D5.6 – Final Evaluation, Impact Assessment and Adoption Guidelines

Version 1.0

Due date: 30.06.2019  Delivery Date: 29.06.2019

Author(s): Dimitrios Miltiadou, Konstantinos Perakis, Stamatís Pitsios (UBITECH), Marios Phinikettos (SUITE5), Cesar Lisi, Alessandro Testa (HDI), Alexander Stocker, Christian Kaiser (VIF), Elisa Rossi (GFT), Anna Maria Hounti. Evgenia Alexopoulou, Fotios Manesis (KONKAT), Alexandru Ormenisan, Mahmoud Ismail (KTH), Fabian Kirstein (Fraunhofer)

Editor: Dimitrios Miltiadou (UBITECH)

Lead Beneficiary of Deliverable: UBITECH


Internal Reviewers: Gianluigi Viscusi (EPFL), Alessandro Testa (HDI)
EXPLANATIONS FOR FRONTPAGE

Author(s): Name(s) of the person(s) having generated the Foreground respectively having written the content of the report/document. In case the report is a summary of Foreground generated by other individuals, the latter have to be indicated by name and partner whose employees he/she is. List them alphabetically.

Editor: Only one. As formal editorial name only one main author as responsible quality manager in case of written reports: Name the person and the name of the partner whose employee the Editor is. For the avoidance of doubt, editing only does not qualify for generating Foreground; however, an individual may be an Author – if he has generated the Foreground - as well as an Editor – if he also edits the report on its own Foreground.

Lead Beneficiary of Deliverable: Only one. Identifies name of the partner that is responsible for the Deliverable according to the AEGIS DOW. The lead beneficiary partner should be listed on the frontpage as Authors and Partner. If not, that would require an explanation.

Internal Reviewers: These should be a minimum of two persons. They should not belong to the authors. They should be any employees of the remaining partners of the consortium, not directly involved in that deliverable, but should be competent in reviewing the content of the deliverable. Typically this review includes: Identifying typos, Identifying syntax & other grammatical errors, Altering content, Adding or deleting content.
AEGIS Key Facts

Topic: ICT-14-2016 - Big Data PPP: cross-sectorial and cross-lingual data integration and experimentation

Type of Action: Innovation Action

Project start: 1 January 2017

Duration: 30 months from 01.01.2017 to 30.06.2019 (Article 3 GA)

Project Coordinator: Fraunhofer

Consortium: 10 organizations from 8 EU member states

AEGIS Partners

Fraunhofer Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.

GFT GFT Italia SRL

KTH Kungliga Tekniska högskolan

UBITECH UBITECH Limited

VIF Kompetenzzentrum - Das virtuelle Fahrzeug , Forschungsgesellschaft-GmbH

NTUA National Technical University of Athens – NTUA

EPFL École polytechnique fédérale de Lausanne

SUITE5 SUITE5 Limited

KONKAT ANONYMOS ETAIREIA KATASKEVON-TECHNIKON ERGON, EMPORIKON, VIOMICHANI KonKAI NAUTILIKAON EPICHEIRISEON KON'KAT

HDIA 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

The scope of deliverable D5.6, which is the final deliverable of WP5, is to document the efforts undertaken within the context of the tasks 5.3, 5.4, 5.5, and 5.6. Towards this end, the deliverable builds on top of the works and outcomes of all deliverables of WP5, namely D5.1, D5.2, D5.3, D5.4 and D5.5, in order to report the final evaluation of both the AEGIS demonstrators and the AEGIS platform taking into consideration of all the activities that were performed from M7 till M30 of the project. The final evaluation presents the summary of the key outcomes of the holistic quantitative and qualitative evaluation that performed during the project, as dictated by the AEGIS evaluation framework.

Within the context of the deliverable D5.6, an overview of all the demonstrators activities is documented, presenting all the detailed information with regards to the execution scenarios, the demonstrators’ implementation phases, the demonstrators’ execution, and the demonstrators’ evaluation. For each of the three demonstrators, a detailed overview is presented, highlighting all the demonstrators’ activities during the project, the digital artefacts that were implemented within the context of each demonstrator, as well as the achievements and the lessons learnt from each demonstrator.

Following the final evaluation of the AEGIS demonstrators, the deliverable also presents the final evaluation of the AEGIS platform. The deliverable presents the aggregated results of all three quantitative and qualitative evaluations that were performed during the project period, as well as the description of the insights that were extracted from this holistic evaluation. In addition to quantitative and qualitative results, the deliverable discusses the security assessment of the final version of the platform that was performed in order to evaluate the security mechanisms of the platform.

Finally, the deliverable presents the AEGIS platform documentation and adoption guidelines. In detail, a comprehensive documentation of the AEGIS platform usage is documented, providing useful insights in all the functionalities of the platform and the necessary guidelines for the development activities that can be performed in the platform. Following the practical guidelines, the deliverable presents the AEGIS platform’s adoption guidelines that are based on the knowledge obtained from the partners involved during the implementation phase of the demonstrators. The guidelines are presented in order to be exploited from the potential users of the platform beyond the project consortium.
Table of Contents

EXPLANATIONS FOR FRONTPAGE ........................................................................................................... 2
AEGIS KEY FACTS ............................................................................................................................... 3
AEGIS PARTNERS ............................................................................................................................... 3
EXECUTIVE SUMMARY ...................................................................................................................... 4
LIST OF FIGURES ............................................................................................................................... 6
LIST OF TABLES ................................................................................................................................... 9
ABBREVIATIONS .................................................................................................................................. 10

1. INTRODUCTION .................................................................................................................................. 11
   1.1. OBJECTIVE OF THE DELIVERABLE ............................................................................................ 11
   1.2. INSIGHTS FROM OTHER TASKS AND DELIVERABLES ................................................................. 11
   1.3. STRUCTURE .................................................................................................................................... 11

2. AEGIS DEMONSTRATORS FINAL EVALUATION ............................................................................. 13
   2.1. DEMONSTRATORS OVERVIEW .................................................................................................... 13
   2.2. AUTOMOTIVE DEMONSTRATOR OVERVIEW, ACHIEVEMENTS AND LESSONS LEARNT .......... 19
   2.3. SMART HOME AND ASSISTED LIVING DEMONSTRATOR OVERVIEW, ACHIEVEMENTS AND LESSONS LEARNT .............................................................................................................................................. 27
   2.4. INSURANCE DEMONSTRATOR ACHIEVEMENTS AND LESSONS LEARNT ................................ 32

3. AEGIS PLATFORM FINAL EVALUATION ......................................................................................... 36
   3.1. QUANTITATIVE EVALUATION ....................................................................................................... 36
   3.2. QUALITATIVE EVALUATION ......................................................................................................... 46
   3.3. PLATFORM SECURITY ASSESSMENT ............................................................................................. 48

4. AEGIS PLATFORM DOCUMENTATION AND ADOPTION GUIDELINES ............................................ 50
   4.1. AEGIS PLATFORM USAGE DOCUMENTATION .............................................................................. 50
       4.1.1. AEGIS Usage ........................................................................................................................... 50
       4.1.2. AEGIS Jupyter Tools .............................................................................................................. 51
       4.1.3. Extended Metadata ................................................................................................................ 71
       4.1.4. Jobs ............................................................................................................................................ 74
       4.1.5. Kafka ....................................................................................................................................... 76
       4.1.6. Model Serving ......................................................................................................................... 78
       4.1.7. Cleansing Tool ......................................................................................................................... 79
       4.1.8. Anonymisation Tool ............................................................................................................... 85
   4.2. AEGIS PLATFORM ADOPTION GUIDELINES .............................................................................. 89

5. CONCLUSION ...................................................................................................................................... 92
LIST OF FIGURES

Figure 2-1: Automotive demonstrator dataset overview .......................................................... 20
Figure 2-2: Data workflow of the Automotive Demonstrator (V1, V2, V3) ............................ 21
Figure 2-3: Broken Road Indicator (Automotive Demonstrator V1) ...................................... 22
Figure 2-4: Safe Driving Indicator I (Automotive Demonstrator V2) .................................. 22
Figure 2-5: Visualisation of all individual risk scores of a single driver (Automotive Demonstrator V2) ................................................................. 23
Figure 2-6: Visualisation of aggregated risk scores of all drivers(Automotive Demonstrator V2) .................................................................................................................. 23
Figure 2-7: Regional Driving Risk Estimator Indicator (Automotive Demonstrator V3) ....... 24
Figure 2-8: Automotive Demonstrator: Mini Use Case Driver Classification ...................... 25
Figure 2-9: Automotive Demonstrator: Driver App I ................................................................. 26
Figure 2-10: Automotive Demonstrator: Driver App II ........................................................... 26
Figure 2-11: SHAL Web App individuals’ dashboard ................................................................. 28
Figure 2-12: SHAL Web App list of Personas with action buttons .......................................... 28
Figure 2-13: SHAL Web App aggregated stats of a Persona .................................................. 29
Figure 2-14: SHAL iOS Dashboard and Graphs ........................................................................ 29
Figure 2-15: SHAL Android Dashboard and Graphs ................................................................. 30
Figure 2-16: SHAL Display of third-party sensor measurements (smart home) .................... 30
Figure 2-17: SHAL Notification as shown in an iPhone locked screen ................................. 31
Figure 2-18: Insurance demonstrator - scenarios overview ................................................. 33
Figure 2-19: Insurance demonstrator - technical components and user(s) for each component ...................................................................................................................... 34
Figure 4-1: AEGIS User Main Page ........................................................................................... 50
Figure 4-2: Project Main Page .................................................................................................. 51
Figure 4-3: AEGIS Tools ........................................................................................................ 52
Figure 4-4: Query Builder - Initial state .................................................................................. 52
Figure 4-5: Query Builder - Hide code ................................................................................... 53
Figure 4-32: Algorithm Execution Container - Output folder ................................................. 69
Figure 4-33: Algorithm Execution Container - Overview ......................................................... 70
Figure 4-34: Algorithm Execution Container - Results .......................................................... 70
Figure 4-35: Algorithm Execution Container - Select existing model .................................... 71
Figure 4-36: Metadata Form for Projects .................................................................................. 72
Figure 4-37: Metadata Form for Datasets .................................................................................. 73
Figure 4-38: Metadata Form for Files ......................................................................................... 74
Figure 4-39: The Jobs tool .......................................................................................................... 75
Figure 4-40: Adding a new Job ................................................................................................. 75
Figure 4-41: new Job is added ..................................................................................................... 76
Figure 4-42: Kafka ....................................................................................................................... 76
Figure 4-43: Create new Kafka topic ......................................................................................... 77
Figure 4-44: Kafka schemas tab ............................................................................................... 78
Figure 4-45: Create new Avro schema ..................................................................................... 78
Figure 4-46: Model Serving ........................................................................................................ 79
Figure 4-47: Create a new Serving ............................................................................................ 79
Figure 4-48: Cleansing tool - Landing page .............................................................................. 80
Figure 4-49: Cleansing tool - Login page .................................................................................. 80
Figure 4-50: Cleansing tool - Set provider ................................................................................. 81
Figure 4-51: Cleansing tool - Rules creation ............................................................................. 81
Figure 4-52: Cleansing tool - Validation rule ............................................................................ 82
Figure 4-53: Cleansing tool - Cleansing rules ......................................................................... 82
Figure 4-54: Missing values rules ............................................................................................. 83
Figure 4-55: Cleansing tool - Upload dataset .......................................................................... 83
Figure 4-56: Cleansing tool - Log files ..................................................................................... 83
Figure 4-57: Cleansing tool – Log file view .............................................................................. 84
Figure 4-58: Cleansing tool - Rest API .............................................................. 85
Figure 4-59: Anonymisation tool - Configuration screen ................................... 86
Figure 4-60: Anonymisation tool - Data selection ............................................. 86
Figure 4-61: Anonymisation tool - Indicative functions ........................................ 87
Figure 4-62: Anonymisation tool - Custom user-defined function ............................ 87
Figure 4-63: Anonymisation tool - Output ........................................................... 88
Figure 4-64: Anonymisation tool - Preview and Risk analysis ............................... 88
Figure 4-65: Anonymisation tool - Exposed API ................................................. 89
Figure 4-66: Anonymisation tool - API response ............................................... 89

LIST OF TABLES

Table 2-1: AEGIS Demonstrators execution scenarios ......................................... 14
Table 2-2: AEGIS Demonstrators test cases ......................................................... 17
Table 3-1: Summary of the AEGIS platform quantitative evaluation results ............. 45
**ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAL</td>
<td>Active and Assisted Living</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>CO</td>
<td>Confidential, only for members of the Consortium (including the Commission Services)</td>
</tr>
<tr>
<td>CPU</td>
<td>Central Processing Unit</td>
</tr>
<tr>
<td>CSP</td>
<td>Care Service Provider</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma Separated Values</td>
</tr>
<tr>
<td>D</td>
<td>Deliverable</td>
</tr>
<tr>
<td>DoW</td>
<td>Description of Work</td>
</tr>
<tr>
<td>H2020</td>
<td>Horizon 2020 Programme</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HVAC</td>
<td>Heating Ventilation and Air Conditioning</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicator</td>
</tr>
<tr>
<td>PSPS</td>
<td>Public Safety and Personal Security</td>
</tr>
<tr>
<td>R</td>
<td>Report</td>
</tr>
<tr>
<td>RTD</td>
<td>Research and Development</td>
</tr>
<tr>
<td>SHAL</td>
<td>Smart Home and Assisted Living</td>
</tr>
<tr>
<td>UI</td>
<td>User Interface</td>
</tr>
<tr>
<td>VOC</td>
<td>Volatile Organic Compounds</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language</td>
</tr>
<tr>
<td>WP</td>
<td>Work Package</td>
</tr>
<tr>
<td>Y2</td>
<td>Year 2</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

The scope of this section is to introduce the deliverable and familiarise the reader with its contents. To this end, the current section summarises the objective of the deliverable, its relationship with the other work packages and corresponding deliverables and analyses its structure.

