HORIZON 2020 - ICT-14-2016-1

AEGIS

Advanced Big Data Value Chains for Public Safety and Personal Security

WP1 - AEGIS Data Value Chain Definition and Project Methodology

D1.3 – Final AEGIS Methodology

Version 1.0

Due date: 31.03.2018 **Delivery Date**: 12.04.2018

Author(s): Spiros Mouzakitis, Evmorfia Biliri, John Tsapelas (NTUA), Cinzia Rubattino, Elisa Rossi (GFT), Alessandro Testa (HDI), Alexander Stocker (VIF), Gianluigi Viscusi (EPFL), Dimosthenis Tsagkrasoulis (HYP), Dustin Stadtkewitz, Yury Glikman (Fraunhofer), Marina Da Bormida (Ethical Advisory Board), Sotiris Koussouris, Marios Phinikettos, Spyridon Kousouris (SUITE5), Konstantinos Perakis (UBITECH), Alessandro Testa(HDI)

Editor: Spiros Mouzakitis (NTUA)
Lead Beneficiary of Deliverable: NTUA

Dissemination level: Public **Nature of the Deliverable:** Report

Internal Reviewers: Yury Glikman (Fraunhofer), Dimitrios Miltiadou (UBITECH)

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
HYPERTECH	HYPERTECH (CHAIPERTEK) ANONYMOS VIOMICHANIKI EMPORIKI ETAIREIA PLIROFORIKIS KAI NEON TECHNOLOGION
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.

EXECUTIVE SUMMARY

The document at hand, entitled "Final AEGIS Methodology", constitutes a report of the performed work and the produced results of all WP1 tasks. The scope of the current report, which concludes the WP1 activities, can be described in the following axes:

- An updated analysis of the stakeholders' needs is provided, mainly based on the results of a survey conducted by the AEGIS consortium. The survey was the second survey conducted to elicit requirements from people involved in big data enabled services in the Public Safety and Personal Security domains and as such it was more targeted to specific actors, specifically managers, IT technical operators and data analysts. It was aimed to understand on one hand, the needs of the potential AEGIS stakeholders, in order to develop a Big Data analytics platform of real added value, on the other hand how the AEGIS platform impacts the market considering all the steps of the AEGIS Big Data Value Chain, spanning from data collection to data and service sharing in real scenarios.
- The AEGIS Big Data Value Chain, initially described in D1.1, is refined. The steps remain the same, but their scope is adjusted to match the stakeholder requirements and the consortium's perception on how each of the entailed data handling processes should be addressed in the context of the project. In order to ensure that all important data characteristics have been taken into consideration when updating the data value chain, data sources relevant to public safety and personal security are grouped based on the stakeholder they are of interest to and the technical challenges imposed by them are highlighted.
- The final version of the AEGIS methodology towards data-driven innovation in the domains of Public Safety and Personal Security is defined, describing the user interactions and workflows to be supported by the AEGIS system. Two examples are provided to concretely showcase how the methodology is instantiated to support various types of users in leveraging the available big data analysis functionalities. Additionally, a discussion on the AEGIS Minimum Viable Product (MVP) to be supported by the final methodology is provided, based also on insights from the stakeholders' perspective and from the technical progress of the project. The integrated AEGIS methodology and the MVP insights will be leveraged as input for the tasks of the project regarding implementation and exploitation of the solution.
- The project's strategy towards ethical and data privacy and IPR considerations is defined in detail, taking into consideration all data privacy requirements and existing regulatory instruments, concluding with the definition of the AEGIS Ethical, Privacy and Data Protection Strategy.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AAL	Active and Assisted Living
AEGIS	Advanced Big Data Value Chains for Public Safety and Personal Security
API	Application Programming Interface
CEO	Chief Executive Officer
COPD	Chronic Obstructive Pulmonary Disease
D	Deliverable
DoA	Description of Actions
EAB	Ethics Advisory Board
HVAC	Heating, Ventilation and Air Conditioning
IAQ	Indoor Air Quality
ICT	Information and Communication Technology
IP	Intellectual Property
IPR	Intellectual Property Rights
IT	Information Technology
MVP	Minimum Viable Product
OBD	On Board Diagnostics
PAYG	Pay As You Go
PSPS	Public Safety and Personal Security
REST	REpresentational State Transfer
SME	Small and Medium-sized Enterprise
T	Task
WP	Work Package

1. INTRODUCTION

1.1. Objectives of the deliverable

This deliverable is related to the activities performed during the second iteration of all WP1 tasks (T1.1-T1.5) and concludes all activities performed in this work package. Its main objectives are to:

- Update the previously provided analysis of the stakeholders' needs by conducting a new stakeholder survey. This survey will elicit requirements from people involved in big data enabled services in the Public Safety and Personal Security domains, targeting specific actors that have been identified as more relevant to the project's scope, specifically managers, IT technical operators and data analysts. It will help to understand on one hand, the needs of the potential AEGIS stakeholders, in order to develop a Big Data analytics platform of real added value, on the other hand how the AEGIS platform impacts the market considering all the steps of the AEGIS Big Data Value Chain, spanning from data collection to data and service sharing in real scenarios.
- Refine the AEGIS Big Data Value Chain and describe how each step is supported inside the project, making sure that the expected data sources have been identified and their characteristics and potential technical challenges have been foreseen and addressed.
- Discuss the collected feedback from the initially described high-level AEGIS usage scenarios and combine this input with the updated stakeholder requirements and the insights from the technical evolution of the project towards defining the final AEGIS methodology for data-driven innovation in the PSPS domains. The final methodology definition will encompass all workflows to be supported for all users of the project's platform, both when working independently but also for collaborating teams. The envisioned AEGIS Minimum Viable Product (MVP) will be also updated.
- Update the work performed in the first iteration towards identifying, monitoring and analysing legal and regulatory legislation relevant to AEGIS innovations and implementation. This work includes the updated identification of relevant national, regional, legislation and regulatory instruments, the extraction of new data protection requirements, where necessary, and the final definition of the AEGIS Data Policy Framework

1.2. Insights from other tasks and deliverables

The current deliverable builds directly on top of the previous two WP1 deliverables, namely D1.1 "Domain Landscape Review and Data Value Chain Definition" and D1.2 "The AEGIS Methodology and High Level Usage Scenarios". Specifically, the current deliverable (1) updates the stakeholder analysis and needs identification first presented in D1.1, (2) refines the AEGIS Big Data Value Chain which was defined also in D1.1, (3) updates and finalises the integrated project methodology first defined in D1.2 and (4) updates the work related to AEGIS Ethical, Privacy and Data Protection Strategy, the first version of which was also provided in D1.2.

It should be noted that all AEGIS partners are actively collaborating in the WP1 activities and therefore progress and insights from them are also taken into consideration during work performed in the current work package. As such, the current deliverable also builds on top of the feedback collected from the technical development of the platform (mainly WP3 but also WP4), the pilot scenarios described in D5.2 and also the activities performed in WP7 which are closely linked to the AEGIS MVP discussed in the current document.

1.3. Structure of the deliverable

The deliverable is organised in six main sections. Apart from the first introductory section and the last section that presents the deliverable conclusions, each of the other sections corresponds to a different objective from the ones described above. More specifically, section 2 presents the survey results and the updated stakeholder requirements. Section 3 provides some insights on the data sources related to PSPS and also presents the updated AEGIS big data value chain. Section 4 describes the updated project methodology and provides two instantiation examples to facilitate its understanding. Finally, section 5 describes the AEGIS ethical, privacy, data protection and IPR strategy.

