the language of textual metadata items, English is preferred. All metadata is licensed under CC license (Creative Commons 'No Rights Reserved'). The property rights or ownership of a result does not change by uploading it to Zenodo.

All public results generated or collected during the AEGIS project will be uploaded to Zenodo for long-term storage and open access. At the end of the project we will consider the option of using an institutional research data repository to further disseminate the project outcomes. The AEGIS project profile has been established in the platform.

Web link: <https://zenodo.org/communities/aegisproject/?page=1&size=20>

### 3.5. GitHub

As part of the exploitation process, the members of the consortium will disseminate some of the project outcomes as Open Source products. While Zenodo is the repository for archiving and preserving publications and datasets, it provides also a connector to GitHub that supports open collaboration for source code. Therefore, the aim of the consortium is to preserve a GitHub channel for disseminating the open source outcomes of the AEGIS project.



GitHub is a well-established online repository which supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables world-wide collaboration between developers and provides also some facilities to work on documentation and to track issues.

GitHub provides paid and free service plans. Free service plans can have any number of public, Open Access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan. Many open-source projects use GitHub to share their results for free. The platform uses metadata like contributors' nicknames, keywords, time, and data file types to structure the projects and their results. The terms of service state that no intellectual property rights are claimed by the GitHub Inc. over provided material. For textual metadata items, English is preferred. The service is hosted by GitHub Inc. in the United States. GitHub using a rented Rackspace hardware infrastructure where data is backed continuously to different locations.

All source-code components that are implemented during the project and decided to be public will be uploaded to an open access GitHub repository, linked also with the Zenodo account of the project.

Web link: <https://github.com/>

We have presented above the list of platforms to be considered for disseminating the project outcomes. We proceed in the next section with the early definition of the datasets/ documents that will be disseminated as part of the Data Management Handling Plan.

#### 4. DATASETS AND PUBLICATIONS FOR DMP

In this section, a list of all existing or foreseeable results for dissemination is presented, separated into public deliverables, publications and open research data. For each result and in accordance to the FAIR data management guideline (European Commission, 2013) we provide a description, name the standards used for storage and metadata (to make data findable & interoperable), and define which open access platform is chosen. Data Security aspects are also defined in this document, while the detailed ethics management policy of AEGIS project is defined in WP9 which provides detailed analysis on the way to handle the datasets generated in the project. In summary, the AEGIS partners will comply with the ethical principles as set out in Article 34 of the Grant Agreement, which asserts that all project activities must be carried out in compliance with:

- (a) Ethical principles (including the highest standards of research integrity - as set out, for instance, in the European Code of Conduct for Research Integrity - and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct)
- (b) Applicable international, EU and national law.

##### 4.1. AEGIS Project Public deliverables

We are considering the AEGIS Project public deliverables as part of the data management plan. The following table presents the list of public deliverables of the AEGIS project.

<b>D_ID</b>	<b>Title</b>	<b>Due Date</b>
D1.1	Domain Landscape Review and Data Value Chain Definition	31-Mar-17
D1.2	The AEGIS Methodology and High Level Usage Scenarios	31-May-17
D1.3	Final AEGIS Methodology	31-Mar-18
D2.1	Semantic Representations and Data Policy and Business Mediator Conventions	31-Aug-17
D2.2	AEGIS Data Value Chain Bus Definition and Data Analysis Methods	31-Aug-17
D2.3	Update on Semantic Representation and Data handling and Analytics Methods	30-Jun-18
D3.1	Technical and User Requirements and Architecture v1.00	30-Jun-17
D3.2	AEGIS Components, Microservices and APIs Design v1.00	30-Nov-17
D3.3	AEGIS Components, Microservices and APIs Design v2.00	31-Mar-18
D3.4	AEGIS Components, Microservices and APIs Design v3.00	30-Sep-18
D3.5	AEGIS Components, Microservices and APIs Design v4.00	31-Jan-19
D4.2	AEGIS Platform - Release 2.00	30-Jun-18
D4.3	AEGIS Platform - Release 3.00	31-Dec-18
D4.4	AEGIS Platform - Release 4.00	30-Jun-19
D5.1	Demonstrators and Project Evaluation Framework	31-Dec-17
D5.3	Demonstrators Evaluation and Feedback–v1	28-Feb-18
D5.4	Demonstrators Evaluation and Feedback–v2	31-Dec-18
D5.5	Demonstrators Evaluation and Feedback–v3	30-Jun-19
D5.6	Final Evaluation, Impact Assessment and Adoption Guidelines	30-Jun-19
D6.1	Plan for Dissemination, Communication and Stakeholder Engagement	31-Mar-17
D6.2	Project Website and Web 2.0 Channels	31-Mar-17