1.1. Objective of the deliverable

The scope of deliverable D5.6, which is the final deliverable of WP5, is to document the efforts undertaken within the context of the tasks 5.3, 5.4, 5.5, and 5.6. Within the context of this deliverable, both the AEGIS demonstrators’ final evaluation and the AEGIS platform’s final evaluation are presented, summarizing the most relevant aspects of the holistic quantitative and qualitative evaluation that was conducted in the course of the project as an overall assessment of the AEGIS platform. Additionally, it presents the results of a security assessment conducted with the final version of the AEGIS platform.

Furthermore, the deliverable includes for each of the three demonstrators (Automotive, Smart Home & Assisted Living, and Insurance demonstrators), an overview of all demonstrators’ activities that were performed during the project. In this overview, the digital artefacts developed within the context of the demonstrator implementation are presented, a summary of the most relevant information regarding each demonstrator execution scenarios and implementation phases is documented, as well as the lessons learned during the comprehensive demonstrator implementation phase on the AEGIS platform.

Furthermore, the deliverable presents the AEGIS platform documentation and adoption guidelines as practical guidelines on how to exploit the value of the AEGIS platform in the best possible way. It includes a comprehensive documentation of the AEGIS platform usage, as well as a guideline on how to develop your own data driven applications with the AEGIS platform. This information was compiled considering the vast knowledge of the data scientists within the AEGIS project implementing the three demonstrators on the AEGIS platform, which is to be transferred to potential users beyond the project consortium.

1.2. Insights from other tasks and deliverables

The deliverable builds on top of the work reported in WP5. In particular, the previous outcomes of the work performed in WP5, as reported in D5.1, D5.2, D5.3, D5.4, and D5.5 provided the AEGIS evaluation framework, as well as the methodology on how to implement the framework during the implementation phase of the three demonstrators of the AEGIS project.

The outcomes of the deliverable D5.2 served as guidance on how the evaluation of both the AEGIS platform and the AEGIS demonstrators will be performed. The AEGIS platform evaluation plan, as well as the scenarios defined for each demonstrator and the documented evaluation plan for each demonstrator have driven the assessment performed during all versions of the demonstrators.

The outcomes of the deliverables D5.3, D5.4, and D5.5, documenting the evaluation of the first (early), second (medium), and third (final) versions of the demonstrators, provided the baseline for D5.6.

1.3. Structure

Deliverable D5.6 is organised in five main sections as indicated in the table of contents:
The first section introduces the deliverable. It documents the scope of the deliverable and briefly describes how the document is structured. It also documents the relation of the current deliverable with the other deliverables, and how the outcomes of other deliverables are received as input to the current deliverable.

Following the introductory section, the second section provides a summarized result of the AEGIS demonstrators’ final evaluation, covering both the quantitative and the qualitative evaluation. It provides an overview of each implemented demonstrator along with the activities performed during the project and the artefacts developed.

The third section summarizes the results of the evaluation of the AEGIS platform that was conducted during the project covering both the quantitative and the qualitative evaluation. The section presents the aggregated results of all three evaluations that were performed during the AEGIS project execution phase as an overall assessment of the AEGIS platform as well as the insights that were extracted. Additionally, it presents the results of a security assessment which was conducted for the final version of the AEGIS platform.

Section four is dedicated to present the AEGIS platform documentation and adoption guidelines and presents the general usage of the AEGIS platform, as well as the usage of the different AEGIS Jupyter tools Query Builder, Visualiser and Algorithm Execution Container. Furthermore, it introduces into the use of jobs, model serving (TensorFlow), the Cleaning tool, and the Anonymisation tool. Finally, this section provides the AEGIS platform adoption guidelines as a practical how-to for creating own data science applications on the AEGIS platform.

Section 5 concludes the deliverable. It outlines the main findings of the deliverable, which will guide the future research and technological efforts of the consortium.
2. AEGIS Demonstrators Final Evaluation

2.1. Demonstrators overview

In this section, an overview of the demonstrator activities that were performed during the project is provided. It includes information with regards to the execution scenarios, the demonstrators’ implementation phases, the demonstrators’ execution, and the demonstrators’ evaluation (interviews, focus groups, workshops).

In the following table, the demonstrator-specific execution scenarios and required functionalities that were implemented within each demonstrator are summarized. In total, 12 scenarios were implemented.

<table>
<thead>
<tr>
<th>Scenario ID</th>
<th>Scenario</th>
<th>Functionalities</th>
<th>Demonstrator Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Broken Road Indicator</td>
<td>The broken road indicator scenario provides the baseline for the two further scenarios. Data is collected by a data logger developed at VIF. 1. Create project 2. Upload driving data (in bulk) and do minimal pre-processing Raw data files of individual sensors are merged, and all trips contained in the data are extracted. 3. Transform driving data and save results Extracted trips are resampled to a fixed, regular-spaced time grid of 10Hz. Coordinate system of the sensors is aligned with coordinate system of the vehicle for each of these trips. 4. Identify road damage and save results (An artificial event-signal is computed. This event-signal has large values in presence of a speedbump/pothole and low values in other situations.) 5. Provide visualisation of road damage</td>
<td>Early</td>
</tr>
<tr>
<td>2</td>
<td>Safe Driving Indicator</td>
<td>Additional functionalities required are: 1. Identify safe driving events and save results 2. Assess driving risk and save results Trip-specific and driver-specific risk scores (safe driving scores are computed) Provide visualisation of save driving events</td>
<td>Medium</td>
</tr>
<tr>
<td>3</td>
<td>Regional Driving Safety Risk Estimator</td>
<td>Additional functionalities required are: 1. Assess regional driving safety risk with a geographic risk estimator and visualise it in a heat map</td>
<td>Advanced</td>
</tr>
</tbody>
</table>
### Smart Home and Assisted Living Demonstrator

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Execution Scenarios</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(At-risk) Individuals data fetching, processing and classification</td>
<td>Registration of personal devices (for data collection and aggregation of activity and health data), registration of external data streams (events, weather, etc.), data processing, creation of personas based on data classification, definition of generic rules per persona, definition of outlier detection algorithms, generation of simple alerts to (at-risk) individuals.</td>
<td>Early</td>
</tr>
<tr>
<td>2</td>
<td>Smart home data monitoring and processing</td>
<td>Registration of smart home data streams, Smart Home Data Monitoring, Pre-processing and normalization.</td>
<td>Early</td>
</tr>
<tr>
<td>3</td>
<td>Notifications and alerts for (at-risk) individuals</td>
<td>More information-rich Personas, definition of medical rules for the notification and alert engine, advanced and non-personalized alerts to (at-risk) individuals, personalised tracking of (at-risk) individuals following their consent and simple notifications to carers</td>
<td>Medium</td>
</tr>
<tr>
<td>4</td>
<td>Smart home comfort profiling and notifications</td>
<td>Thermal and Visual Comfort Profiling, notification and alert services for adverse indoor environmental conditions</td>
<td>Medium</td>
</tr>
<tr>
<td>5</td>
<td>Personalised notifications and recommendations for (at-risk) individuals</td>
<td>Definition of personalized medical rules, personalised alerts and recommendations to (at-risk) individuals and notifications to informal carers</td>
<td>Advanced</td>
</tr>
<tr>
<td>6</td>
<td>Smart home automation services</td>
<td>Optimization and automated control of HVAC and lighting devices</td>
<td>Advanced</td>
</tr>
</tbody>
</table>

### Insurance Demonstrator

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Execution Scenarios</th>
<th>Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Financial impact, customer support and services</td>
<td>Definition of the financial impact that an event (weather or socio-political) should have on the company. Identification of the customers that should be involved and customer support service initialisation.</td>
<td>Early</td>
</tr>
<tr>
<td>2</td>
<td>Personalised early warning systems for asset protection</td>
<td>Detection of a foreseen event that could impact on the HDI customers, identification of the customers, notification to their Mobile App and to their agent. Personalised offers based on the guarantee hold and on the event.</td>
<td>Medium</td>
</tr>
<tr>
<td>3</td>
<td>Business plan and marketing strategy</td>
<td>Business analysis request for analysis on the customers’ portfolio in different years in order to define an accurate marketing strategy as well as a successful business plan.</td>
<td>Advanced</td>
</tr>
</tbody>
</table>

Table 2-1: AEGIS Demonstrators execution scenarios
In the following table, all demonstrator specific test cases and their execution states are summarized. All test cases have been successfully executed. In total, 58 test cases were executed. The test cases were presented in detail in the deliverables D5.3, D5.4 and D5.5.

<table>
<thead>
<tr>
<th>Test case ID</th>
<th>Test case Name</th>
<th>Execution Status</th>
<th>Demonstrator version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario1 - 1.1</td>
<td>Create project</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.2</td>
<td>Upload driving data (in bulk)</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.3</td>
<td>Transform driving data and save results</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.4</td>
<td>Identify road damage and save results</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.5</td>
<td>Provide visualisation of road damage</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario2 - 2.1</td>
<td>Identify safe driving events and save results</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario2 - 2.2</td>
<td>Assess driving risk and save results</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario2 - 2.3</td>
<td>Provide visualisation of safe driving events</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario3 - 3.1</td>
<td>Assess regional driving safety risk with a geographic risk estimator and visualise it in a heat map</td>
<td>Completed</td>
<td>Advanced</td>
</tr>
</tbody>
</table>

**Automotive Demonstrator**

**Smart Home and Assisted Living Demonstrator**
<table>
<thead>
<tr>
<th>Scenario3 - 3.2</th>
<th>Design of outlier notification model</th>
<th>Completed</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario3 - 3.3</td>
<td>Execution of Outlier Notification Model</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario3 - 3.4</td>
<td>Mapping Actions to Notifications for personas of (at-risk) individuals</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario3 - 3.5</td>
<td>Mapping Actions to Notifications for carers</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario4 - 4.1</td>
<td>Software development for comfort profiling and notifications</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario4 - 4.2</td>
<td>Receive alerts regarding uncomfortable or health-endangering conditions</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario5 - 5.1</td>
<td>Enhanced (at-risk) individuals profile and provision of access to CSPs for personalised notifications</td>
<td>Completed</td>
<td>Advanced</td>
</tr>
<tr>
<td>Scenario5 - 5.2</td>
<td>Registration of medical rules</td>
<td>Completed</td>
<td>Advanced</td>
</tr>
<tr>
<td>Scenario5 - 5.3</td>
<td>Re-Classification of Individuals</td>
<td>Completed</td>
<td>Advanced</td>
</tr>
<tr>
<td>Scenario5 - 5.4</td>
<td>Dynamic Dashboard following Algorithm Execution</td>
<td>Completed</td>
<td>Advanced</td>
</tr>
<tr>
<td>Scenario5 - 5.5</td>
<td>Persona and Personalised notifications</td>
<td>Completed</td>
<td>Advanced</td>
</tr>
<tr>
<td>Scenario6 - 6.1</td>
<td>Smart home automation recommendation system</td>
<td>Completed</td>
<td>Advanced</td>
</tr>
<tr>
<td>Scenario6 - 6.2</td>
<td>Smart home automation implementation</td>
<td>Completed</td>
<td>Advanced</td>
</tr>
</tbody>
</table>