2. UPDATED STAKEHOLDER ANALYSIS AND NEEDS IDENTIFICATION

2.1. AEGIS Value Chain and Stakeholders Analysis

The overall objective of this section is to update the review presented in D1.1 of the application of Big Data technology in the target sectors for the AEGIS platform.

In D1.1 we identified 11 stakeholder groups as shown in the following table from D1.1.

Table 1: Stakeholder groups overview

Stakeholder group	Types
SG1 - Smart Insurance	Insurance Companies Financial institutions Insurance brokers
SG2 - Smart Home	Electronics Smart home technology providers Safety and security Energy and Utilities
SG3 - Smart Automotive	Car manufacturer Car dealers Electronics GPS Navigation System Providers
SG4 - Health	Nursing homes Hospitals Doctors
SG5 - Public Safety / Law Enforcement	Police Emergency Medical Service Fire Service Search and Rescue Military
SG6 - Research Communities	Students Professors Research institutes
SG7 - Road Construction Companies	
SG8 - Public Sector	Municipalities Public Authorities
SG9 - IT Industry	IT software companies Data scientists Data Industries

SG10 - Smart City	Electronics Smart City technology providers Smart City planners
SG11 - End Users	Citizens

Smart Insurance

The Insurance sector is one of the identified stakeholder groups and is represented in the AEGIS Consortium by the Italian Insurance Company HDI Assicurazioni.

In recent years, data and analytics have become essential tools for insurers in designing more sophisticated approaches across all aspects of their operations. With an incredible amount of data flowing in from multiple new digital channels, the insurance industry is undergoing a paradigm shift in the way they function – right from product planning to pricing, introduction, marketing, customer self-service and claim processing.

In D1.1 the main benefits of Big Data Analysis towards Insurers were identified as: Fraud Detection and Prevention, Smart Finance, Customer Loyalty and Retention, Telematics, Reputation and Brand Analysis, Claims Management and Social Network Analysis. The impact of AEGIS will be evaluated through some of these features (e.g. Smart Finance, Customer Loyalty and Retention, Claims Management) on our demonstrator.

Furthermore, it is possible to recognise other interesting application of Big Data Analytics to the Smart Insurance, these applications are described hereinafter.

Risk based pricing and support – Risk analysis to help proactively monitor risks to minimise customer losses. This kind of analytics applied to real-time streaming data could be useful in particular for Life and Vehicles policies, allowing the insurer to make real-time decisions that manage risk. In addition, predictive analysis could be applied towards this end: it helps in offering the right premium for the right risk, helping retain profitable customers.

Marketing and Sales – Through Big Data Analytics it is possible to gain complex information, as an advanced analysis from different data sources and different type of data (e.g. structured and unstructured). The outputs of the analysis lead to an accurate, site-based strategy that involves both of the marketing side (targeted advertising campaigns) and the sales strategy (pricing adjustment).

Personalised products and offerings – Predictive, analytical models applied to heterogeneous data sources allow the identification of cross-selling opportunities for a wide-range of products by better anticipating customer needs and interests while allowing to take corrective actions to run targeted loyalty programs and predict financial commissions for an insurance major. Analytics could be applied also to understand the customer risk profile and offer tailored products, while social media data provide deeper insights about customers leading to better lifestyle analysis, personalizing products and pricing models all of which lead to business growth.

Agent management – Big Data analytics could be useful in order to optimise the Agent distribution and workload on the territory, to evaluate their performance and also to analyse the agency sales performance and drive revenue growth.

Business plan – From a finance perspective, by combining past losses with predictive analytics on future losses, the ability to manage loss reserves is also greatly enhanced. Funds can be more efficiently allocated, improving margins through improved cash management.

Smart Home

The Smart Home stakeholders group is comprised of interacting parties related to automated and optimized control of the in-home energy systems (e.g. HVAC, lighting), utilizing correlated information from sensory data, such as occupancy patterns and weather signals. The detailed description of the particular group can be found in D1.1. Here we concentrate on few updates produced within the course of the project.

In addition to what was reported in the previous deliverable, it was identified that a Big Data Platform can particularly contribute to the pre-processing, normalization and transformation of raw smart home data coming from the monitoring sensors, so as to alleviate the need to perform these actions within dedicated hardware (e.g. smart home gateways), which commonly have limited computational power and are not tailored towards intensive data analysis. A demonstrative example can be drawn with respect to occupancy data and the workflow required to extract such information from PIR sensors. Motion data generated from such sensors require a relatively simple but high-frequency processing flow in order to be transformed to occupancy events. This frequent processing can become a bottleneck in low-performance devices where, in addition, energy autonomy is critical. As such, transposition of such functionality to the platform, where further correlation of events may generate increased value in terms of insights over the aggregated datasets, can be of particular help to smart home service providers.

Smart Automotive

The automotive industry – and especially original equipment manufacturers (OEMs) will face a radical transformation in their business, while a series of new players mainly from the information and communication technology industry are entering the market. Two concrete challenges related to the digitalisation of the automotive industry are the transformation of current vehicles into automated vehicles, making extensive use of machine learning and artificial intelligence technologies, as well as the emergence of new businesses resulting from the provision of third party services built on top of vehicle data (as well as other vehicle-related data).

Data-driven services and connected car platforms will therefore play a substantial role in future automotive ecosystems for different stakeholders. In recent years, the ongoing digitalization of the automotive industry has already emerged a number of new players. In analogy to the quantified-self-movement, the IT industry in the USA has evolved a number of quantified car start-ups, which are backed by enormous amounts of risk capital, reaching far more than 20 million USD in some cases. These developments demonstrate that investors perceive a high market value of digital services based on vehicle data.

Linking to the challenge of vehicle development and production, a further impact of digitalization in the automotive industry is the shift from traditional business models (e.g. vehicle as a product) to new, data-driven business models (e.g. transportation as a service, digital services based on vehicle operation data), which is a strong driver for innovation and automotive market re-organization.

The following stakeholder-groups are especially relevant for the automotive demonstrator:

Transferring the quantified-self phenomenon to the vehicle domain, **vehicle drivers** are directly targeted as potential beneficiaries by the project. Vehicle drivers are the most important target group for a quantified vehicle ecosystem, as they - as the owners of the data their vehicle generates – may share their driving data for in return receiving a benefit through services created based on their data. Hence, vehicle drivers who have a major interest in assessing e.g. their (personal) driving style and benchmarking their driving style with those of their peers through advanced driving analytics features will benefit greatly. Furthermore vehicle drivers may be provided with a more detailed analysis on how they drive in the field (as current vehicles are capable of providing to them), receiving a visualisation of safe as well as of unsafe driving events which were detected in their driving data on a geographic map. As there are many external factors affecting driving style and driving behaviour, taking external data sources (e.g. weather data) into account will generate additional impact on this analysis. Finally, drivers who are interested in learning how to drive more economically friendly are also directly targeted by the project results.

Automotive data scientists (stemming from automotive research communities) are another direct valuable stakeholder-group for the project results. They are provided with both the capability for vehicle data analysis through the platform created as well as with data from real vehicles operated in the field, which is available on the platform created and can be used by them for their own research. They can finally use the platform to test and implement algorithms for e.g. driving behaviour and driving style analysis as well as for detecting different events in data streams (e.g. road conditions, driving risks, driving distraction). The platform allows them not only to explore the data, but also to create and evaluate their own algorithms on real data.