D6.3	Data Management Handling Plan	30-Jun-17
D6.4	Dissemination, Communication and Stakeholder Engagement Report and Plan – Interim Version	30-Jun-18
D6.5	Dissemination, Communication and Stakeholder Engagement Report and Plan – Final Version	30-Jun-19

Table 2: List of AEGIS Public Deliverables

The template for the management of public deliverables is provided:

**Data set reference and name:** The name and the ID of the deliverable

**Data set description:** A short description of the content of the deliverable

**Standards and metadata:** The type of the document format and any type of metadata associated with the content of the document.

**Data sharing, archiving and preservation (including storage and backup):** how data will be preserved, how to archive and preserve the open datasets of the project.

An overview of the public documents prepared so far by the consortium are presented:

#### *4.1.1. D1.1 Domain Landscape Review and Data Value Chain Definition*

##### **Data set reference and name**

Domain Landscape Review and Data Value Chain Definition (Report D1.1)

##### **Data set description**

The document provides an overview and analysis of tools and systems for Big Linked Data and the data sources and data value chains that will be defined in the project with special focus about Public Safety and Personal Security applications. The tools and systems identified in this document facilitate us to build a platform for an open ecosystem in which PSPS actors can safely and securely share data. As such, this document includes a state-of-art analysis of the Big Data landscape in terms of frameworks and tools that straddle the boundary with Linked Data. In the second part of this document, we identify the stakeholders who could benefit from AEGIS, their needs and requirements, and preliminary features along with the definition of the initial data value chain of the project.

##### **Standards and metadata**

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organizations, and keywords that classify this report (Big, Linked and Open Data, Semantic Web, RDF, SPARQL, Entity Recognition Tools, Stakeholder analysis, Data Sources and Value Chain)

##### **Data sharing, archiving and preservation**

The document will be published openly on the AEGIS webpage (following E.C. review and approval). All earlier versions of the document are archived on the project-internal BSCW

repository. The repository is backed on a regular basis by Fraunhofer. The document will be added to Zenodo for long term preservation at the end of the project.

#### *4.1.2. D6.1 Plan for Dissemination, Communication and Stakeholder Engagement*

##### **Data set reference and name**

Plan for Dissemination, Communication and Stakeholder Engagement (Report D6.1)

##### **Data set description**

This document, entitled “Plan for Dissemination, Communication and Stakeholder Engagement”, provides the detailed and well-targeted dissemination plan that will uptake all possible and appropriate dissemination channels and means also exploiting the collaboration and open deliberation capabilities provided by contemporary social networking. The scope of this document is to report the work performed in Task 6.1 on Planning Dissemination, Communication and Community Engagement. The main goals of this task are:

- To produce the strategy and the plans for dissemination, communication and stakeholder’s engagement throughout the project’s period
- To provide the 1st list of potential dissemination, communication and users’ engagement activities to be performed by the consortium during the 1st project period (M1- M18)

##### **Standards and metadata**

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organizations, and keywords that classify this report (Dissemination & Communication, Stakeholders Engagement, DCP, Dissemination & Communication Means and Opportunities)

##### **Data sharing, archiving and preservation**

The document will be published openly on the AEGIS webpage (following E.C. review and approval). All earlier versions of the document are archived on the project-internal BSCW repository. The repository is backed on a regular basis by Fraunhofer. The document will be added to Zenodo for long term preservation at the end of the project.