**Insurance Demonstrator**

<table>
<thead>
<tr>
<th>Scenario1 - 1.1</th>
<th>Event Detection tool configuration and training</th>
<th>Completed</th>
<th>Early</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario1 - 1.2</td>
<td>Event Detection Tool Streaming</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.3</td>
<td>Create Account and Project</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.4</td>
<td>Anonymised in-house dataset upload</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.5</td>
<td>Identification and visualisation of the possibly involved customers</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.6</td>
<td>Priority list creation</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.7</td>
<td>Priority list (report) de-anonymization and assignment, risk exposure evaluation</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario1 - 1.8</td>
<td>Priority list (report) de-anonymization and assignment</td>
<td>Completed</td>
<td>Early</td>
</tr>
<tr>
<td>Scenario2 - 2.1</td>
<td>Event Detection tool training (version 2)</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario2 - 2.2</td>
<td>Event Detection notification configuration</td>
<td>Completed</td>
<td>Medium</td>
</tr>
<tr>
<td>Scenario2 - 2.3</td>
<td>Event Detection Tool Streaming (version 2)</td>
<td>Completed</td>
<td>Medium</td>
</tr>
</tbody>
</table>
Scenario2 - 2.4 | Create Project | Completed | Medium
---|---|---|---
Scenario2 - 2.5 | Uploading datasets | Completed | Medium
Scenario2 - 2.6 | Mobile App data enrichment | Completed | Medium
Scenario2 - 2.7 | Identification and visualisation of the possibly involved customers | Completed | Medium
Scenario2 - 2.8 | Priority list creation | Completed | Medium
Scenario2 - 2.9 | Priority list (report) de-anonymization, sharing and personalised offer | Completed | Medium
Scenario3- 3.1 | Business analysis request | Completed | Advanced
Scenario3- 3.2 | Create project | Completed | Advanced
Scenario3- 3.3 | Uploading anonymised datasets | Completed | Advanced
Scenario3- 3.4 | Business Analysis – open datasets search | Completed | Advanced
Scenario3- 3.5 | Data preparation with Query Builder | Completed | Advanced
Scenario3- 3.6 | Data analysis with Algorithm Execution Container | Completed | Advanced
Scenario3- 3.7 | Report visualisation | Completed | Advanced
Scenario3- 3.8 | Report sharing | Completed | Advanced

Table 2-2: AEGIS Demonstrators test cases

A series of evaluations have been made during the progress of the demonstrator development. Evaluations concerning the demonstrators (and not the AEGIS platform) are summarized in this section, while evaluations concerning the platform are provided in section 3.

With respect to the automotive demonstrator an evaluation involving a focus group of drivers was conducted after the development of the medium demonstrator, safe-driving indicator, targeting at drivers as beneficiaries. In general, the drivers (who were also providing their driving data for the demonstrator) understood the concept and liked the services provided, visualising trips and driving styles and providing a risk score as a safety benchmark on trip and driver level. They would use such a service to further improve their driving style in order to learn how to driver safer, if it is commercially available. Furthermore, they provided a list of ideas and suggestions on how to further improve the demonstrator including e.g. self-exploration of the data using a mobile application for drivers, increase the number of interactive elements of data exploration and visualisation, provide further contextual information to drivers, benchmark the own driving style against the driving styles of other drivers and detect also speeding events. Many of their ideas and suggestions (e.g. benchmarking of driving styles, mobile driver app) have been implemented in the later phases of the project.

Another evaluation of the automotive demonstrator was conducted involving a focus group of people related to traffic planning after the development of the advanced demonstrator, regional driving safety risk estimator, targeting at such stakeholders as beneficiaries. This evaluation
was conducted together with a focus group of persons knowledgeable in traffic planning to generate additional feedback on how the experts experienced the service. Results have shown that in general the information provided to the traffic planning experts in the dashboard and the approach pursued are perceived to be useful to better understand driving risks within urban areas and to further improve interventions to mitigate them. The intense discussion between the experts triggered a plethora of useful ideas, many of them going even beyond the scope of the demonstrator. In general, the experts were interested to get more contextual information about areas of Graz covered by the heatmap and about the underlying trip data. Displaying road use as an (additional) heatmap or route-based visualisation would be a great add-on. Having a context menu leading to further information while marking certain areas of the heatmap would be interesting to better judge the severity of the indicated risks. Traffic flow prediction and visualisation were topics of great interest for the participating experts. Adding a speeding event (besides braking, acceleration and curving/cornering) would be interesting as speeding leads to many risky situations and accidents. Too many events are used in the calculation of the heatmap (based on too sensible set thresholds for event detection, which was required for better testing the AEGIS Platform).

With respect to the smart home and assisted living demonstrator, an internal focus group was organized among data scientists and developers after the development of the medium demonstrator, notifications and alerts for (at-risk) individuals & smart home comfort profiling and notifications, discussing advantages of the service offered, perception of the development of the service application, and the level of refinement of the offered services. One major result was the perceived usefulness of the data scientists that the utilization of the big data analytics platform really provides benefits such as high scalability of services offered or the utilization of components to help overcome difficulties in exploring and visualizing data. The demonstrator development process has so far been relatively free from unexpected issues. The service application workflow was stable and streamlined and the UI experience of the demonstrator was perceived to be good supporting both mobile and web interfaces.

Another evaluation of the smart home and assisted living demonstrator with a focus group of healthcare experts was conducted after the development of the third demonstrator version. These persons were experienced users of (similar) healthcare services and so collecting their unbiased feedback was very valuable. The participants showed a high interest in the pursued approach using the features of the AEGIS platform (especially the notebook feature), as well as on the designed classification model and its interactions with the underlaying rule engine. They acknowledge the persona approach, the classification of the (at-risk) individuals on these personas and the ability of a care service provider to define a set of medical rules on each persona. They also liked the utilisation of wearable and smart home device for a near to real-time monitoring of an individual’s well-being and home conditions. The evaluation also led to many new ideas on how to enhance the demonstrator with new features, which will be considered in the exploitation phase. They also foresee a number of challenges to solve such as challenges related to the continuous data retrieval, processing and analysis, extensibility of the data pipeline with new devices, import of the anonymized SHAL database in AEGIS with growing database size, data protection and anonymization, and the inclusion of a 'right to be forgotten'.

With respect to the insurance demonstrator an internal focus group involving data scientists and demonstrator developers has been conducted after the development of the medium demonstrator. One result was the effective run of the query builder and visualiser components.
in order to be provided with a better overview on customers affected by particular (detected) events using geographic maps. A developed web application allows a seamless information exchange between the actors involved in the process. Data scientists have gained more experience of using the platform and problems encountered during the first version of the demonstrators have been fixed. The status of the insurance demonstrator is on a good direction with clearly defined steps. The event detection tool has been trained in Italian language and the respective machine learning algorithm has been significantly improved. The anonymiser was exploited for managing sensitive in-house data following the rules of the ethical advisory board. Furthermore, the developed mobile app was tested and found to work satisfactorily. Finally, suggestions to new functionalities such as enhanced technical support in the app, or information about traffic and statistics about vehicle usage were made. Security and privacy regulations seem to be the main issues, such as customers as app users have to give consent for their geolocation to be used. Finally, a close cooperation between HDI and the technical team has found to be fundamental and fruitful during demonstrator development.

Another evaluation of the insurance demonstrator was conducted based on a workshop held at HDI with insurance experts and data scientists of different business domains, where the AEGIS project and the insurance demonstrator were presented, followed by a demo session of the insurance demonstrator and a feedback phase from the experts. On very interesting wish of the participants was to explore the AEGIS platform on its own, showing the huge interest in the project. All of them recognized that the implemented scenarios bring value to the company and the insurance business. Work practices, which currently need many manual steps based on personal relationships could be automated using the features developed. In general, the experts were satisfied with the workflow implementation and the insurance demonstrator (as well as the use of the AEGIS platform) were positively evaluated from a business perspective.

2.2. Automotive Demonstrator overview, achievements and lessons learnt

The automotive demonstrator was developed according to three different scenarios, (1) broken road indicator, (2) safe driving indicator, and (3) regional driving safety risk estimator. All three scenarios have been successfully implemented and evaluated.

The figure below shows the automotive demonstrator project page as well as all datasets created on the AEGIS platform, which are necessary to run the three demonstrators. In total 2,163 trips were processed and analysed on the AEGIS platform resulting in ~47 GB Vehicle Raw Data, ~17 GB TripData_raw and ~18 GB TripData_prepared and ~22 GB TripDensity in the corresponding folders as shown in the figure below.
The automotive demonstrator (V1, V2, V3) features a comprehensive data processing pipeline implemented on the AEGIS platform. Thereby the four steps (01) data extraction & preparation, (02) event calculation & aggregation, (03) analytics & result generation and (04) output preparation are executed before the result is visualized, using the AEGIS Visualiser component. The figure below explains all created datasets (raw_data, weather_data, raw_trips, trip_data, trip_events, trip_densities, event_densities, road_damage_quality, regional_risk_density, driving_score_data, final_potholes, final_regional_risk, final_trips, and final_driving_scores) including their substructures as well as all python scripts executed to transform data from one stage to another. The figure also includes all scripts required to compute the required data for the Visualiser component (01a_extract_trips, 01b_prepare_trips, 02a_compute_events, 02b_compute_trip_density, 03_compute_damage_density, 03b_compute_event_density, 03_compute_risk_density, 03b_compute_driving_scores, 04_compute_pothole_vis, 04_compute_regional_risk_vis, 04_compute_trip_vis, and 04_compute_driving_score_vis). Finally, the script visualizer is run to show the results.
Figure 2-2: Data workflow of the Automotive Demonstrator (V1, V2, V3)
Though each demonstrator version can be used independently, they require similar data extraction & preparation as well as event calculation and aggregation steps using the data processing pipeline presented above.

The first demonstrator, **broken road indicator**, creates either a marker visualisation of road damage (potholes), or a heatmap visualisation of road damage (potholes) for the Visualiser, as shown in the figure below, to further discuss it with road maintenance experts.

![Marker Visualisation of Road Damage (all drivers)](image1)
![Heatmap Visualisation of Road Damage (driver-id3)](image2)

**Figure 2-3: Broken Road Indicator (Automotive Demonstrator V1)**

The second demonstrator, **safe driving indicator**, creates a (combined) trip and marker visualisation of all trip-specific safe driving events detected within one trip to further discuss it with drivers.

![Marker and Route Visualisation of Safe Driving Events (one trip)](image3)

**Figure 2-4: Safe Driving Indicator I (Automotive Demonstrator V2)**

Furthermore, the second demonstrator, **safe driving indicator**, calculates individual risk scores for all trips and all drivers, as well as the aggregated risk scores, for all drivers for driver/trip...
benchmarking, which can be visualised with time series visualisation (all risk score of all trips) and scatter plot (all aggregated risk scores of all drivers) to further discuss it with drivers.

**Figure 2-5: Visualisation of all individual risk scores of a single driver (Automotive Demonstrator V2)**

The third demonstrator, **regional driving risk estimator**, creates a heatmap of all safety relevant events identified in the complete data for the visualizer to further discuss it with traffic planning experts.

**Figure 2-6: Visualisation of aggregated risk scores of all drivers (Automotive Demonstrator V2)**
During the implementation of the automotive demonstrator, many challenges were raised, which are described in the deliverables D5.3-D5.5. One particular challenge to mention was the alignment of the demonstrator development with the platform development. However, during the progress of the project, the cooperation quality between demonstrator developers and the platform developers highly increased. Furthermore, after exploring the platform’s capabilities by developing the first demonstrator, broken road indicator, a redesign of data structures, workflows, and scripts has been initiated, as experiences with the AEGIS methodology have increased, too. This redesign process lasted until the end of the AEGIS project and the resulting final data workflow of the automotive demonstrator has been already presented in this deliverable.

The evaluation with users (of the services) generated many relevant insights; however, not all of them can be implemented within the project’s runtime. Many of them are out of scope and therefore relevant for the latter exploitation phase of the project with respect to automotive data science. They will be carefully considered in further exploitation actions. Furthermore, the three demonstrators enabled VIF to meet virtually and face to face with external stakeholders of the automotive ecosystem (car manufacturers, engineers, and technology providers) and engage them in commercially relevant discussions for future cooperation in automotive data science projects. This is something which would not have been possible without the results of the AEGIS project, as these clearly indicates the data science competencies of VIF leading to three relevant demonstrator’s versions.

During the cause of the demonstrator implementation, even a mini use case on the topic of driver/vehicle classification was implemented to explore the machine learning/AI capabilities of the AEGIS platform in the course of a computer science seminar project using the data of the automotive demonstrator too. Thereby the following research questions have been formulated to set the direction: Can a deep neural network identify the driver given a trip segment of raw data? Is the network able to construct its own relevant features? Is the network able to analyse only the comparable parts when given a larger segment? What accuracies can be achieved? How do recurrent networks compare to non-recurrent ones? The figure below
shows already quite promising results for driver classification, using the data of vehicle standstill events calculated on the AEGIS platform. The project will continue till the end of July 2019.