Another target groups are **road/city planners** and **road maintenance teams**, who directly benefit through being provided with an overview on critical areas within their regions with regard to driving practices and road conditions. Both may use the reports created for them on the platform to identify driving safety-related areas within their cities and road networks (e.g. areas where harsh driving is a dominating driving style) and then act accordingly by setting infrastructural measures (e.g. limit the speed in this area).

Health

In recent years, due to the ongoing technological advancements, vast amounts of health-related data have been accumulated. Their characteristics, as has been described previously are variety, volume, structure, annotation, as well as level of data privacy. The goal within the project is to address a number of the aforementioned issues, especially with respect to patient behaviour and clinical data. In particular, it is actively investigated how expert rules, coming from medical

experts can be incorporated to a data analysis framework towards the development of a personalized notification system.

Public Safety/Law Enforcement

Big Data can provide additional insights to governments and businesses (intelligence agencies) to help them keep cities safe, and better respond to disasters when they strike. Public safety is a wide challenge that involves many fields, from the crimes to roads and bridges maintenance. Big Data analytics has the potential to make intelligence more efficient, by detecting patterns in huge volumes of data, revealing connections between documents, people, recurring events and the circumstances that lead to events that threaten security, safety, and property.

Research centers/communities

With the advent of data science (i.e. data-driven science) the demand for algorithms, methods and processes to extract value from (research data) has increased significantly and research communities are eager to contribute to data-driven challenges (e.g. to data hackathons), too. Due to initiatives including open science or science 2.0 to name two of them the availability of research data to a wider public is also slowly but steadily increasing. This is aimed to provide data to all scientists who are not willing – or not able – to collect data on their own.

While in the meantime the EU has also spawned a plethora of platforms to open and share research data (e.g. Zenodo.org), the availability of platforms for working with the data according to the Big Data Value Chain (i.e. transforming the data, analyzing it, extracting interesting events, and visualizing them) is still very limited. Hence many researchers have to develop proprietary, hard-coded solutions to finally showcase their research results via the Web.

Due to the ongoing digitalization, the demand of applied researchers coping with data-driven industrial challenges has increased dramatically. This holds for many industrial domains and especially for industry 4.0 (the continuing digitalization of manufacturing) and automotive (the digitalization continuing of transport). The demand for platforms to showcase how to extract value from manufacturing and supply chain data (cf. industrial data space) or transportation-related data (c.f. vehicle data markets) is steadily growing.

Road Construction Companies

The road network is a crucial part of the transportation system and maintaining its functionality and road safety is of a high importance. Especially after the winter season, broken roads have to be repaired. Potholes and cracks have to be handled. This currently involves a series of periodic manual inspecting actions (sometimes also in response to reports from the public) to identify these hotspots of reduced safety and functionality in the road networks. Some rely on special vehicles for road monitoring to identify them.

Automatic road damage detection is therefore a hot topic, as it could reduce or limit manual inspection works. Automatic road damage detection approaches can for instance be based on high resolution (satellite) images or video data. Furthermore, road maintenance companies are interested in being provided with an overview on critical areas within their road network with

regard to traffic patterns and road conditions to even predict the emergence of critical situation within their road network.

Public Sector

As was stated in the previous deliverable, a Big Data platform can contribute to four major areas related to public sector productivity and organization. In particular: advanced analytics, through automated algorithms; improvements in effectiveness, providing greater internal transparency; improvements in efficiency, where better services can be provided based on the personalization of services; and learning from the performance of such services. Furthermore, through the provision of Big Data services and associated applications related to security and safety, we can positively affect further the public sector's effectiveness and yield.

IT Industry

The IT Industry is one of the most technological for its nature, with appropriate investments in most of cases for both of the hardware and the software to manage Big Data and analytics. Their employees are skilled or at least with a background that allows them to quickly respond to new market needs.

Smart Cities

The notion of smart cities encompasses and extends that of smart homes. In specific, it is reasonable to consider as an extension to the self-learning home automation framework, a more general service that takes into account not only the in-house presence, but also data coming from our outdoor activities and city conditions. This holistic approach can further increase the public safety and civil security, through medical and social support, as well as complementary services, such as traffic and transportation advice, and risk alerts.

2.2. Questionnaires

While the AEGIS project is moving to its second part, a second version of the survey was jointly developed by the partners of the Consortium, to collect further suggestions from the potential stakeholders, evaluating their concerns and expectations related to Big Data.

The first survey was sent at the beginning of the project (M3) in order to define the preliminary user requirements and information sources of the stakeholders that are potentially interested in AEGIS Data Value Chain. The analysis of the first survey was discussed in D1.1 and it is available also as a blog post at https://www.aegis-bigdata.eu/what-is-the-current-and-expected-use-of-big-data-technologies-a-glimpse-to-our-aegis-questionnaire-results/ .

The survey is aimed to understand on the one hand, the needs of the potential AEGIS stakeholders, in order to develop a Big Data analytics platform that could be a real added value for them, on the other hand how the AEGIS platform impacts to the market considering all the steps of the AEGIS Big Data Value Chain, from the collection to the sharing in real scenarios.

The survey was meant to involve three types of participants (Manager, IT Technical Operator and Data Scientist), with different points of view, knowledge and needs.

For the second iteration, it has been decided to design a questionnaire which is offered to all different target groups (11 stakeholder groups) pointed out in D1.1 (section 3.1) and reported in section 2.1 of the present deliverable, targeted on the role of the participant in his/her organisation. The roles identified are:

- Manager = a person responsible for controlling or administering an organization or group of staff, he/she has a high-level point of view about Big Data analytics but is the person that could benefit from them. He/she has a focus on business intelligence.
- IT Technical Operator = person responsible for the management of the data storage, curation and collection, he/she knows which the critical points of these tasks could be.
- Data Scientist = person that extracts information from data, using Big Data analytic tools, for instance following the instructions of the manager. He/she has the proper skills for data analysis and could identify the deficiencies of the existent tools.

We tried to collect general information on the responder's organisation, then depending on the role of the participant, his/her point of view/approach regarding the various steps of the AEGIS Value Chain. The following Table (Table 2) shows for each role the main topics/features investigated.

Table 2: Overview of the main topics investigated through the survey for each of the defined roles

Role	Main Features
Manager	Effort and resources involved in data analytics Type of data involved in the analysis Data treatment agreement Sharing of the analysis
IT Technical Operator	Data collection Data sources Hardware Sharing of the analysis
Data Scientist	Type of data involved in the analysis Data sources Analytic tools and algorithms Sharing of the analysis

Once agreed upon the questions and the structure of the questionnaire, an online version (powered by Easy Feedback) has been provided and it is still available at the following link: https://indivsurvey.com/aegis/117873/8il3tU. The survey can be found in Appendix A.

Email invitations among the audience of stakeholders collected within the partners' direct links and contacts have been sent directly from each partner.

The results from the questionnaire are now being examined trying to understand the needs of each actor involved in the process and looking for a solution that optimises them.

2.3. Questionnaire results and reflections

The received valuable replies to the questionnaire were 33. The respondents covered almost all of the target groups of AEGIS project, even if the major part (almost 50%) of them is coming from IT industry and this is probably due to the motivation to participate to these types of studies (Figure 2-1). All the participants' sectors with less than four answers were joined as 'Other'. Most of the participants belong to a private company (Figure 2-2), while there is a regular distribution of respondents in SMEs and large entities. The geographical distribution of the participants is highly related to the partner distribution: the main part of the answers comes from Austria, Greece and Italy.