#### *4.1.3. D6.2 Project Website and Web 2.0 Channels*

##### **Data set reference and name**

Project Website and Web 2.0 Channels (Report D6.2)

##### **Data set description**

The purpose of this report is to briefly document the AEGIS website design and deployment, as well as the creation of AEGIS accounts in popular social media. The design and implementation of the AEGIS website falls under the activities of WP6 “Dissemination and Communication” and in particular of Task T6.2 “Project Communication Activities”.

According to the DoA, the objectives of WP6 are (among others) to establish an effective online presence and to communicate the project outcomes to the intended audiences, in a way that is consistent with the project's branding and scope.

### **Standards and metadata**

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organizations, and keywords that classify this report (AEGIS Website, Social Media Channels)

### **Data sharing, archiving and preservation**

The document will be published openly on the AEGIS webpage (following E.C. review and approval). All earlier versions of the document are archived on the project-internal BSCW repository. The repository is backed on a regular basis by Fraunhofer. The document will be added to Zenodo for long term preservation at the end of the project.

#### *4.1.4. D1.2 The AEGIS Methodology and High Level Usage Scenarios*

### **Data set reference and name**

The AEGIS Methodology and High Level Usage Scenarios (Report D1.2)

### **Data set description**

This deliverable is related to the activities to be performed under Task 1.4 T1.4-Regulatory Framework for Data Protection, IPR and Ethical Issues and 1.5- T1.5-Methodology Elaboration and High Level Usage Scenarios and reports the AEGIS integrated methodology, showing how components and concepts interrelate while also displaying high level usage scenarios of the concept. In addition, this report includes a legal evaluation and assessment of the AEGIS technologies as well as the integrated systems, assessing to what extent the legal requirements have been taken into account.

### **Standards and metadata**

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organizations, and keywords that classify this report (Ethics, Legal aspects, software components, usage scenarios, AEGIS Most Valuable Product (MVP))

### **Data sharing, archiving and preservation**

The document is published openly on the AEGIS webpage. All earlier versions of the document are archived on the project-internal BSCW repository. The repository is backed on a regular basis by Fraunhofer. The document will be added to Zenodo for long term preservation at the end of the project.

## 4.2. Research Datasets and Software Components

The analysis covers the two types of data streams defined early in this section: the Open Source AEGIS components and the datasets that set the inputs/ outputs of the project. The analysis starts with the analysis of the open source components defined early in the project.

### 4.2.1. Open Source Software Components

The analysis is performed by taking into account the list of exploitable components, defined in D7.1 as open source components. A summary is presented in the following table.

S_ID	Title	Responsible Partner
S01	Semantic Models	NTUA, Fraunhofer
S02	Micro-Services and Orchestrator Engine	UBITECH, SUITE5
S03	Cross-Sector Analytics and Visualization	GFT, UBITECH, SUITE5
S04	Data handling algorithms	Fraunhofer, UBITECH
S05	Business Brokerage Engine	GFT, UBITECH, SUITE5
S06	AEGIS As a whole	All

Table 3: List of Open Source software components

While AEGIS project is committed to submit most of the platform software as Open Source software components, the final decision will be made along with the definition of the IPR issues in WP7. Thus, we are providing a typical template about Open Source Software dissemination within the context of DMP.

**Data set reference and name:** The name of the software component

**Data set description:** A short description of the content of the software application

**Standards and metadata:** The dev. tools utilized for the development and the deployment of the application

**Data sharing, archiving and preservation (including storage and backup):** how the source code will be preserved, how to archive and preserve the open source code of the project.

The example for an open source software development of the project is presented.

#### 4.2.1.1. Data handling algorithms

##### Data set reference and name

Data handling algorithms

## Data set description

With the use of the AEGIS big data value chain technologies, the AEGIS consortium is contributing to challenges of the Big Data community: which are to solve a number of hitherto difficult to solve safety and security problems the business world was facing so far. The challenges and difficulties lie in the justification of algorithm correctness, on analysis of suitable algorithm performance, and on choosing the right data structure for the problem at hand.