Finally, a lightweight mobile application for drivers has been developed as a further mini demonstrator to visualise their driving data (computed on the AEGIS platform) without having to access the AEGIS platform. The figure below shows screenshots of the application running on an Android phone (Samsung Galaxy S9 Plus). This was also a requirement voiced by the voluntary drivers providing data to the AEGIS platform so that they can investigate their driving data without the assistance of a VIF data scientist operating the AEGIS platform.

The driver can visualise an overall risk score, as well as the individual risk scores of the last 20 trips. Finally, the driver receives a summary of the total number of trips examined, the total distance travelled, the total time driven, the total number of events detected in the driving data and the total number of events detected per category (brake, acceleration, standstill), as shown in the three screenshots of the driver app below.

Figure 2-8: Automotive Demonstrator: Mini Use Case Driver Classification
Figure 2-9: Automotive Demonstrator: Driver App I

The driver can select a trip from all trips of today, the current week or the current month and visualise it with the mobile app. All trip-specific events and a trip-specific risk score are displayed. Finally, the driver can compare his individual driving style with that of his colleagues, as the three screenshots of the driver application below show.

Figure 2-10: Automotive Demonstrator: Driver App II
2.3. Smart Home and Assisted Living Demonstrator overview, achievements and lessons learnt

As part of the Smart Home and Assisted Living (SHAL) demonstrator, the responsible partners developed an infrastructure that collects health, well-being, activity and smart-home data of (at risk) individuals and makes use of the AEGIS platform to perform certain analyses and retrieve data towards producing services that are able to improve the safety of such individuals. Although the end-beneficiary of the service are (at-risk) individuals and their informal carers, the services are offered to Care Service Providers (CSPs) that are in need of a system that is able to collect data from individuals they treat, and to use big data infrastructures such as the AEGIS platform to minimise the cost of running in-house analytics.

In more detail, the system that has been developed makes use of the data coming from individuals and that are (after consent) offered by those individuals to CSPs, in order to help data analysts working in such organisations to combine such heterogenous data coming from diverse proprietary and public sources, and conduct analyses that may result to better motoring and more evidence-based recommendations to such user groups, targeting also personalisation as well.

During the course of the demonstrator, the responsible partners successfully developed the demonstrator envisaged in the project’s DoA, and in total a number of 24 test cases were executed in the three stages of the demonstrator, which has been constructed in such a procedural manner to allow for better testing the assumptions that were made during its initial conceptualisation.

Within the context of the SHAL demonstrator the key main artefacts produced were: (a) the SHAL Web App that is offering the extended list of functionalities to all stakeholders of the demonstrator (CSPs, informal carers and individuals), (b) the mobile applications that were developed for both Android and iOS devices that is offering the main functionalities for the (at-risk) individuals and (c) a sophisticated lightweight but still powerful backbone was developed in order to orchestrate the rest of components, but also to support the required backend functionalities such as the storage, homogenisation and anonymisation of the collected data and the connectivity with the AEGIS platform in order to leverage from the platform’s big data capabilities.

The Web App is offering an interface for both individuals, as well as their CSPs (and informal carers) to visually explore some of the data that are collected by sensors that individuals register to the system (wearables, smart home sensors, etc). From the point of individuals, they are served with a dashboard that contains such information and allows them to share it with certain CSPs, either anonymously (in that case personal data is stripped from personal identifiable information and is aggregated under a pre-selected “persona” under which the user fits best), or even in full, by granting access to their dashboard to CSP. Moreover, they are also able to link to carers, which are receiving only notifications that have to do with the wellbeing of each individual (the exact same notifications that individuals receive), but are not able to access any other data that relates to those people.
For CSPs, the web app provides a similar interface, where they are able to witness stats that relate to “personas”, as showing in the next figures, where individuals are anonymously categorised and aggregated stats of the persona are displayed. Options exist to extract the data of the persona for further analysis on the AEGIS platform, or locally, as well as to manually send notifications to groups of a persona.

Figure 2-11: SHAL Web App individuals’ dashboard

Figure 2-12: SHAL Web App list of Personas with action buttons
The mobile applications are offering an interface where at risk individuals can check some of the more important values that have to do with their daily activities and biometrics, such as heartrate, blood pressure, steps walked, etc., while also information from smart home devices that they have linked to their profile is displayed.
The App works as a data hub, which collects information from various sources that are connected to a smartphone and pushes these data to the backend SHAL server. Moreover, it retrieves other data available in SHAL to showcase to the user, such as for example indoor and outdoor conditions.

Moreover, the app allows the receipt of notifications, which are pushed to such individuals either automatically based on the trained models that relate to health issues or smart home comfort zones, or manually, once this is instructed by a CSP to inform the people monitored.
Apart from the successful technical results of the demonstrator, that were tested during this period, of special importance for the partners implementing the demonstrator was the feedback received by potential stakeholders of the domain that the demonstrator addresses, which are CSPs that have been engaged in face-to-face meetings and in an online focus group (webinar), as well as individuals that have tested the infrastructure as volunteers. Both those groups were very positive both on the concept and the solution developed and were eager to use the solution to understand better its operation and its potential impact.

From a technological perspective, the consortium has overcome some challenges that had to do with data homogenisation in order to facilitate analyses, while the privacy and anonymisation features have been successfully tackled by the chosen approach. The existence of the AEGIS platform and its analytics processing features, as well as the presence of some open datasets allowed the partners to develop a solution that is lightweight and requires little processing resources, which is ideal for CSPs of medium to small sizes, that do not want to invest in computing resource, but outsource data analysis and data linkage tasks. One major technical challenges faced, which has not been fully tackled due to external conditions have to do with the real-time flow of information (as most sensors and mobile devises do not support the continuous streaming of measurements mostly due to energy constrains).

CSPs commented that the overall concept as well as the insights that are delivered by the platform see to be of great importance, as they allow CSP personnel after performing certain analyses to better categorise subjects, not only on pure medical and health data, but also on other activity which are in the majority of cases unknown to CSPs and they are not able to track them down. Although such data are by themselves not of great significance to CSPs, when combined with the medical data and the health profiles of individuals, they become an extra
asset to understand the behaviour of individuals and to offer more informed and personalised recommendations.

Individuals are also providing very positive comments for the SHAL idea, both as a concept, but also liked the current deployment which allows them to understand better their day-to-day routines, and also receive recommendation for tasks they would otherwise either ignore. In principle all of the individuals which voluntarily participated in the demonstrator were concerned about the use of their personal data, nevertheless the anonymisation method used in SHAL, as well as the option for them to choose with who to share their data and to which extent (e.g. anonymised or actual data), played down their concerns, whereas they all agreed to share more data for getting more personalised notifications.

Finally, it has to be noted that the overall operation of the demonstrator has showcased that there is a huge need for data availability and homogenisation, as the more and better data available, the better the insights can be. As an example, the availability of (open) APIs from third party sources that provide information about weather and environmental conditions are at the moment quite high, nevertheless the data they provide are quite poor; they are restricted to certain regions, updating times are slow and data frequencies are quite low, while the accuracy of those is poor as well. In principle, the partners believe that the availability of such data would further improve the impact of infrastructures that have similar concepts with SHAL, and that the sharing of those data could be facilitated by a platform as AEGIS, where data providers would be able to distinguish a real need for their data, and work towards improving them in return for some value.

2.4. Insurance Demonstrator achievements and lessons learnt

The overall goal of the AEGIS Insurance demonstrator is to exploit the AEGIS platform big data technologies in order to access and analyse information coming from diverse and heterogeneous data sources including the in-house data (e.g. customers location, customers portfolio). Exploring with the AEGIS tools weather, news and crime open data, the HDI data scientists would be able to manage in an efficient way events (that are going to happen or just happened), while the use of the AEGIS analytic tools would allow the company to set a strategy to minimise the impact of the event on the company itself, while offering a support to the customers.

Summarizing, the expected benefits of the AEGIS adoption by HDI Assicurazioni include the possibility to rapidly analyse and process huge volumes of cross domain data, an improvement of the customer’s satisfaction through personalised offering and information about natural or social events of interest. Moreover, AEGIS services would constitute a key for the development of an accurate and successful business plan through the pricing analysis and the analysis of the trend of the company.

The Insurance demonstrator has been presented with three scenarios, as summarized in Figure 2-18:

- **Scenario 1**: Financial impact, customer support and services
- **Scenario 2**: Personalised early warning services for asset protection
- **Scenario 3**: Marketing strategy and pricing support services
Each scenario aims to profit of the AEGIS platform functionalities towards:

- Improving the customer centricity (Scenario 1 and 2),
- Preventing frauds (Scenario 1),
- Exploiting the internet of things and real-time notifications (Scenario 1 and 2),
- Setting the pricing and selling strategies (Scenario 2 and 3),
- Performing risk analysis (Scenario 3).

The main technical components of the Insurance demonstrator are depicted in Figure 2-19:

- The Anonymiser that is an offline tool locally installed. With the use of the Anonymiser, the Data Scientist anonymizes the in-house datasets before their upload on the AEGIS platform. This data pre-processing is needed for compliance with the privacy regulation. The Anonymiser is provided as an offering of the AEGIS project.
- The HDI Mobile App that provides the customers geolocation to the Web App (Scenario 2) and receives notifications from the Web App (Scenario 1 and 2). A Google service (Firebase) is called by the Mobile App Back-End in order to enable the push notifications. The HDI Mobile App was enhanced for the scenario execution with the two functionalities described.
- The HDI Web App that is an HDI environment (accessed by all the HDI actors defined) that allows the communication between different roles. It has been developed within the
project by the demonstrator in order to implement the correct workflows between the actors in the three scenarios. From the HDI Web App it is possible to gather the geolocation data from the HDI Mobile App (Scenario 2). Finally, the HDI Expert is informed that an event has been detected (Scenario 1 and 2) through a notification that is sent by the AEGIS platform.

- The Event Detection Tool/ Data Harvester (Scenario 1 and 2): through the creation of a Job on the AEGIS platform, the Data Harvester uploads the event data (detected by the Event Detection tool) on the AEGIS platform. The Kafka service of AEGIS detects this and sends notifications to the Web App.
- The AEGIS platform notebooks (Query Builder, Algorithm Execution Container and Visualizer) are the actual tools used by the Data Scientists to perform his/her analysis (Scenario 1, 2 and 3).

Figure 2-19: Insurance demonstrator - technical components and user(s) for each component

Summarising, from this list of components, the following have been developed within the context of the demonstrator for the execution of the three scenarios of the Insurance demonstrator:

- the HDI Web App, for the sharing of information within the HDI actors. The application has the aim to ease the communication between different departments of the company.
- the existing HDI Mobile App has been enriched with some functionalities, in particular the geolocation feature and push notifications.
- the Event Detection Tool has been trained for four events (flood, whirlwind, hailstorm and socio-political) in the Italian language.
- the Query Builder and the Visualizer have been customized to fulfil the demonstrator requirements: customers priority feature, the earthquake risk Heatmap and the customers pin related with the kind of policy held.

It is important to point out that the scenarios have been developed and executed in different times according to the demonstrator implementation and execution plan: on M18 for the early demonstrator, Scenario 1 was executed, on M24 for the medium demonstrator Scenario 2 was executed, and on M30 for the advanced demonstrator, Scenario 3 was executed. They have been deeply described respectively in D5.3, D5.4 and D5.5 and their execution results reflected the version of the platform at that time.

While the platform matured, some corrections have been done to the workflows developed. The major achievement that has not yet been described is the real-time notification service that enhances the workflow of the Scenario 1 and 2. Through a Job on the AEGIS platform in fact, from the HDI Web App it is possible to receive notifications by the Kafka service of the platform when the Event Detection tool detects a new event. The increasing integration of the platform and its components, lead to store the events data detected by the Event Detection Tool as files through the Data Harvester on AEGIS. If these files are updated, the Kafka service subscribed with a Java class by the HDI Web App retrieves this change and pushes a notification.

The scenarios execution has been performed by GFT and HDI, however during the development phases the consortium partners that developed the three notebooks, the Anonymizer and the Even Detection Tool, namely UBITECH, NTUA and SUITE5, collaborated with the demonstrator development team to facilitate the process.

For the final evaluation of the Insurance demonstrator, as described also in D5.5, a workshop with six HDI people that were not involved in the team of the AEGIS project has been organised. The participants were three Data Scientists/Analysts and three Experts (belonging to different HDI departments: claims, commercial, portfolio). They had the opportunity to have an overview of the project with a focus on the Insurance demonstrator, a demo and some time to use the AEGIS platform and the HDI Web App.