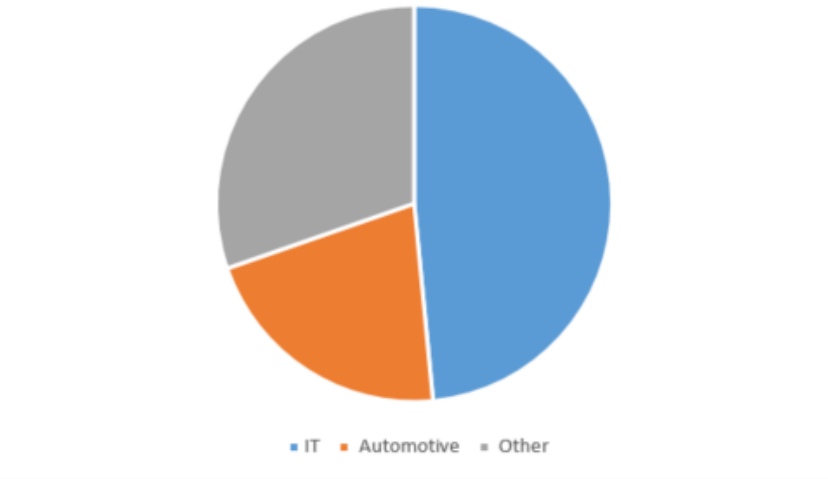

Figure 2-1: Chart of the survey's participants organisation sector

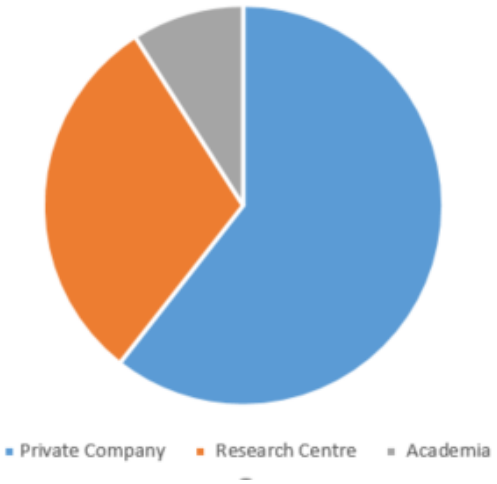

Figure 2-2: Chart of the survey participant's type of organisation

According to the table below, almost 70% of the organisations represented are effectively using Big Data, a number which shows that the AEGIS community comprises, to a large extent, high tech companies. This result is also in agreement with the main purpose of this second iteration of the questionnaire, i.e. to target the participants on 'Big Data experts'.

Table 3: Level of Experience of the organisation with Big Data

Effectively using Big Data	67%
Beginning in the use of Big Data	18%
Planning to use Big Data	12%
No experience	3%

As described in section 2.2 the survey was intended to be filled by three different types on participants; the collected answers come from Managers (48.5%), Data Handling Operators/Data Scientists (39.4%) and IT Technical Operators/Computer Systems Analysts (12.1%).

Hereinafter the answers of each role will be individually examined, while in the next section (section 2.4) a combined analysis will be provided.

Manager

The responses from the Managers were 16; three of them belong to organisations that are planning to use Big Data: they do not have a designed team (internal or external) to perform data analysis. Among the others, considering who is beginning to use Big Data, they prefer to perform analysis through external consultants or an internal team, but not as main activity, while the organisations that are effectively using Big Data have an internal team of Data Scientists.

The people working on data analysis in the organisations in most of cases are from 2 to 7 out of 12 answers (Figure 2-3); this is not related to the dimension of the organisation itself. Moreover, it is important to point out that the 50% of the participants declared that even if they have a dedicated budget for Big Data and Analytics, the investment is not adequate, while only one out of 12 considered the investment appropriate; the 25% plans to dedicate a budget on Big Data and related Analytics. The 58.3% of the participants' organisations has the proper hardware to manage Big Data, the 25% has not and the 16.7% doesn't know. These last two groups match with who considered the investment on Big Data and Analytics as not adequate or with who is planning to invest on Big Data and Analytics.

Figure 2-3: How many people work on Data Analysis in your organisation?

The added values of Big Data analysis and the main issues related to Big Data handling from the Managers point of view are reported in the following tables.

Table 4: Which are from your point of view the added values of Big Data Analysis?

Predictive analysis	75%
Cross-domain analysis	68.7%
Real time analysis	37.5%
Fast decision making	37.5%
Improvement of the offered services	25%
Cost reduction	18.7%
Better customer service	18.7%
Competitive advantages over rivals	18.7%
More effective marketing	12.5%

Provide insights for research	6.2%

Table 5: Which are the main issues related to Big Data handling in your organisation?

Difficulty of handling Big Data	50%
High management cost	31.2%
Difficulty of finding trained staff in Data Analysis	25%
Legislation about privacy and security	25%
Lack of performance of the available tools	12.5%
Heterogeneity of available data	6.2%
Lack of confidence in the real benefit	6.2%

Considering the AEGIS Big Data Value Chain, one of the questions regarded which of its steps are implemented in the organisation: 11 participants out of 12 carry out Data Acquisition and Analysis, 8 out of 12 really use the results of the analysis, 8 out of 12 have a Data Storage while only 5 out of 12 have a step of Data Curation. The respondents that are 'effectively using Big Data' have already been implementing all the steps of the AEGIS Big Data Value Chain.

The origin of the data involved in the analysis is reported in the following table.

Table 6: Which are the data involved in the analysis of your organisation?

External, customers (e.g. data from social media, sensors)	66.6%
External, open data	50%
Internal of the organisation	33.3%
External, real-time data	33.3%
Internal, related to customers	25%
Purchased data	16.6%

Regarding the contract between the organisation and the data providers, the participants that declared the use of external or purchased data have an agreement that includes a reference to

further processing of previously collected personal data. Furthermore, all the participants agreed on the importance of interlinking datasets from different domains/data sources for the analysis: the 66.6% uses to perform that kind of analysis, while the 33.3% finds it useful but it does not yet perform it.

Only 9% of the respondents uses alerts, warnings or monitoring systems based on Big Data Analytics to support after an event, 9% does not know, and 81.8% does not use that kind of automated feedback.

Finally, considering the use of the reports of an analysis, the sharing is mainly with the customers (63.3%) and with colleagues of the same team (54.4%).

Table 7: The analyses are shared

With customers	63.3%
With colleagues of the same office/department/team	54.4%
With colleagues of other offices of the same organisation	45.4%
As open data	18.2%
I don't share analysis	9%
With external, entities	-

IT Technical Operator

The replies from IT Technical Operators were four; two of them are part of an organisation that is effectively using Big Data, while the other two belong to an organisation that is planning/beginning to use them. The number of answers is not enough meaningful to allow effective analysis, and could not lead to the identification of habits/issue/expectations of IT Technical Operators.

Nevertheless, both of the participants that come from an organisation that is effectively using Big Data have an internal team working on Big Data Analytics with more of seven people involved on it. They both identified as issue related to Big Data handling, the legislation about privacy and security.

Data Scientist

The Data Scientists that participated in the questionnaire are 13. The organisations of the participants are effectively using Big Data, and the number of people employed in Data Analysis are mainly related to the dimension of the organisation itself. Seven respondents out of 13 (53.8%) declared that there are different restrictions about data visibility in their organisation,

while the 15.3% declared that there are not restrictions about data visibility (the 30.7% does not know).

The added values of Big Data analysis from the Data Scientists point of view are reported in the following table.

Table 8: Which are from your point of view the added values of Big Data Analysis?