To face that demand AEGIS is developing an elastic / adaptive infrastructure with libraries of data handling algorithms to analyse transactional, structured and unstructured data on a single platform. The AEGIS data handling algorithms include a library of sector specific algorithms powered by Spark on top of the HDFS and Hadoop. More specifically this library includes the following features:

- Uniform data access that allows a common way to access a variety of data sources, including Hive, Avro, Parquet, ORC, JSON, and JDBC
- Run unmodified Hive queries on existing data
- Access Hive meta-store, that allows full compatibility with existing Hive data, queries, and UDFs
- Cost-based optimizer, columnar storage and code generation to enable fast queries
- Scale to thousands of nodes and multi hour queries using the Spark engine to provide full mid-query fault tolerance.

The algorithmic framework towards handling the different datasets in a structural way is considered as the open source outcome of the project.

## Standards and metadata

Different programming languages may be used to implement the data handling processes. The responsible technical partners (Fraunhofer & UBITECH) contributing at the development of the open source software component will decide which parts are trust and may be open sourced. The source code should be reviewed before it is published while information about platform security (e.g. server names), user privacy, or business secrets will be deleted. The documentation of the source code along with the associated metadata is provided in the code itself.

The details of the open source data sets will be decided during the implementation and updated in the following versions of the data management plan.

## Data sharing

Part of the source code will be published at the end of the project as open source using GitHub. The access is free for everyone and without restrictions.

## Archiving and preservation (including storage and backup)

Each partner in the project uses their own state of the art version management repositories to store their source code contributions. Each partner is backing its repository on a regular basis. All open sourced components will be added to GitHub for long term preservation during or at the end of the project.

#### 4.2.2. AEGIS Project Datasets

We are presenting different data types to be considered as part of the Data Management Plan following the initial identification of AEGIS datasets in D1.1 & D1.2. The review is not including the list of open datasets that are already publicly available but private datasets gathered from the different demonstrators (Smart Home and AAL, Smart Automotive, Smart Insurance). Also, we need to point out that the AEGIS platform will act as a data marketplace, accessible by external stakeholders towards retrieving different data types (fully preserving ethical and IPR issues defined in the project). Therefore, the platform itself may be considered as a means of disseminating datasets. We are providing a typical template about datasets presentation within the context of DMP.

**Data set reference and name:** The name of the dataset

**Data set description:** A short description of the content of the dataset

**Standards and metadata:** The format utilized for making the datasets publicly available

**Data sharing:** How the dataset will be preserved during the project period

**Data archiving and preservation (including storage and backup):** How to archive and preserve the datasets of the project

##### 4.2.2.1. Smart Automotive Demonstrator

We have defined in D1.1 the list of datasets for Smart Automotive Demonstrator .We selected the preliminary list of data attributes that may be useful for external stakeholders and thus are part of the AEGIS Data Management Handling Plan.

#### Anonymous statistics about on board diagnostics data

##### **Data set reference and name**

Anonymous statistics about on board diagnostics

##### **Data set description**

The goal of AEGIS smart automotive demonstrator is to correlate information coming from on board diagnostics with contextual information (GPS, location, trips, position) towards providing useful services for the end users (drivers, automotive industry etc....).

Vehicle sensor data like calculated engine load, engine coolant temperature, intake manifold, engine speed, vehicle speed, etc..., available from the on board diagnostics device, could be a



very useful dataset to 3<sup>rd</sup> parties for further experimentation. Toward this direction, we are considering further dissemination of this dataset as part of AEGIS Data Management Plan.

### **Standards and metadata**

The data will be stored in a platform-independent format (.CSV). Any names or features that may be used for single user identification are removed at the data source layer and thus and thus anonymous time-series data will be available. Part of the metadata extracted from the raw streams of data will be made also available, though we need first to define the list of metadata generated in the project.

### **Data sharing**

User-related data that may lead to single user identification will not be published. Only abstract time-series and statistics that do not allow for single user identification will be made openly accessible. Any anonymous user data (e.g. statistics) will be published openly using Zenodo. The access is free for everyone and without restrictions. A sample dataset is already accessible via: <https://zenodo.org/record/820576#.WVUGCoR97cs>

### **Archiving and preservation (including storage and backup)**

The data types defined above are stored in project-related databases, which are hosted by the project partners. Considering privacy and security concerns, partner of the consortium restrict access to private information by following the security policy as defined in the project. Databases are backed on a regular basis by each partner. The user statistics as defined in the DMP will be added to Zenodo for long term preservation during or at the end of the project.