Due to the privacy regulation, as stated also in the previous deliverables of WP5, it was not possible to involve the HDI customers in the scenarios execution since a large amount of them could have signed a specific consent to make the results significant (for further details see section 6.3.1 of D5.5 - Demonstrators Evaluation and Feedback – v3).

During the three scenarios development phase, it was acknowledged that apart from the ‘accessory tools’, the platform functionalities have been exploited almost as they are offered by the AEGIS platform and only minor changes were needed to the Jupyter notebooks. This easy adaptation to different scenarios from the available default options of the three notebooks has been pointed out also during the workshop as the extended usability level of the platform. Furthermore, during the workshop the events identified for the EDT training have been evaluated as suitable which highlight the usefulness of the Event Detection Tool. All the workshop’s participants showed interest on the project and on the platform, and evaluated successfully the usability both of the AEGIS platform and of the HDI Web App. They expressed their interest to use more the platform; in fact the two Data Scientists that participated to the workshop have already started using the platform.
3. AEGIS Platform Final Evaluation

3.1. Quantitative Evaluation

This section presents an overview of the quantitative evaluation of platform that was conducted during the project. As instructed by the AEGIS evaluation framework, which was documented in deliverable D5.1 and updated in deliverable D5.2, the holistic evaluation of the platform was performed in three different iterations, one for each of the phases of the demonstrators, giving valuable feedback to the AEGIS platform developers during the different implementation phases of the AEGIS platform. Especially for the third evaluation of the platform, as it was performed in the final month of the project (M30), in which the final release of the AEGIS platform was also delivered, the valuable feedback collected will drive the enhancements and adjustments that will be introduced in the platform during the project’s exploitation phase.

The quantitative evaluation of the platform was based on a set of technical Key Performance Indicators (KPIs) that were defined based on the eight characteristics (and its’ sub-characteristics) of the international standard ISO/IEC 25010:2011: Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE). The scope of the quantitative evaluation was to provide the required quality assurance and control during the implementation phase of the platform.

Table 3-1 presents the aggregated results of the three evaluations that were performed during the evaluation phase. From these results, the following main insights were extracted:

- As the implementation of the platform progressed, some key characteristics were significantly improved towards the final version of the platform, such as the performance efficiency, the usability and the reliability characteristics. This indicates the incremental maturity of the platform, as well as the focus of the development team to offer additional value to the platform’s stakeholders.

- For several key characteristics, the platform has showcased the added value that it offers from the early releases, such as the functional suitability, the portability and maintainability, and it is obvious that the development team focused on providing a stable platform with advanced characteristics from the very beginning. Furthermore, in some cases these characteristics were further improved indicating that the development team provided additional features towards the better addressing of the stakeholders’ requirements.

- Critical characteristics such as the security and compatibility of the platform remained in high standards from the early releases. Nowadays, security has been one of most critical aspects of any software solution and the results indicate the robustness and effectiveness of the applied security mechanisms. Additionally, as the technologies change at such rapid rate, the high-level of compatibility of a platform is also a critical aspect for its success and sustainability in the future.
<table>
<thead>
<tr>
<th>Sub-characteristics</th>
<th>KPIs</th>
<th>Calculation Type</th>
<th>Mandatory/Optional</th>
<th>1st evaluation</th>
<th>2nd evaluation</th>
<th>3rd evaluation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Functional suitability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Functional completeness</td>
<td>Portion of completed User Stories</td>
<td>[Completed User Stories] / [Iteration Cycle of User Stories] * 100%</td>
<td>M</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>All use cases planned for each version were successfully executed.</td>
</tr>
<tr>
<td>Functional correctness</td>
<td>Portion of User Stories without reported bugs</td>
<td>[Completed User Stories without bugs] / [Iteration Cycle of User Stories] * 100%</td>
<td>M</td>
<td>90%</td>
<td>93%</td>
<td>95%</td>
<td>During each implementation phase a small series of bugs were identified and they were successfully addressed.</td>
</tr>
<tr>
<td>Functional appropriateness</td>
<td>Straightforward task accomplishment</td>
<td>Are tasks completed without the use of unnecessary steps? [Yes/No]</td>
<td>O</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Due to the nature of the accomplished tasks, assistance from the respective persons is required in some cases.</td>
</tr>
<tr>
<td><strong>Performance efficiency</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time behaviour</td>
<td>Average latency</td>
<td>[Total response time] / [Number of requests]</td>
<td>M</td>
<td>~1.2 sec</td>
<td>~1.1 sec</td>
<td>~1.0 sec</td>
<td>Average latency was measured with tools such as Chrome Dev Tools.</td>
</tr>
<tr>
<td>Resource utilisation</td>
<td>Mean CPU Utilisation</td>
<td>$[\Sigma[%CPU utilisation probes]] / [\text{Number of probes}]$</td>
<td>M</td>
<td>&lt;45%</td>
<td>&lt;40%</td>
<td>&lt;38%</td>
<td>Based on the resource monitoring tool of the platform</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------</td>
<td>---------------------------------------------------------------</td>
<td>---</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Mean memory usage</td>
<td>$[\Sigma[\text{RAM Megabytes used in each probe}]] / [\text{Number of probes}]$</td>
<td>M</td>
<td>&lt;20%</td>
<td>&lt;18%</td>
<td>&lt;20%</td>
<td>Based on the resource monitoring tool of the platform</td>
</tr>
<tr>
<td></td>
<td>Maximum memory usage</td>
<td>Maximum % RAM Memory utilisation recorded</td>
<td>M</td>
<td>37%</td>
<td>40%</td>
<td>42%</td>
<td>Based on the resource monitoring tool of the platform</td>
</tr>
<tr>
<td></td>
<td>Maximum processing power used</td>
<td>Maximum % CPU utilisation recorded</td>
<td>M</td>
<td>90%</td>
<td>90%</td>
<td>90%</td>
<td>As the resource management is performed by YARN (see deliverable D3.5), the appropriate resource allocation is always performed according to the provided configuration.</td>
</tr>
<tr>
<td>Capacity</td>
<td>Maximum file size upload</td>
<td>Total number of Kilobytes of files</td>
<td>M</td>
<td>250MB</td>
<td>450MB</td>
<td>5.1 GB</td>
<td>Note: This is size of the current biggest individual file available.</td>
</tr>
<tr>
<td></td>
<td>Maximum file system size (^1)</td>
<td>Total number of Kilobytes of files</td>
<td>M</td>
<td>76GB</td>
<td>89GB</td>
<td>163.5 GB</td>
<td>Note: This is the current size of HopsFS that can</td>
</tr>
</tbody>
</table>

\(^1\) AEGIS platform utilises the distributed file system HopsFS. Thus, the database size metric was modified.
### Compatibility

<table>
<thead>
<tr>
<th>Co-existence</th>
<th>Ability to Co-Exist (host in a single environment)</th>
<th>Can the AEGIS platform operate in shared environment? [Yes/No]</th>
<th>O</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interoperability</td>
<td>% of APIs coverage</td>
<td>[Number of integrated systems exposing or consuming data through API] / [Total number of integrated systems] * 100%</td>
<td>M</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Ability to handle different datasets</td>
<td>Can the AEGIS platform consume datasets from different formats (e.g. CSV, JSON, XML files)? [Yes/No]</td>
<td>M</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Can the AEGIS platform provide datasets in various formats (e.g. HopsFS can store)</td>
<td>M</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

scale according to the needs of the project.

All integrated components / services are integrated through APIs.

No limitations on the file formats HopsFS can store. Files can be processed using the appropriate libraries by the data scientist.

No limitations on the file formats HopsFS can store and provide.
<table>
<thead>
<tr>
<th>Usability</th>
<th>% Positive feedback on appropriateness based on the available documentation</th>
<th>[Number of positive response] / [Total number of responses] * 100%</th>
<th>O</th>
<th>Not applicable.</th>
<th>80%</th>
<th>90%</th>
<th>The final version of the documentation of the platform is delivered as part of D4.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriateness recognisability</td>
<td>% Coverage of features with learning documents</td>
<td>[Unique number of help documents mentioning a feature] / [Total number of features available] * 100%</td>
<td>M</td>
<td>Not applicable.</td>
<td>90%</td>
<td>100%</td>
<td>In the final version of the documentation all the features of the platform are properly documented.</td>
</tr>
<tr>
<td>Technical Learnability</td>
<td>% Coverage of features with learning documents</td>
<td>[Unique number of help documents mentioning a feature] / [Total number of features available] * 100%</td>
<td>M</td>
<td>Not applicable.</td>
<td>90%</td>
<td>100%</td>
<td>In the final version of the documentation all the features of the platform are properly documented.</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>Dashboard availability</td>
<td>Is there an available dashboard or wizard with easy navigation? [Yes/No/Partially]</td>
<td>O</td>
<td>Partially</td>
<td>Partially</td>
<td>Yes</td>
<td>The platform offers an intuitive and user friendly UI</td>
</tr>
<tr>
<td>User error protection</td>
<td>% Coverage of input fields with error protection methods</td>
<td>[Number of error protected fields] / [Total number of critical input fields] * 100%</td>
<td>M</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>All input fields in the UI are protected.</td>
</tr>
<tr>
<td>User interface aesthetics</td>
<td>% Positive feedback on user interface aesthetics poll</td>
<td>O</td>
<td>80%</td>
<td>85%</td>
<td>90%</td>
<td>The platform offers an intuitive and user friendly UI, with enhanced aesthetics and user experience.</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------------------------------</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Number of supported screens] / [Total number of different screens] * 100%</td>
<td>M</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>No inaccessible or malformed screens were identified.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Responsive</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Number of supported screens] / [Total number of different screens] * 100%</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technical Accessibility</td>
<td>WCAG 2.0 Conformance Level(^2)</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[None/ A/ AA/ AAA]</td>
<td>M</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
<td>M</td>
<td>27</td>
<td>31</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum Concurrent users</td>
<td>M</td>
<td>Max Containers allocated = 12, Max applications = 5 Concurrent application, requests to filesystem =</td>
<td>Max Containers allocated = 17, Max applications = 9 Concurrent application, requests to filesystem =</td>
<td>Max Containers allocated = 22, Max applications = 11 Concurrent application, requests to filesystem =</td>
<td>Based on the resource monitoring tool of the platform</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum number of concurrent users recorded</td>
<td>M</td>
<td>Max Containers allocated = 12, Max applications = 5 Concurrent application, requests to filesystem =</td>
<td>Max Containers allocated = 17, Max applications = 9 Concurrent application, requests to filesystem =</td>
<td>Max Containers allocated = 22, Max applications = 11 Concurrent application, requests to filesystem =</td>
<td>Based on the resource monitoring tool of the platform</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maximum number of simultaneous requests</td>
<td>M</td>
<td>Max Containers allocated = 12, Max applications = 5 Concurrent application, requests to filesystem =</td>
<td>Max Containers allocated = 17, Max applications = 9 Concurrent application, requests to filesystem =</td>
<td>Max Containers allocated = 22, Max applications = 11 Concurrent application, requests to filesystem =</td>
<td>Based on the resource monitoring tool of the platform</td>
<td></td>
</tr>
</tbody>
</table>

\(^2\) WCAG 2.0: https://www.w3.org/WAI/WCAG20/quickref/
<table>
<thead>
<tr>
<th>Availability</th>
<th>% Monthly availability</th>
<th>11 concurrent requests.</th>
<th>15 concurrent requests.</th>
<th>20 concurrent requests.</th>
<th>All downtime recorded was due to infrastructure upgrades</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[1-{Downtime in minutes} / [Total month minutes]} * 100%</td>
<td>M</td>
<td>&gt;95%</td>
<td>~97%</td>
<td>~98%</td>
</tr>
</tbody>
</table>

| Success rate | [Number of correctly completed requests] / [Total number of requests] | M | ~90% | ~95% | ~96% | All problematic requests were successfully addressed with bug fixing. |

| Fault tolerance | % of identified Software problems affecting the platform | [Critical Software Issues] / [Total number of Software faults detected] * 100% | M | ~25% | ~22% | ~19% | All problems have been identified and fixed. |
|                |                                                      |                                         |                              |                          |                          |
|                | % of identified Hardware problems affecting the platform | [Critical Hardware Issues] / [Total number of Hardware faults detected] * 100% | M | 100% | 100% | 100% | Critical hardware issues were identified and fixed in short time. |

| Recoverability | Mean recovery time from Software problems | [Total recovering time from Software issues] / [Total number of Software issues in need of recovery] | M | ~1 hour | ~1 hour | ~1 hour | |