Predictive analysis	69.2%
Cross-domain analysis	61.5%
Improvement of the offered services	46.1%
More effective marketing	46.1%
Cost reduction	38.4%
Fast decision making	38.4%
Better customer service	30.7%
Real time analysis	15.3%
Competitive advantages over rivals	15.3%

Considering the AEGIS Big Data Value Chain, the steps coverage is shown in the following table. Only two out of 13 of the participants organisations have already been implemented all the steps of the AEGIS Big Data Value Chain.

Table 9: Which of these steps have already been implemented in your organisation?

Data analysis	84.6%
Data usage	53.8%
Data acquisition	46.1%
Data storage	46.1%
Data curation	23%

Such results lead to consider that the data used for the analysis are not directly coming from a proper acquisition/storage process, but this statement is inconsistent with the other answers of the same participants. First, considering the results of the answer '*Which are the data involved in the analysis of your organisation?*', the data processed come mainly from customers (e.g. data from social media or sensors) – 61.5%, or from internal data related to customer (e.g. contracts)

– 53.8%. The other categories (open data – 38.4%, data internal of the organisation – 23%, real-time data – 15.3% and purchased data – 7.6%) scored percentages considerably lower. Second, all of the participants asserted that they use to acquire data when needed, and through scheduled streaming – it is not possible to extract meaningful data about the frequency of these streaming. These two questions bring on the possibility that the lack of the step of 'Data acquisition' is not properly true, for instance the participants considered that they do not have a structured step of data acquisition, but they acquire data only when needed.

The types of data mainly used for the analysis are logs and sensors data, while the data types not yet exploited but that the participants would like to use are: geospatial data, phone usage, emails, transactions, social media, audio, RFID scans or POS data, Earth observation and space. In general there are more types of data 'I would like to use' than the 'I use', while 11 participants out of 13 indicated more than one data type as the ones used, in agreement with the importance of the analysis between datasets from different domains/data sources (9 respondents perform that type of analysis while 3 of them consider it as important).

The 53.8% of the participants answered that has the proper analytic tools related to his/her needs; the most popular tools used are R, Matlab, Python, other tools mentioned are Pandas, MS Excel, Spark and SAS Base. The algorithms adopted/that the respondents would like to adopt for the analysis are reported in the following table, while the outputs of the analysis are mainly in tabular format (75%).

Table 10: Which are the algorithms involved/would you like to involve in your analysis?

Algorithm	I use	I would like to use
Linear regression	61.5%	30.77%
Predictive analysis	46.1%	30.77%
Clustering algorithms	46.1%	38.46%
Simulations	46.1%	30.77%
Estimation of correlation between variables	38.4%	30.77%

Only 5 out of 13 of the participants declared that they have scheduled automated analysis, 6 out of 13 declared that they don't have scheduled automated analysis and 3 out of 13 don't know, but only 2 out of 13 said that in their organisation there are alerts, warnings or monitoring systems based on Big Data Analytics.

Finally, considering the use of the reports of an analysis, the sharing is mainly with colleagues of the same team (76.9%) and with the customers (53.8%).

Table 11: The analyses are shared

With colleagues of the same office/department/team	76.9%
With customers	53.8
With colleagues of other offices of the same organisation	38.4%
With external, entities	30.7
I don't share analysis	15.4%
As open data	7.7%

The last question for each participant was the same, aimed to understand if the AEGIS platform is an interesting tool; the main AEGIS features/functionalities were listed and the respondents could assign a value corresponding to their level of interest about the functionalities, varying from 'Not at all' to 'Very'. The following tables show the results for each role.

Manager

Table 12: Would you been interested on a tool...

Feature/Functionality	Level of interest			
	Not at all	Slightly	Moderately	Very
Online and free	1	2	2	**8**
Where you can buy and sell assets	1	1	4	**5**
With a set of open assets			4	**7**
Where you can connect in-house streaming datasets	1	2	5	**6**
Where you can store your analysis and assets		2	5	**7**

Where you can manage the metadata related to your data		2	4	**8**
Where you can query your datasets and access a set of related visualisations	1	3	2	**8**
Where you can set and save the steps of your analysis	1	1	4	**8**
Where you can share the information with a selected group of users	1	4	2	**6**
Where you can set different restrictions about data visibility	1	2	4	**7**

IT Technical Operator

Table 13: Would you been interested on a tool...

Feature/Functionality	Level of interest			
	Not at all	Slightly	Moderately	Very
Online and free		1	**2**	1
Where you can buy and sell assets	1	**2**	1	
With a set of open assets		**3**	1	
Where you can connect in-house streaming datasets		1	**3**	
Where you can store your analysis and assets		1	**3**	

Where you can manage the metadata related to your data		1	**3**	
Where you can query your datasets and access a set of related visualisations		**2**		**2**
Where you can set and save the steps of your analysis		**2**	1	1
Where you can share the information with a selected group of users		1	**3**	
Where you can set different restrictions about data visibility		**2**	1	1

Data Scientist

Table 14: Would you been interested on a tool...

Feature/Functionality	Level of interest			
	Not at all	Slightly	Moderately	Very
Online and free	3		2	**6**
Where you can buy and sell assets	3	**5**	2	1
With a set of open assets	3		**7**	1
Where you can connect in-house streaming datasets	3	1	**4**	2
Where you can store your analysis and assets	2	2	**5**	1

Where you can manage the metadata related to your data	2		**7**	3
Where you can query your datasets and access a set of related visualisations	1	1	**6**	3
Where you can set and save the steps of your analysis	2	1	4	4
Where you can share the information with a selected group of users	2	1	**6**	2
Where you can set different restrictions about data visibility	2	3	**5**	1

2.4. Updated Needs Identification

The second iteration of the survey is aimed to understand the actual usage of Big Data and related Analytics, with a particular focus on all of the steps of the AEGIS Big Data Value Chain. The questionnaire was designed keeping in mind the features and the functionalities of the AEGIS platform that has been built following the requirements pointed out from the first survey iteration (ref. D1.1) where there were analysed the potential stakeholders' point of view, and from the user stories, further elaborated as user requirements (ref. D3.1) where there were analysed the requests of the three demonstrators of the project.

It is important to evidence that the first survey was developed and circulated at the early beginning of the AEGIS project (the results were delivered at M3) when the AEGIS platform has not yet a well-defined shape as now. Hence, compared to the first questionnaire (ref. D1.1) the questions of the second are more specific, for that reason it has been decided to target the survey on people with a good knowledge/experience with Big Data and Analytics. For that reason, most of the participants of the first iteration were not involved in the second iteration, as well as most of the questions have been changed. The following figure, comparing the extension of the experience with Big Data in the participants' organisations between the two iterations, highlights this aspect.

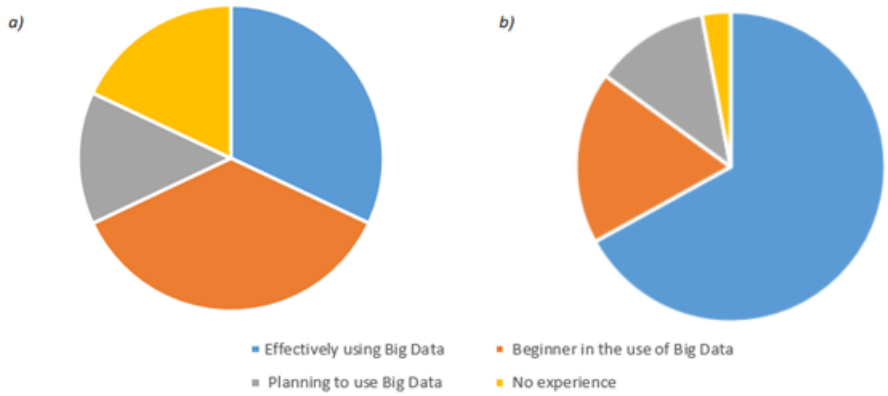

Figure 2-4: To what extent does your organisation have experience with Big Data? a) Results of the first survey iteration; b) results of the second survey iteration

Table 4 and Table 8 show the benefits of Big Data Analysis from, respectively, the Managers and the Data Scientists points of view. In general, the scores are similar, but despite the expectations, the proposed answers that could be associated with a business view (i.e. cost reduction, more effective marketing and improvement of the offered services) scored a higher percentage among the Data Scientists' answers than among the Managers' answers (38.4% vs. 20%, 46.1% vs. 13.3% and 46.1% vs. 26.6%). This discrepancy could be due to a more trust on the worth of Big Data Analytics from those daily experiencing them than the final 'beneficiaries' as it is possible to consider the Managers.