#### **4.2.2.2. Smart Home Environment & Ambient Assisted Living Demonstrator**

We have defined in D1.1 the list of datasets for Smart Home Environment & Ambient Assisted Living Demonstrator. We selected here the preliminary list of data attributes to be available for further dissemination.

##### Anonymous statistics about temperature & humidity

#### **Data set reference and name**

Anonymous statistics about temperature & humidity data

#### **Data set description**

The goal of AEGIS smart home demonstrator is to provide non-intrusive profiling techniques that will be employed to enable the definition of accurate ambience preference profiles of elderly people in the indoor environment. Such profiles will set the basis for the deployment of human-centric, personalised smart automation strategies over their heating/ cooling and lighting devices, to ensure optimal comfort levels and compliance with special, ambience-related, health requirements.

Towards this direction sensing data (temperature, humidity among others) will be processed in combination with outdoor environment data to ensure (through automated control) that indoor ambient conditions always fit the personal preferences of the elderly occupants.

### **Standards and metadata**

The data will be stored in a platform-independent format (e.g. XML or JSON). Any names or features that may be used for single user identification are removed at the data source layer and thus and thus anonymous time-series data will be available. Part of the metadata extracted from the raw streams of data will be made also available, though we need first to define the list of metadata generated in the project.

### **Data sharing**

User-related data that may lead to single user identification will not be published. Only abstract time-series and statistics that do not allow for single user identification will be made openly accessible. Any anonymous user data (e.g. statistics) will be published openly using Zenodo. The access is free for everyone and without restrictions.

### **Archiving and preservation (including storage and backup)**

The data types defined above are stored in project-related databases, which are hosted by the project partners. Considering privacy and security concerns, partner of the consortium restrict access to private information by following the security policy as defined in the project. Databases are backed on a regular basis by each partner. The user statistics as defined in the DMP will be added to Zenodo for long term preservation during or at the end of the project.

#### Anonymous statistics about energy consumption

### **Data set reference and name**

Anonymous statistics about energy consumption data

### **Data set description**

One of the main objectives of AEGIS project and Smart Home Environment demonstrator is to properly analyse energy consumption data and correlate them with energy pricing data in order to avoid extraordinary energy costs and optimize balancing (“sweet spot”) with personal comfort/ health.

Towards this direction energy metering data per device type will be available at the project demonstrator, enabling that way the implementation of the aforementioned use case.

### **Standards and metadata**

The data will be stored in a platform-independent format (e.g. XML or JSON). Any names or features that may be used for single user identification are removed at the data source layer and thus and thus anonymous time-series data will be available. Part of the metadata extracted from the raw streams of data will be made also available, though we need first to define the list of metadata generated in the project.

## Data sharing

User-related data that may lead to single user identification will not be published. Only abstract time-series and statistics that do not allow for single user identification will be made openly accessible. Any anonymous user data (e.g. statistics) will be published openly using Zenodo. The access is free for everyone and without restrictions.

## Archiving and preservation (including storage and backup)

The data types defined above are stored in project-related databases, which are hosted by the project partners. Considering privacy and security concerns, partner of the consortium restrict access to private information by following the security policy as defined in the project. Databases are backed on a regular basis by each partner. The user statistics as defined in the DMP will be added to Zenodo for long term preservation during or at the end of the project.

We presented above the preliminary list of AEGIS datasets that may be published openly using either existing institutional repositories or Zenodo. The kind of data to be generated in the project has still to be decided during the design and implementation phase and thus the updated version of the deliverable will include the final list of datasets for dissemination.

## 4.3. AEGIS Scientific Publications

Along with the dissemination of project deliverables and datasets, we are considering as part of the Data Management Handling Plan, further dissemination of project Scientific Publications.

There are publications already submitted by the consortium partners in scientific papers and conference events, some of them is within the context of open access and thus an indicative reporting of the work performed in the project.