WP5 – AEGIS Data Value Chain
Early Community Demonstrators  © AEGIS Consortium  Page 42 of 92
<table>
<thead>
<tr>
<th></th>
<th>Mean recovery time from Hardware problems</th>
<th>[Total recovering time from Hardware issues] / [Total number of Hardware issues in need of recovery]</th>
<th>M</th>
<th>~ 1 hour</th>
<th>~ 1 hour</th>
<th>~ 1 hour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Security</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidentiality</td>
<td>Incidents of ownership changes and accessing prohibited data</td>
<td>Number of recorded incidents</td>
<td>M</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Integrity</td>
<td>Incidents of authentication mechanisms breaches</td>
<td>Number of recorded incidents</td>
<td>M</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>Non-repudiation</td>
<td>% Activities reporting</td>
<td>[Number of log categories] / [Total number of system operations]</td>
<td>M</td>
<td>90%</td>
<td>95%</td>
<td>97%</td>
</tr>
<tr>
<td>Accountability</td>
<td>User actions traceability</td>
<td>Are usernames included in each activity log entry uniquely? [Yes/No]</td>
<td>M</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Authenticity</td>
<td>Level of User authenticity</td>
<td>Can you identify that a subject is the one it claims to be? [Yes/No/Partially]</td>
<td>M</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
## Maintainability

<table>
<thead>
<tr>
<th>Modularity</th>
<th>% of modularity</th>
<th>Modularity</th>
<th>100%</th>
<th>100%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of modularity</td>
<td>[Number of components that can operated individually] / [Total number of components] * 100%</td>
<td>M</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reusability</th>
<th>% of reusable assets</th>
<th>Reusability</th>
<th>100%</th>
<th>100%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of reusable assets</td>
<td>[Number of assets that can or are reused] / [Total number of assets] * 100%</td>
<td>M</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Analysability</th>
<th>Level of analysability</th>
<th>Analysability</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of analysability</td>
<td>Can the changes in the performance of the AEGIS platform be efficiently evaluated after each upgrade? [Yes/No]</td>
<td>O</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The system offers monitoring tools with performance indications.

<table>
<thead>
<tr>
<th>Modifiability</th>
<th>% of update effectiveness</th>
<th>Modifiability</th>
<th>95%</th>
<th>95%</th>
<th>97%</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of update effectiveness</td>
<td>[Number of updates performed without operational issues] / [Total number of updates] * 100%</td>
<td>M</td>
<td>95%</td>
<td>95%</td>
<td>97%</td>
</tr>
</tbody>
</table>

The updates were performed successfully with minor issues.

<table>
<thead>
<tr>
<th>Testability</th>
<th>Level of testing</th>
<th>Testability</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of testing</td>
<td>Are tests able to probe the behaviour of the</td>
<td>M</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The system offers monitoring tools with performance indications.
<table>
<thead>
<tr>
<th>Portability</th>
<th>AEGIS platform? [Yes/No]</th>
<th>Portability</th>
<th>AEGIS platform? [Yes/No]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adaptability</strong></td>
<td>Mean number of errors per hardware change/upgrade</td>
<td>M</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>[Total number of errors recorded] / [Total number of hardware changes]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean number of errors per software change/update</td>
<td>M</td>
<td>&lt;2%</td>
<td>&lt;2%</td>
</tr>
<tr>
<td>[Total number of errors recorded] / [Total number of software changes]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Replaceability</strong></td>
<td>% of software products replaceability within AEGIS platform</td>
<td>M</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>[Number of replaceable software components] / [Total number of used software components] * 100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3-1: Summary of the AEGIS platform quantitative evaluation results**
3.2. Qualitative Evaluation

This section provides an overview on the results of the qualitative evaluation of the platform performed after each demonstrator development phase. Thereby interviews and/or small focus groups with data scientists in charge of demonstrator implementation as well as with demonstrator stakeholders (e.g. end users) have been conducted. In general, all data scientists in charge of demonstrator development have already gained experiences with different tools and languages for data analysis as well as experiences in analysing data within their own industrial domains. They are well qualified to execute the implementation of the demonstrators and to provide feedback on the AEGIS platform based on their experiences from using it.

The following summarized evaluation results have been gained after the implementation of the early demonstrator, by using a very early version of the AEGIS platform, providing only basic features to support data scientists in their demonstrator development. It can be said that the data scientists liked the ‘platform as a service concept’ (Hadoop file system / Spark / Zeppelin / Jupyter). It clearly makes sense and saves data scientists a lot of work (since setting such an environment up and maintaining it requires a lot of experience and infrastructure and should not be the work of data scientists). They see in AEGIS a great platform that combines a lot of things and capabilities that people working in the data science sector need and tools that are hot in the current data industry. However, for less experienced users (e.g. business users with no background in programming) the platform might not seem as the most appropriate tool to solve their challenges. Furthermore, the extendibility of the platform was questioned, as extending the platform would in its current state just be possible with notebooks, or tools that are outside the platform. Data scientists working with geo-referenced data (as for example in automotive) were interested in been provided with more specialised tools for geographic data visualisation by the platform. Some data scientists criticised user guidance and usability of the platform as well as the platform performance, which should be both further improved. For the first version of the demonstrator, the required code was mainly developed offline (in R or Python) first and then ported to python and/or simply copied into a notebook on the platform. So far, the data scientists were already satisfied with the platform, its capabilities and the current development process (besides some technical issues where the platform development team had to be contacted).

Challenges and recommendations made after the first evaluation mainly concerned an improvement of the user interface of the platform, an improvement of the standard settings of the platform with respect to performance, a user guide for the platform to reduce initial efforts of trying out things on the platform, improved capabilities for geo-referenced data visualisation, and a redesign of the platform’s landing page.

The following summarized evaluation results have been gained after the implementation of the medium demonstrator, where a more mature version of the AEGIS platform has been used providing improved features to support data scientists.

Data scientists acknowledged the improvement of the platform in terms of comfortability and usability of the platform, as well as usefulness of the platform for performing the tasks required for the implementation of the demonstrator. Offerings of the platform were significantly increased with the list of predefined tools such as the Query Builder, the Visualiser and the Algorithm Execution Container. They also acknowledged the improved look-and-feel of the platform, which was a major issue in the first evaluation. However, they still see further
potential for graphical user interface improvement and enhancement. Furthermore, the improved Visualiser offers more capabilities for visualising geo-referenced data (e.g. heatmaps or enhancements to marker visualisations), which was requested by some data scientists (e.g. from automotive demonstrator). Another perceived major advantage of the platform update was the improved configuration of Jupyter, where certain options like adding execution memory to a project can be more easily accessed. The shift from Zeppelin to Jupyter was perceived to be a good decision as Jupyter seems to run more smoothly. One issue which has been voiced by the data scientists was the cooperation between the platform team and the demonstrator developers, which could be further improved. Demonstrators must meet certain deadlines for implementation and evaluation (and so does the platform development). If critical platform updates conflict with demonstrator deadlines, demonstrator developers face a challenge. Furthermore, the update process of the platform could be improved, because in case of an update, demonstrator owners must (re)create their user accounts on the platform, then re-create the project, and the datasets (folders), before data (files) can be restored into the correct datasets (folders) by the platform team. Finally, cooperation between the platform team and the demonstrator developers was perceived to be good, although some deadlines were not met causing some issues in the demonstrator’s processes. A closer coordination between the teams should have been considered.

**Challenges and recommendations made after the second evaluation**

Challenges and recommendations made after the second evaluation mainly concerned the alignment of the update process of the platform/testbed with the milestones of demonstrator development, a general improvement of the update process of the platform, the fine-tuning of the platform’s user interface so that it becomes even more appealing to the data scientists, automated enabling of Query Builder, Visualiser and Algorithm Execution Container tools within a project for users, improved or better documented accessing and processing of available datasets from the Jupyter notebook, improved navigation through the various functionalities and services offered by the platform, automated process of restoring user accounts/datasets following a platform upgrade, and automated import of notebooks when creating a new project.

The following summarized evaluation results have been gained after the implementation of the advanced demonstrator, where the final version of the AEGIS platform has been used. Having used the final version of the AEGIS platform for demonstrator implementation, the data scientists’ perception of the platform in terms of usefulness, usability and ease-of-use has again improved significantly. The redesign of the user interface of the platform was well received, and the data scientists acknowledged that the new interface is far more appealing, while also offering an improved user experience. It even facilitates the usage of the platform for non-experienced users focused on the data science sector (e.g. who show interactive visualisations provided by the visualiser). Furthermore, the documentation of the platform has improved significantly and reduces the training period for new platform users. The services of the platform have become more customisable and can be better tailored to the developers’ or data scientists’ needs. Furthermore, the collaboration between the demonstrator developers and the platform developers was further improved. Technical challenges during demonstrator implementation, such as restarting the core or increasing the computing power or RAM required for data processing workflows, have been quickly solved by the platform team increasing both user satisfaction and trust. Because the AEGIS platform received several refinements and updates during the demonstrator development period, minor issues were faced, especially in the user interface environment that was completely redesigned, and developers had to wait for the stable intermediate release of the platform before they could continue their development activities, which is understandable as both, the demonstrators, and the platform, are developed.
in parallel. Finally, the upgrade process has been further improved and most of the manual processes that were identified in the previous version were resolved, which was a major concern during the first and second evaluation. Thanks to the weekly technical telephone conferences demonstrator deadlines have been better aligned with platform development deadlines and updates.

**Challenges and recommendations made after the third evaluation** mainly concerned to remove any manual intervention in the update process of the platform for users, provide a guideline on how to best use the platform for own data science projects and expand the documentation, implement minor improvements in the user interface based on further experiences made while using the platform, further highlight the capabilities of the platform such as pre-defined or scheduled job execution.

### 3.3. Platform Security Assessment

As part of the final evaluation of the AEGIS platform, and as per the comments received during the first project review, the consortium decided to perform a thorough security assessment of the platform in order to ensure that the security mechanisms that are incorporated in the platform are operating efficiently and effectively. Furthermore, the scope of the security assessment conducted was to identify security weaknesses and deficiencies in order to be resolved and to provide possible recommendations on how to correct these identified weaknesses or deficiencies with the aim to reduce or eliminate any vulnerabilities.

In detail, the security assessment was performed by the specialised security team of UBITECH that analysed the security standards and tools utilised in the platform in order to: (a) identify any possible vulnerabilities, risks or threats that may result in a possible exploitation that may result in an intentional or unintentional compromise of the platform and (b) to identify the ways that these vulnerabilities, risks or threats can be exploited in order to defeat the deployed security mechanisms of the platform. A thorough penetration test was performed, which included amongst others the following methods (attacks): SQL injection, File traversal, Server-Side Request Forgery (SSRF), Insecure Direct Object References (IDOR), Carriage Return Line Feed (CRLF) injection, Cross-Site Scripting (XSS), Remote Code Execution (RCE), Cross-Site Request Forgery (CSRF) and Open Redirect.

During this thorough test, the platform performed well in the attempted attacks, as all of them failed, with the exception of the SSRF method. More precisely, during testing, a set of endpoints was discovered with a SSRF vulnerability, which could allow the attacker to force the server to perform requests on their behalf, for example invoking internal services. On the following request, the development team tried to mitigate this by searching if the word /static/ is included in the URL, but that can easily be bypassed as shown below:

```
REQUEST
GET /hopworks-api/yarnui/http://IP:8080/test;/static/ HTTP/1.1

RESPONSE
HTTP/1.1 200 OK
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 3.2 Final//EN"><html>
<title>Directory listing for /</title>
<body>
```
As it can be seen from the example, the request was performed on a test server that the security team of UBITECH controlled, however if the IP is changed to “127.0.0.1” requests can be performed on services listening on localhost. The spotted vulnerability was reported immediately to the development team of the AEGIS platform and at the time of writing of this deliverable, the development team has already started working on the corrective actions towards the elimination of this vulnerability.
4. AEGIS Platform Documentation and Adoption Guidelines

4.1. AEGIS Platform Usage Documentation

The AEGIS platform provides data management, analysis, and processing as well as user management through the use of Hopsworks integrated services. On top of that, the AEGIS platform provides tools for data anonymization and cleansing as well as data harvesting. AEGIS uses the notions of Project, Dataset, and User to enable multi-tenancy within the context of data management. The AEGIS platform is equipped with a list of services for data processing including data parallel processing frameworks such as MapReduce, Spark, and Flink, as well as interactive analytics using interactive notebooks such as Jupyter. In addition, it offers real time analytics using the Kafka service, full-text search using the Elasticsearch service, as well as development and serving of deep learning models using the Tensorflow service. Moreover, the AEGIS platform provides tools for data validation, visualisation, exploration, and processing through the use of the AEGIS integrated notebooks.