Considering the data sources used for the analysis, during both of the iterations there were identified the same: logs, sensors and events. On the other hand, the data not yet exploited but identified as most promising are very different: even if logs data, sensor data and events data were identified as the most exploited, they were identified also as the data desired in 5 years (see Table 9, from D1.1). The present questionnaire otherwise revealed a great interest for the future on geospatial, audio, earth observation and space, RFID and POS, phone usage, social media, emails and transactions data. It is interesting to point out that while in the first iteration the role of the respondent was not investigated, in the second the respondent of the same answer were all Data Scientists (Table 16).

Table 15: Summary of the most relevant data types. Percentage of participants collecting and analysing them (from D1.1)

Data type	Collected	Analysed	Forecast (5 years)
Log	67%	50%	83%
Social media, open data PSI, events, sensors, transactions, external feeds	40-60%	10-25%	75-80%

Free-form text, geospatial, images/video	25%	10%	50-60%

Table 16: Summary of the most relevant data types. Percentage of participants using and would like to use them

Data type	I use	I would like to use
Log, sensors, events	40-60%	40-55%
Open data/public sector information, external feeds, free-form text, emails	10-25%	65-75%
Transactions, social media, phone usage, reports to authorities, Earth observation and space, audio	10%	85-90%

Due to the role-based survey developed, the results of the question "*Does your organisation have the right analytical tools to handle Big Data*" are fairly different: in the first iteration only the 24.3% of the participants declared to have the proper tools, while in the second the 53.8%. The Big Data analytic tools identified in the first survey were Hadoop (21%) and Microsoft power BI (17%). In the second questionnaire, each of the respondents (Data Scientists) has indicated more than one tool, the most popular are: Python and R (50%), Pandas and Matlab (33%).

The following table is the summary of the answers of the last question of the survey for each role, "*Would you been interested on a tool*", where the main features of the AEGIS platform were listed.

Table 17: Would you been interested on a tool...

Feature/Functionality	Level of interest			
	Not at all	Slightly	Moderately	Very
Online and free	4	3	6	**15**
Where you can buy and sell assets	5	**8**	7	6

With a set of open assets	3	3	**12**	8
Where you can connect in-house streaming datasets	4	4	**12**	8
Where you can store your analysis and assets	2	5	**13**	8
Where you can manage the metadata related to your data	2	3	**14**	11
Where you can query your datasets and access a set of related visualisations	2	6	8	**13**
Where you can set and save the steps of your analysis	3	4	9	**13**
Where you can share the information with a selected group of users	3	6	**11**	8
Where you can set different restrictions about data visibility	3	7	**10**	9

The features that mainly attracted the respondents were the possibility to have a tool where you can manage metadata, where you can set and save the steps of your analysis and the fact that the tool is online and free. A good attention would be also for a tool with a set of open assets, where it is possible to connect in-house streaming datasets, where it is possible to store the analysis performed and assets, and that allow querying datasets and accessing a set of related visualisations.

The functionality that scored the lowest level of interest is the possibility to buy and sell assets. The lack of interest could be explained with the low percentages of purchased data (12%) and with the tendency to share the analysis within the organisation or at least with customers (Table 7, Table 11 Table 8) from both of the Manager' and the Data Scientist' points of view. This match in turn with the data privacy issues evidenced in Table 5.

The following table highlights the main outcomes of the survey's analysis.

Table 18: Main survey outcomes

Role	Main outcomes	Business requirements
Manager	The investments on Big Data and related analytics are not adequate.	A targeted dissemination strategy: the AEGIS platform is online and free, the investment for the organisations to use it will not be high.
	The step of Data Curation is not handled within most of the organisations.	Emphasise and develop a strong and reliable Data Curation step, enrichment of data with metadata, interlinking of datasets (planned).
	The reports are in most of cases shared with colleagues.	Allow sharing features (planned – ref. TR58, D3.1).
IT Technical Operator	N/A	N/A
Data Scientist	Importance of different restrictions about data visibility.	Definition of different levels of visibility and details of datasets and analysis' results (planned – ref. FR_RT5, FR_DS6, D3.1).
	The step of Data Curation is not handled within most of the organisations.	Emphasise and develop a strong and reliable Data Curation step, enrichment of data with metadata, interlinking of datasets. (planned).
	Need to exploit a wide range of data types that have not already used.	Make available on the platform a variety of trusted data sources (both free and available for purchase), spanning through the interests of all of the stakeholder groups identified (planned).
	The reports are in most of cases shared with colleagues.	Allow sharing features (planned – ref. TR58, D3.1).

3. AEGIS DATA VALUE CHAIN

3.1. Insights on PSPS data sources

AEGIS touches upon big data analysis in a variety of domains inside the spectrum of public safety and personal security (PSPS). Due to this wide scope, an exhaustive list of potentially relevant data sources cannot be drafted, especially in today's constantly evolving data landscape. D1.1 provided a long list of identified data sources that could be of interest to one or more of the 11 AEGIS stakeholder groups (described in Section 2.1 of the current deliverable) with concrete dataset examples and also performed an annotation of the datasets according to certain important attributes, e.g. data format, real time vs historic etc.

It should be stressed that activities reported in D1.1 were carried out during the first three months of the project and therefore aimed to explore the landscape in order to fuel subsequent tasks with the necessary insights to guide the project's decision making. Given the current progressed state of the project, the aim is not to extend an already extensive list but to group the available information and gained experience in a more useful way that will accelerate the identification of new, previously unseen data sources and provide insights regarding the expected difficulties in handling them, which will further serve as requirements for the project's methodology on big data analysis. In this scope, the types of data sources presented here are not, as in D1.1, organised under the stakeholder group that generates/ provides them, but under the stakeholder group that may benefit from their analysis according to the AEGIS progress so far. Furthermore, data sources of video and image content are not examined, as AEGIS has decided to focus on text data, a decision, which is fully aligned with the elicited stakeholder requirements.