### 4.3.1. *Quantified Vehicles: Novel Services for Vehicle Lifecycle Data*

#### Data set reference and name

AEGIS Project Research Paper 1: “Quantified Vehicles: Novel Services for Vehicle Lifecycle Data”.

#### Data set description

This is a full research paper titled “Quantified Vehicles: Novel Services for Vehicle Lifecycle Data” published by the partner of the consortium VIF in the highly ranked journal Business & Information Systems Engineering (BISE)

The article, titled Quantified Vehicles: Novel Services for Vehicle Lifecycle Data:

- defines the term Quantified Vehicles,
- explains why car manufacturers are not in the pole position for inventing novel digital services for vehicles,
- discusses stakeholder types for digital vehicle services within an ecosystem,

- provides an overview about available services from start-ups and discuss why they usually are located in the U.S. and, amongst others,
- mentions aspects like privacy, security or standardization.

The overall framework as defined in the paper is in line with the definition of the Smart Automotive demonstrator examined in the AEGIS project.

### **Standards and metadata**

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organization, and keywords that classify this research paper. For indexing linked references, citable DOI numbers are added by the publisher.

### **Data sharing**

This research paper is published in the highly ranked journal Business & Information Systems Engineering (BISE). While the document is not an open access publication, this will be freely available after a 12 month embargo period (April 2018). The paper will be available in ResearchGate channel of the project.

#### *4.3.2. Quantified Cars: Novel Digital Services and Business Models Based On Vehicle Data*

### **Data set reference and name**

AEGIS Project Research Paper 2: “Quantified Cars: Novel Digital Services and Business Models Based On Vehicle Data”.

### **Data set description**

This is a full research paper titled “Quantified Cars: Novel Digital Services and Business Models Based on Vehicle Data” presented by the partner of the consortium EPFL in the 2<sup>nd</sup> International Conference on New Business Models (held in Graz on 20-23 June 2017).

It describes the Quantified Car topic which currently pops up on the European market and has reached attention of car manufacturers. New business models are underway and aim to revolutionize the automotive industry. The overall framework as defined in the paper is in line with the definition of the Smart Automotive demonstrator examined in the AEGIS project.

### **Standards and metadata**

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organizations, and keywords that classify this research paper. For indexing linked references, citable DOI numbers are added by the publisher.

### **Data sharing**

This research paper will be published within the proceedings of the International Conference on New Business Models (held in Graz on 20-23 June 2017). The conference proceedings are published as the Institute of Systems Sciences, Innovation, and Sustainability Research Report (former “ISIS-Report”, ISSN 2305-2511, print; ISSN 2308-1767, online) under a Creative

Commons copyrights license. Therefore, the paper is easily accessible for the mass audience. The paper is also available in ResearchGate channel of the project and the project portal.

We presented above the project outcomes (publications issued by the project consortium, white papers published, open source code generated, mock-up datasets used for supporting the development process) towards communicating and spread knowledge of project results to all interested communities and stakeholders. The analysis is tightly connected with the definition of open repositories for data management and dissemination as defined in previous section.

## 5. SUMMARY AND CONCLUSIONS

The 1<sup>st</sup> version of Deliverable 6.3 Data Management Handling Plan is provided in M6 and describes the steps for data management which are to be followed during the execution of AEGIS project. The current document gives preliminary information on the data types used and generated by the project consortium partners including focus on the means of sharing data captured by AEGIS framework and further specifies the methods of data storage thus providing general view over the complete data management life cycle. As this report is generated in the early stage of the project execution is considered as a living document which will be further supplemented.

**APPENDIX A: LITERATURE**

- [1]. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, [http://www.gsrt.gr/EOX/files/h2020-hi-oa-data-mgt\\_en.pdf](http://www.gsrt.gr/EOX/files/h2020-hi-oa-data-mgt_en.pdf)
- [2]. Guidelines on FAIR Data Management in Horizon 2020, [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- [3]. How to create a DMP Plan, <https://www.openaire.eu/opendatapilot-dmp>
- [4]. How to select a repository?, <https://www.openaire.eu/opendatapilot-repository>