4.1.1. AEGIS Usage

The AEGIS platform can be deployed on-premise or in a cloud environment. Once deployed, the AEGIS platform is accessible through a web interface. A user can create an account that is later verified by an administrator. Once a user is verified, he/she can create projects on the platform by clicking on the New Project button as shown in Figure 4-1.

![Figure 4-1: AEGIS User Main Page](image)

Once the project is created, the user can navigate to the project main page as shown in Figure 4-2, browse its default datasets, as well as create new datasets and upload data. The project is pre-populated with a list of default datasets such as Logs, Resources, and Jupyter. The left side pane shows the different project associated settings and tools. Firstly, the user can manage the project general settings such as the description of the project by navigating to the Settings page.
Additionally, the user can view, add, and delete members of the project as well as changing members’ roles using the Members page. The activity stream page provides an overview over the changes happening during the project lifetime since its creation such as creating, sharing, and deleting a dataset. The Extended Metadata tool is part of the AEGIS metadata service and provides a user interface for users to associate metadata to the project. The AEGIS tools page provides the three default integrated notebooks offered by the AEGIS platform. The Jobs page offers the ability to create and manage data processing jobs using Spark and Flink. The Kafka page offers the ability to create and manage Kafka topics and their associated schemas. The Model Serving page provides the ability to serve your deep-learning models using the Tensorflow serving.

![Figure 4-2: Project Main Page](image)

### 4.1.2. AEGIS Jupyter Tools

The AEGIS Jupyter tools contain the three default tools provided by AEGIS to help users to easily perform data exploration, validation, processing, and visualisation, namely the Query Builder, the Visualiser and the Algorithm Execution Container, that are presented in the paragraphs below.
Query Builder

The main purpose of the Query Builder is to enable the capability of interactively defining and executing queries on top of the data available in the AEGIS platform.

Once the user clicks on the Query Builder tool, a preconfigured Jupyter notebook is opened as shown in the following figure.
The Query Builder is now in its initial state. All code cells are hidden using the hide code plugin and all output has been cleared. If the user is not interested in reviewing or altering the code (it is not required in the standard workflow) he/she may proceed with hiding the cell toolbars. In order to do that, the user can select "View" from the menu, then "Cell Toolbar" and click on the "None" option, as shown below.

![Query Builder - Hide code](image)

**Figure 4-5: Query Builder - Hide code**

By clicking on this option, the user is able to see the standard view of the Query Builder as depicted in the following screen.
In order to proceed with the usage of the Query Builder, the user must initialize the notebook. To perform this the user must run the first cell. To achieve this, the user must make sure that the first cell is selected and either click on the "Run" button from the top Jupyter menu, or press Ctrl + Enter. At this point, the user should be able to see a button appearing as shown below.

Figure 4-6: Query Builder - Standard view

Figure 4-7: Initialise Query Builder
The user should click on the "Initiate QB" button and wait for the execution to stop. When this is done, the user should see a screen like the one below, where all cell prompts have been replaced by numbers and the file selection UI has appeared.

![Figure 4-8: Query Builder options](image)

Note that the output regarding the Spark application will only appear when the interpreter is started. This means that if the "Initiate QB" button is pressed again, the message will disappear, but the tool will be still functional. Now the query builder is initialized and the user can start working with different files.

User can see a list of all "Available Datasets" in the current project. The list can also be refreshed. Select the desired dataset and click "Open dataset", this will populate the "Available Files" list.
Figure 4-9: Query Builder Local Browser

Once the user selected a file and presses the "Open file" button, the file is loaded (as a PySpark Dataframe) and become available for further processing/querying.

Dataframes are essentially data tables. At each moment, Query Builder has two Dataframes: the temporary dataframe (TempDF) which holds the result of the last performed action, i.e. the contents of a file or the result of applying a filter/query on the previous contents of TempDF and the master dataframe (MasterDF) which is explicitly updated by pressing a button that moves the TempDF contents to the MasterDF.
When a file is loaded, a form appears with several options for simple data filtering and processing. Each selected action will open a new form with the necessary parameters. Once the user fills the required information and presses OK a new filter will be added in the queue to be applied.
when the user wants. As an example, if the user chose the "Rename Column" filter he/she will then need to select which column to rename and what should be the new name.

![Query Builder - Rename column](image)

**Figure 4-12: Query Builder - Rename column**

Once the user presses “OK”, the filter will be added in the queue, under the "Selected Filters" label.

![Query Builder - Selected filters](image)

**Figure 4-13: Query Builder - Selected filters**
The user may add more than one filter before executing them. They will be added in the same list:

![Query Builder - Multiple filters](image1)

**Figure 4-14: Query Builder - Multiple filters**

When the user decides to apply them, he/she should click on the "Apply Filters" button. When the execution is finished, all successfully applied filters will change colour:

![Query Builder - Apply filters](image2)

**Figure 4-15: Query Builder - Apply filters**
The user may continue adding new filters after that point. The list will show which filters have been already applied and which are pending:

![Query Builder - Add more filters](image1.png)

**Figure 4-16: Query Builder - Add more filters**

By pressing the "Refresh temp" button, the user will get the first 40 lines of the TempDF. This is useful both when the user first loads a file but also to review the result of applying some filters.

![Query Builder - Review temp dataset](image2.png)

**Figure 4-17: Query Builder - Review temp dataset**
A "Refresh Master" button is also available to preview the contents of the master dataset. Keep in mind that the two dataframes may hold completely different data. The user may also join the two dataframes through the corresponding option in the filters form. The result will be kept in the TempDF.

![Image of Query Builder - Preview temp and master Dataset](image)

**Figure 4-18: Query Builder - Preview temp and master Dataset**

The user can use the "Update Query Output" button which will reveal a ready to copy and use python snippet which corresponds to the processes that have been applied on the TempDF.
Figure 4-19: Query Builder - Preview Query Output

By clicking the "Save master to CSV" button, the Save configuration form will appear where the user can select where to store the MasterDF's data.

Figure 4-20: Query Builder – Save dataframe to CSV
The user may load, process and combine data following the previous steps in any order, until he/she is satisfied with the end-result. At any time, the user may press the "Initialise QB" button to clear the screen and start over. When the user has completed his/her work with the Query Builder, he/she may either directly exit the Notebook or clear all the output first so that he/she can have a fresh start without any leftovers the next time. This can be achieved from Jupyter's top menu. Choose Cell -> All Output -> Clear.

**Visualiser**

Once the user clicks on the Visualiser tool, a preconfigured Jupyter notebook is opened as shown in the following figure.

![Visualiser initial state](image)

**Figure 4-21: Visualiser initial state**

The Visualiser is now in the initial state. All code cells are hidden by using the hide code plugin and also all output has been cleared. To start the tool, the user should simply execute the first paragraph by either selecting in and press ctrl+enter or by clicking the “Run” button on the top menu.

At that stage, the Visualizer is set and can be used, the first widget that appears looks like a file picker, where the user can navigate between the files stored in the current project and open a desired one. Note that only CSV files are supported.
Once a file is opened, a small preview in tabular format will be shown.

Depending on whether or not the selected file has been associated with extended metadata, a small section with a few recommended chart types will be displayed.

Afterwards, a list with the available visualisation types will be displayed. The user can pick a desired type and click on the Apply button.
For every visualization type, a widget with various options depending on the type will be opened. The user should fill the mandatory fields with the desired values and then click on the Visualize button.

The user should pay attention to the parameter called Position. This actually refers to the position on the dashboard where the chart will be placed. In the previous figure, there are five empty cells in the end of the notebook that act as placeholders for the dashboard. Currently only up to five different charts at the same time are supported, but this can be easily extended. The dashboard can contain several charts from the same dataset or even from different ones. When
a new dataset is opened, previous charts are not deleted. A sample chart that appears after filling the necessary options can be seen below.

Figure 4-26: Visualiser - Sample chart

At the bottom of every chart, the user is presented with a choice to save the displayed chart. The only thing that needs to be done is to provide a desired name and the press on the “Save Chart as HTML” button. This will store the chart as a dynamic HTML file inside HopsFS.

When the user has completed his/her work with the Visualiser, it is highly recommended to clear all the output, so he/she can do a fresh start without any leftovers the next time. This can be achieved from Jupyter's top menu. Choose Cell -> All Output -> Clear.

Figure 4-27: Visualiser - Clear output
Note: There are two alternatives ways for opening the Visualizer besides the one described earlier. The first one is accessible by visiting the Datasets Page and then open “Jupyter” dataset which is available by default in all projects. The user should browse for the Visualizer notebook, right click on it and then select the “Open Jupyter Notebook” context menu option.

![Figure 4-28: Visualiser - Open with Jupyter](image)

The second way, makes again use of the context menu, but in a completely different way. Again, the user should navigate to the Datasets Page and then open a dataset of her choice. Then, by right-clicking on a specific CSV file, she is given the ability to open that file directly into the Visualizer, as it can be seen in the screenshot above.

![Figure 4-29: Visualiser - Open with context menu](image)
The Algorithm Execution Container is implemented with Python 3.6 as a Jupyter notebook. Upon launch of the corresponding Algorithm Execution Container (hereinafter container), the user has to initialise it through a dedicated button. The initialisation ensures that the Spark interpreter of the AEGIS platform is started and also creates the basic UI of the container, which is a simple five-tab window:

- An input file selection tab
- An algorithm selection and configuration tab
- An output folder selection tab
- An overview tab showing the current user selections and, when ready, the execution results
- A model application tab, to perform regression or classification with existing trained models on other datasets.

Apart from the overview tab, the other three correspond to the basic steps towards applying an algorithm through the container. The first step is to select the file that contains the data to be used by the algorithm, through the interface shown in the next figure, which brings up also a preview of the selected file.

![Figure 4-30: Algorithm Execution Container - Select input](image)

The next step is to select the algorithm to be applied. The provided algorithms are grouped under five categories (algorithm families). Once an algorithm is selected, a form from which to configure its parameters is shown to the user. A basic form validation is done for the case that a selected value for a parameter is not within the specified boundaries. In this point, the final version of the Algorithm Execution Container includes the option to set a grid of parameters,
so that the algorithm may run within those spans, and it automatically selects the best model based on the results produced. A preview of the configuration of the grid parameters is shown in the next figure.

![Algorithm Execution Container – Algorithm configuration](image)

**Figure 4-31: Algorithm Execution Container – Algorithm configuration**

The model that will be created when an algorithm is applied is saved in the user’s datasets, in a folder specified in the last container tab, as shown below:

![Algorithm Execution Container - Output folder](image)

**Figure 4-32: Algorithm Execution Container - Output folder**

Once this last step is concluded, by pressing the “Apply” button, the user is taken to the Overview Tab:
There, by pressing Done, the algorithm is applied and when the execution is completed, some algorithm-dependent results are provided to the user in the same tab (next figure). The final output of the analysis is stored back in the AEGIS Data Store, while the model that was used for the analysis, is also stored alongside with the analysis results.

**Figure 4-33: Algorithm Execution Container - Overview**

<table>
<thead>
<tr>
<th>Input File</th>
<th>Algorithm Selection &amp; Configuration</th>
<th>Output File</th>
<th>Overview</th>
<th>Apply Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selected Input File:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Selected Algorithm:</strong></td>
<td>Logistic Regression</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Selected Output File:** | | | | | Dataset:/Projects/testing/testing_Training_Datasets  
Folder: Classification_Results |

---

**Results Summary**

- falsePositiveRate: 0  
- recall: 1  
- precision: 1  
- fMeasure: 1  
- Output: The created model has been saved in the selected output folder.  
- truePositiveRate: 1  
- accuracy: 1

---

**Figure 4-34: Algorithm Execution Container - Results**

In case the analysis that has been selected is of a classification or a regression, the model that has been created in the previous steps can be re-applied on new datasets, using the “Apply Model” tab, provided that the format of the input file is same as that of the one used to generate the model. This feature is shown in the next figure, where the user is able to select from existing files and re-apply the model.
Figure 4-35: Algorithm Execution Container - Select existing model

4.1.3. Extended Metadata

The AEGIS platform supports the provision and usage of extended metadata based on DCAT-AP and the AEGIS ontology. It is serialized as Linked Data and can retrieved via a SPARQL endpoint or via a RESTful-API. Each entity type (project, dataset and file) have their respective metadata model and forms.