Table 19: Overview of PSPS data sources & relevant insights

Stakeholder Group	Main data sources of interest (in the scope of AEGIS)	Data characteristics & Challenges
SG1 – Smart Insurance	• Customer records • Digital documentation of claims • Various data that can be used as reports on current situation, including weather and events reported by authorities or discussed/implied in public online conversations	• Data variety and variance • Presence of free text & accuracy of analyses derived from natural language processing at scale • Strong data privacy requirements & need to combine data from various sources/authorities to produce valuable insights • Multilingual information & lack of common formats that could facilitate automation of cross-lingual analysis • Guaranteed anonymisation of data to be shared among different organisations

SG2 - Smart home	• Ambient sensors and other smart home appliances that constantly report measurements • Data from wearable devices & smartphone sensors • Health records • Various data that can be used as reports on current situation, including weather and events reported by authorities or discussed/implied in public online conversations • User actions on smart devices (e.g. change of settings)	• Real time data streaming & event detection • Time series data, often in different ranges • Storage of large volumes of data • Combination of diverse data sources (e.g. time series with natural language) • Powerful visual analytics required • Lack of one dominant commonly used standard • Sensitive data • Presence of free text & accuracy of analyses derived from natural language processing at scale
SG3 - Smart Automotive	• Vehicle usage data (on board diagnostics, rotation sensors etc.) • Historic & real time traffic data • Maps • Historic accident data • Weather • Various data that can be used to mine references to traffic related events (congestion, accident…), ranging from official reports from authorities to social media and e-newspapers	• Time series analysis • Large volumes of data • Potentially sensitive data • Real time data streaming and event detection • Combination of diverse data sources (e.g. time series with natural language) • Availability of historic data: low quality when open data, difficult to obtain proprietary data • Visual analytics for interactive maps • Presence of free text & accuracy of analyses derived from natural language processing at scale
SG4 - Health	• Health records, medical results, clinical trials data • Data from activity tracking wearable devices • Medical IoT • Open data and public reports from national authorities (including WHO, Eurostat, OECD) • Ambient sensors • Weather	• Sensitive data • Low quality of open data • Lack of commonly used file templates and reporting structures • Data quality validation cannot be automated for open data – a per dataset approach may be required

SG5 - Public safety / law enforcement	• Open data, both historic and streaming • Public reports from national authorities and public event databases • Various data that can be used as reports on current situation, including weather and events reported by authorities or discussed/implied in public online conversations	• Lack of interoperability among systems from different authorities in many countries • Lack of common templates/terminology in the generated csv files • Data quality validation cannot be automated – a per dataset approach may be required
SG6 - Research communities	Very broad stakeholder group, hence it is potentially relevant to and can benefit from almost every data source reported in the table.	
SG7 – Road Construction companies	• Road condition data • Road maintenance data • Weather data • Historic traffic data	• Data availability • Data usually not produced to be used in this context, hence maybe are missing information and also need to be cleanses and/or pre-processed
SG8 – Public Sector	• Databases and other historic data produced and kept by public authorities • Potentially all data sources of the other stakeholder groups, as all analyses related to PSPS domains are inherently of public interest	• Lack of interoperability among systems from different authorities in many countries • Lack of common templates/terminology in the generated csv files • Data quality validation cannot be automated – a per dataset approach may be required
SG9 - IT Industry	Stakeholders in the IT domain act as service providers and analysis enablers on data related to and identified by stakeholders of the other groups	
SG10 - Smart City	• smart grid sensors & meters • telemetry devices • smart light sensors	• Collection and storage of large volumes of data • Combined analysis of multiple time series possibly in different ranges • Identification of the appropriate data subset to include in an analysis
SG11 – End Users	Very broad stakeholder group, hence it is potentially relevant to and can benefit from almost every data source reported in the table.	

More details on the exact data sources that were used to extract the insights presented in the previous table can be found in Tables 14-27 of D1.1, in the AEGIS pilot scenarios reported in D5.2 and in Section 5.4.2 of the current deliverable.

3.2. Updated data value chain definition

AEGIS adopts the Big Data Value Chain defined by Curry, E. (2016), which comprises five main steps, as follows:

Figure 3-1: Big Data Value Chain

The scope of each step is refined, when needed, to match the project's context and is adapted to the requirements of the stakeholders, as shown by the survey results in Section 2, and as reported by the project's pilots in D5.2. Furthermore, the challenges posed by the data sources presented in Section 3.1 serve as direct input for the definition of the AEGIS data value chain, presented below. The technical requirements identified in the project so far (reported in D3.1) are also considered in order to ensure consistency and balance among the ambitious theory-driven design, the outlined real-life stakeholder needs and the evolution of ICT solutions in data analysis.
For each of the steps, an informative overview is provided together with some considerations, along the following 3 axes, when relevant:
Big Data Vs: All steps in the data value chain are, more or less, affected by the four Vs of Big Data, i.e. the four dimensions initially identified to turn data into big data. The well-known four Vs, in brief, are:

a. Volume, which refers to the scale of data, i.e. having massive amounts of data.
b. Variety, which refers to having different forms of data that need to be analysed in common.
c. Veracity, which is used to highlight the lack of trust and confidence to the data at hand, i.e. refers to the inherent uncertainty (inaccuracies, poor quality etc.) of massive data.
d. Velocity which refers to the speed at which data are being generated or updated and mainly manifests in streaming data.

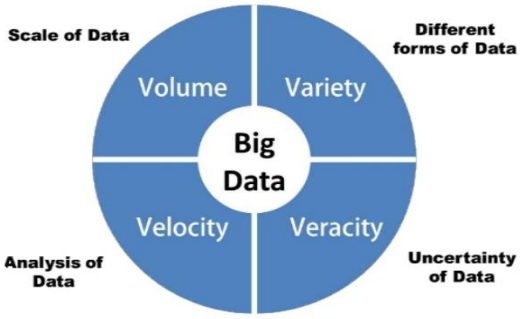

Figure 3-2: The 4 Vs of Big Data

Involved AEGIS users: The stakeholder groups presented and described in Section 2.1 refer to the domains involved in public safety and personal security. However, people active in each domain have different jobs, technical and theoretical background and goals to achieve through the AEGIS solution. Understanding these, as well as other behavioural aspects related to their expectations when performing certain data manipulation tasks, may highlight important differentiating factors that affect the way each step should be supported.
Cross-sectorial data handling considerations: AEGIS spans across multiple domains in the PSPS spectrum. Although data integration is a common task in data analysis, additional complications may need to be considered when combining data from multiple domains with different terminology and semantics.

1. **Data Acquisition** is defined as "the process of gathering, filtering and cleaning data, before any data analysis can be carried out".
   As shown in the previous section, AEGIS aspires to help users analyse and extract meaning from a wide range of data types, including both structured and unstructured data, data from sensors and sensor networks, various streaming data (possibly shared through different protocols) and mined events. Hence, AEGIS will support both real time data acquisition and acquisition of historic non-dynamic data. Data acquisition could also be seen under the legal framework prism, in order to ensure that proper data access control is applied and data privacy and security rules are in place. It should be noted that prior to data upload, for some data sources (e.g. the sensors of the smart home demonstrator, the in-house datasets of the insurance demonstrator) it may be necessary to anonymise sensible data. The relevant technical aspects are examined in the corresponding technical deliverables, whereas the legal and ethics aspects are discussed in Section 5 of the current deliverable.
   Considering the 4Vs of a Big Data Analytic platform (Figure 3-2), data acquisition is a key step for three of the dimensions: big data acquisition entails high speed input/collection of voluminous data in various formats.
   Although many end users can act as data providers and relevant processes and interfaces are foreseen, data acquisition in the current context refers to the backend functionalities of technical components that enable the connection to data sources and the data retrieval and gathering. Hence the active users in this step are IT people involved in the development and administration activities of the system (e.g. developers of the platform and system administrators).

More details on how AEGIS realises this step of the data value chain, e.g. for aspects related to avoiding latency and enabling connectivity to IoT-enabled devices, can be found in the technical deliverables.