The project metadata form can be accessed in Tools menu in right pane: “Extended Metadata”: 
Figure 4-36: Metadata Form for Projects

The metadata form for datasets is available with a right-click on any dataset: “Add or edit extended metadata”: 

![Metadata Form for Projects](image-url)
Figure 4-37: Metadata Form for Datasets
The metadata form for files is available with a right-click on any file: “Add or edit extended metadata”:

![Metadata Form for Files](image)

**Figure 4-38: Metadata Form for Files**

### 4.1.4. Jobs

The AEGIS platform provides support for data parallel processing frameworks such as MapReduce, Spark, and Flink through the use of the Jobs service. To run a job, the user needs to first upload his/her job files (.jar, .py, .ipynb) to a dataset. Then, the user can add a new job by clicking on the New job button as shown in the following figure.
To create a new job, the user provides the job name, type, configuration, and other external libraries as shown in the figure below.

Once the job is created, the user can run the job, and then look at the execution logs listed at the end of the page. Also, the user can rerun, edit, delete, and export the existing job or add new jobs.
4.1.5. Kafka

The AEGIS platform provides real time analytics using the Kafka service. The user can create topics and their respective schemas using the user interface provided by AEGIS, as shown in the following figure.

Figure 4-41: new Job is added

Figure 4-42: Kafka
After clicking on the new topic button, the user is required to fill the topic details such as topic name, number of partitions, number of replicas, and the topic schema, as shown in Figure 4-43. The user can select a schema from the existing schema in the platform or create his own schema first by navigating to the schemas tab, as shown in Figure 4-44. The user can create his/her own schema by clicking on the new avro schema button and then provide the schema’s details, as shown in Figure 4-45.

![Create new Kafka topic](imageurl)
4.1.6. Model Serving

The AEGIS platform uses Tensorflow for developing deep learning models. It provides a model serving service to deploy pre-trained model using the Tensorflow serving, as shown in Figure 4-46. The user can create a new serving by providing a model name, version, and a path, as shown in Figure 4-47.
4.1.7. Cleansing Tool

Data Cleanser is an offline tool that can help data owners to apply a set of validations, cleansing and data completion actions before they upload their data in the main platform. It is offered as a Docker image to allow the easy installation in every system.
In order for a user to get started with it, she must first clone the repository that is publicly available in GitHub. Inside the project’s folder there is a .env file which is used to define a series of environmental variables that are required for the application to run. The environmental variables can be set following the instructions below:

- **MONGO_INITDB_ROOT_USERNAME**: The username to access the MongoDB instance (MongoDB is used to store the rules configuration). The user should set it to value of his/her choice, or leave the default one.
- **MONGO_INITDB_ROOT_PASSWORD**: The password of the MongoDB instance.
- **MONGO_PORT**: The port at which MongoDB will be running on the host machine.
- **SECRET_KEY**: A secret to be used for securing Flask sessions. The user should set it to a long alphanumeric value.
- **JWT_SECRET_KEY**: A secret which is used to sign the JWT created by the application. The user should set it to a long alphanumeric value.
- **FLASK_ENV**: The environment under which Flask application will be running. Possible values are `dev`, `test` and `prod`. The latter is recommended.
- **APP_PORT**: The port under which the application will be running on the host machine.
- **MONGO_URI**: The full MongoDB URI which the application will use to connect to the MongoDB instance.

After environmental variable setup is complete, the user should execute the command `docker-compose -f docker-compose-prod.yml up --build -d`. This will take a couple of minutes. After that, the user should visit `http://[IP]:{PORT}` and the landing page will be displayed:

![Figure 4-48: Cleansing tool - Landing page](image)

In order to use the features of the cleansing tool, one must login first. The default username/password is admin/admin123.

![Figure 4-49: Cleansing tool - Login page](image)
Before starting setting up cleaning rules, the datasets of interest and their variables should be registered first. The tool assumes the following hierarchy. First, the user defines the providers. A provider or data owner is the name of the company/organization/individual who possesses the data. Each provider has a set of datasets and each dataset a set of variables. Note that it is not necessary to register all of the datasets/variables, but only the ones that you are interested to clean.

The tool offers an easy to use UI for creating the aforementioned structure.

![Figure 4-50: Cleansing tool - Set provider](image)

Next, the user creates the cleaning rules of interest.

![Figure 4-51: Cleansing tool - Rules creation](image)

Rules fall under three different categories:

- Validation Rules: They define constraints that should be checked (e.g. if values of a column are in a desired range)
• Cleaning Rules: They define actions that should be taken, in case of a specific validation rules is being violated (e.g. if values of a column are outside a desired range, replace those values with a predefined value).

• Missing Values Rules: Define actions that should be taken, in case where there are empty values in a column.
Having registered the necessary rules, the user can proceed to the cleaning process. The tool offers a simple UI for choosing the necessary provider and dataset and then upload a file to clean. The only supported file types are CSV and XLSX up to 500MB in size. For larger datasets, it is advised to use the tool's API instead.

Once the process completes, a new cleaned file will be returned. Additionally, the cleansing tool offers a feature where the user can check all the actions in detail that took place during the cleaning process. Actions are stored as log files.
Upon opening a specific log file, the user can get a detailed explanation of the actions that took place, as well as a nice dashboard which displays some interesting statistics.

**Figure 4-57: Cleansing tool – Log file view**

Finally, several API endpoints are being exposed from the tool. Those endpoints are documented using Swagger which can be accessed under:

http://{IP}:{PORT}/cleaner/api/v1/docs
4.1.8. Anonymisation Tool

The Anonymisation Tool is an extensible, schema-agnostic plugin that allows real-time efficient data anonymisation. The tool can used as an offline tool on the premises of the user and can configured to provide the anonymised output through a secure web API. The input of the Anonymisation Tool can be provided through a sync with a private dataset or through a file that can be imported to the tool.

In the first screen of the tool, the user is able setup or edit a saved configuration. The configuration is composed by a set of various database back-ends or text files and the corresponding anonymisation functions that will be applied on top of them.

When creating a new configuration, the tool will prompt the user to provide the connection details to the desired private database backend or select the local file to open. Additionally, the user is prompt to select the entities / tables to anonymise.
At the following step, the tool prompts the user to select: a) the fields from the data source that will be exposed to the anonymised set and b) the anonymisation function that will be performed.

The Anonymisation Tool offers a list of predefined anonymisation functions that can be used directly (e.g. city from an exact address, range of values from an integer), as well as a list of aggregation functions (e.g. average).
Furthermore, the tool offers the extension of the supported anonymisation functions with custom anonymisation functions. In this case, the user should provide the corresponding python module.

```python
def address_to_city_helper(address, info=None):
    if not info:
        info = get_address_info(address)
    if not info['results']:
        return '', info
    city_name = ''
    for component in info['results'][0]['address_components']:
        if 'administrative_area_level_1' in component['types']:
            city_name = component['long_name']
        elif 'administrative_area_level_5' in component['types']:
            city_name = component['long_name']
    if city_name:
        break
    return city_name, info
```

Additionally, the tool offers an integrate console to the user in order to perform any queries or test the output of the anonymisation process.
Moreover, the user can assess the risk levels of the selected anonymisation mapping via the Preview and Risk analysis, as depicted in the following figure.

**Figure 4-63: Anonymisation tool - Output**

**Figure 4-64: Anonymisation tool - Preview and Risk analysis**
As described above, the tool offers to option to expose the output data via a secure API to external parties. To enable this, the provisioning of access keys is mandatory.

![Access Keys](image)

**Figure 4-65: Anonymisation tool - Exposed API**

In this case, the user can access the produced anonymised data through the secure API in JSON format by utilising the private access keys configured above.

![JSON response](image)

**Figure 4-66: Anonymisation tool - API response**

### 4.2. AEGIS Platform adoption guidelines

The AEGIS platform has a lot of features to support rapid prototyping and development of data-driven applications for various stakeholders. In general, there are two types of potential users of the AEGIS platform, an experienced user with a data science background who can write and execute data management tasks using notebooks and code generation, or a non-technical user who prefers a simple user interface and uses the application developed by the data scientist on the AEGIS platform to just perform simple, prepared tasks such as loading prepared data into the Visualizer, creating a visualization and interacting with it to support his decision-making process.
This section briefly explains how data scientists can make the most of the potential of the AEGIS platform in own data science projects and how they can use the platform to develop an application meeting the users’ needs.

The first step in a data science project is to identify a concrete need of a target group that a data scientist wants to meet through the data-driven application to be developed. This is a complex challenge and data scientists must position themselves into the target audience, which requires also a lot of communication and requirements engineering skills. It is of utmost important that data scientist understand the domain addressed sufficiently. Moreover, it should be examined, if the data (made available by the target group or received by other means) is capable of meeting the needs of the target group. Several approaches can be chosen for this needs-analysis, e.g. interviews with stakeholders, workshops or focus groups. One practical idea is to document – together with representatives of the target group – the as-is situation and the challenge to solve, the to-be situation to be achieved, the technical solution approach to develop (the data-driven application) and the expected impact of using this application for the target group. Achieving a common understanding of the big picture very early is one success factor of data science projects.

In a second step, the data scientist must comprehensively analyse the needs of the target group to elicit concrete requirements for the data-driven application to develop. Therefore, it is important to determine the system boundary and define which aspects should be covered by the envisaged data-driven application, and which should not. Requirements can be documented in full sentences, however, a user story with some mock-ups of the application should be developed, too, describing the interaction with the application from the perspective of the user. There are several methods which can be applied published in the requirements management literature.

In a third step (which is of course also possible to be conducted at the beginning of the process, but probably the motivation of exploring the AEGIS platform is much higher having a concrete project to implement on the platform in mind), the data scientist must familiarise himself with the AEGIS platform and the available possibilities. The AEGIS documentation effectively guides the data scientist through the first steps of getting to know the platform and its modules. A good idea is to start with creating a demo project, uploading some (known) demo data, loading this demo data in a notebook, doing some simple computation and probably visualizing the result to better know how the platform works. The next step is then to examine the AEGIS modules, the Visualiser, the Query Builder, the Algorithm Execution Container, and the Data Harvester. Using these modules can save data scientists a lot of time. In addition, the AEGIS platform supports data-parallel processing services such as MapReduce, Spark and Flink as well as interactive analyses with notebooks such as Jupyter. Real-time analysis is enabled through the use of the integrated Kafka service. Deep learning is made possible by the use of the included Tensorflow service. Support for the above services allows data scientists to develop different types of applications according to their needs and then seamlessly implement and test them on AEGIS platform. But of course, the data scientist must get familiar with all these features first.

Once the data scientist is sufficiently familiar with the platform, he/she can design a concept for the data-driven application to develop and then carefully plan the steps for implementation. One way to be successful is to describe a set of scenarios with a set of test cases to be implemented on the AEGIS platform.
The next step is then to execute these defined steps on the platform. Once the user is satisfied and confident with the current plan of how to implement the application, he can begin the coding phase by developing the application using his framework of interest from the applications supported by AEGIS (MapReduce, Spark, Flink, Tensorflow and Kafka). It is always recommended to test the data-driven application/service developed closely together with the selected target group already at an early stage to determine whether the application and the way it should be used is suitable for solving the identified needs of the target group. There are several ways to get user feedback, such as thinking aloud, interviewing or focus groups.

Based on the received user feedback and ideas, the data scientist and the target audience can anticipate several phases of application evaluation and application iteration until the data-driven application meets the expectations of both the data scientist and the target audience. This is a usual practice in design-based projects for meeting the requirements of users in a best possible way.
5. CONCLUSION

The scope of deliverable D5.6 as the final deliverable of WP5 was to document the efforts undertaken within the context of the tasks 5.3, 5.4, 5.5, and 5.6 and provide a final evaluation, impact assessment and adoption guideline. It was built on top of the work and outcomes of the deliverables D5.1, D5.2, D5.3, D5.4, and D5.5 towards the aim of providing a summarized assessment of the AEGIS platform and demonstrators, lessons learnt made during demonstrator implementation and finally a how to use the AEGIS platform in own data science projects.

At first, the deliverable provided a summarized result of the AEGIS demonstrators final evaluation, covering both the quantitative and the qualitative evaluation as well as an overview of each implemented demonstrator along with the activities performed.

Following the AEGIS demonstrators’ final evaluation, the deliverable presented the AEGIS platform final evaluation, which was conducted quantitatively and qualitatively, too. Additionally, the document presented the results of a security assessment as the consortium decided to perform a security assessment of the AEGIS platform in order to ensure that its security mechanisms are in are operating efficiently and effectively.

Finally, the deliverable presented the AEGIS platform documentation and adoption guidelines featuring a platform usage documentation, as well as the platform adoption guidelines as a how-to best exploit the potential of the AEGIS platform in own data science projects.

The outcomes and knowledge extracted from this deliverable will serve as valuable feedback and information for the exploitation phase, following the AEGIS project.