2. **Data Analysis** is "concerned with making the raw data acquired amenable to use in decision-making as well as domain-specific usage".
   This step is considered as one of the core steps of the overall data value chain and is able to showcase the actual value of the overall flow, as it goes one step further than simple data sharing and offers knowledge sharing. During this step, data that have been already collected pass through analyses execution containers, and results are generated, allowing data analysts as well as ordinary users to understand the meaning of data sets and the importance of the data that reside in them. However, although the outputs of this step can be offered to different stakeholders depending on the output format chosen (e.g. scientific datasets, visualisation, simple reports, etc.), this step is mostly performed by data analysts that have the necessary background knowledge to choose, design, configure and interpret the results of various algorithms. In AEGIS, data analysts are grouped under two main categories: the ones that are familiar with coding and perform their analysis through a programming language of their choice and the ones more familiar with UI-based tools that can be used to apply various data manipulation methods on the desired data. AEGIS supports both these user types, referred to as coding and non-coding data analysts respectively.
   One of the main tasks in data analysis is correlation mining, i.e. the discovery of dependency patterns, among specific data inputs. In order to gain new insights and true value from the identified correlations, it is important to allow for unforeseen data combinations and means of evaluation, taking in mind that some datasets are difficult for the human mind to interpret, e.g. sensor data or data produced from a first level of analysis that strips them from their human-friendly form (e.g. through transforming natural language text to vectors). As such, during this step, combinations of datasets are brought in for analysis, supporting in this manner a real cross-sector and cross-domain analysis, attributing to a real data value chain and to knowledge the quality of which is highly dependent on the combinations made possible during this step. Data Analysis depends heavily on the V's of Big Data, as for different Vs (such as volume or velocity), different infrastructures and algorithms are necessary to be employed to be able to process the data and perform the desired analysis. Variety of data is also very common, especially due to the cross-domain data combinations in PSPS sectors. As AEGIS addresses the needs of the non-coding analysts as well, semantically-enhanced data manipulation services need to be offered to facilitate users through extracting meaning from combining and cross-examining data of multiple formats.
   It should be stressed that, although PSPS applications require high accuracy levels, there are inherent data features that render the required analysis not only more labour-intensive, but also error prone. Indicatively, in many NLP tasks 80% correctness counts as good quality, whereas in real life applications the propagation of such large errors across the value chain would be disastrous.
   In the scope of AEGIS, Data Analysis involves two discrete, yet interconnected steps, which have to do with the actual analysis that is performed on data that is stored in the AEGIS repository, and the visualisation of certain results that derive out of the analysis. In this context, the first part refers to the utilisation of different data analytics algorithms that reside

in certain analytics and machine learning libraries, and the second to the presentation of results in order to make them more comprehensive to interested stakeholders.
As a conclusion, it becomes evident that the criteria used for the analysis of big data cannot and should not be known a-priori, but only in analysis time, in order to ensure that the extracted value is not limited by early erroneous decisions (according to the principle of late interpretation). Hence, explorative analysis is at the core of the data analysis step. Exploratory analysis builds on the fact that when analysis starts, the questions to be answered are not (always) known. Questions only emerge a-posteriori together with the extracted answers, which is the case in many of the AEGIS envisioned applications and services.
The provision of exploratory analysis capabilities inside the wide field of PSPS is extremely challenging and guides the way the next steps (inside Data Curation) of the value chain are designed.

3. **Data Curation** is "the active management of data over its life cycle to ensure it meets the necessary data quality requirements for its effective usage".
   Data curation is an umbrella term for various processes regarding data organization, validation, quality evaluation, and provenance and multiple-purpose annotation. In the scope of AEGIS, the following data curation services are designed and evaluated along the following axes:
   a. **The definition and measurement of data quality**. Data quality affects the complete value chain since it compromises the value of the final output, regardless of the adopted data processing practices. Data provenance is also considered a significant factor here, especially since AEGIS aspires to build a trustful collaborative environment with clear data exchange and processing mechanisms.
   b. **The need to employ traceable and repeatable curation processes**. This is linked to the volatile nature of big data, which requires existing data curation steps to be verifiable against new versions of data and render the detection of new steps possible. These requirements imply that data curation must be scriptable, but at the same time cannot be fully automated.
   c. **The need to avoid irreversible data restructuring**. This is a requirement of the previously explained need to enable exploratory analysis, which by definition forbids the application of lossy data transformations and compressions, since these may impede future analyses.
   d. **The inherent limitations of applying data curation at scale**. These limitations do not refer to technical challenges, but to the fact that the correctness of the output of such a process can only be validated through evaluating statistical properties of the result and/or through visual analytics. It is therefore crucial to devote the necessary time and be aware of possible errors that should be handled before affecting the final analysis results.
   e. **The benefits of using semantics**. Although time consuming, especially during the first design phases, agreeing on clear semantics and applying appropriate annotations can pave the way for truly smart big data enabled services and assist users (mostly data analysts) in exploring and extracting meaning from data during the steps of data analysis, curation and usage.

   In the context of the project, data analysis and curation are tightly linked processes and are thought to be performed both by data analysts, with the coding group more likely to be involved in the current step. The term data curator is also used to denote the users that are

mostly active in this specific step of the data value chain. Inside AEGIS, data curators are seen as data analysts with more targeted scope of work.
All 4 Big Data Vs are in principle relevant here, although velocity is not expected to affect the methods and services used to perform data curation. Finally, when it comes to cross-domain data curation, AEGIS adopts an approach fuelled by strong semantics expressed in metadata that allow a level of automation and/or the generation of suggestions to the user.

4. **Data Storage** is "the persistence and management of data in a scalable way that satisfies the needs of applications".
   Data storage is a wide area and is extremely important in Big Data ecosystems, since it deals with issues ranging from scalability and performance to data consistency and availability, to data models and security and many others.
   It is obvious that storage is related to all 4 Vs, however the scope of the step is related to the technical architecture and is therefore addressed in the technical deliverables. Regarding the storage of cross-domain data, no additional challenges are posed, if variety and lack of veracity have been foreseen and addressed from a technical perspective.

5. **Data Usage** refers to the "data-driven" business activities that need access to data, its analysis, and the tools needed to integrate the data analysis within the business activity".
   Through the previous value chain steps that enable big data crunching and analytics, information-rich and reduced-volume data are exposed to organizations in a way that makes them useful for value creation, and thus data usage represents the final step in deriving value from data. While the exact manner of data usage is each time inherently associated to the specific business objective at hand, a number of basic principles can be extracted, and constitute the backbone of the data usage value step, adopted in the AEGIS platform. In particular, processed data should comprise the pillar for decision support, in-use analytics, visualization and exploration, which can be integrated into application offerings, as well as revenue analysis.
   As this step encompasses the results of all other steps and therefore the complete chain of actions leading to it, all AEGIS users are expected to be active here. Due to the wide range of possible data usages, the cross-domain requirements and the challenges imposed by the big data Vs are examined and addressed on an application basis. At the very least, interactive visual exploration is always foreseen and supported within AEGIS.

4. AEGIS METHODOLOGY AND MVP DEFINITION – FINAL

4.1. Reflections on initial high-level usage scenarios

D1.2 presented five high-level scenarios to showcase how AEGIS aspired to enable stakeholders in the PSPS domains create added value services through big data analysis. The scenarios were collaboratively produced by the consortium members and were used as input for the initial design steps of the AEGIS solution, since they provided concrete examples of the stakeholders' expectations and hints on required technical components. They were also the foundations on which the first version of the AEGIS methodology was defined. The final AEGIS methodology encompasses insights gained through the activities performed in all project tasks so far, which include updates in the big data analysis state of the art theories and tools, feedback from domain stakeholders internal and external to the project's consortium and also insights from the technical implementation of the AEGIS platform.

An additional valuable input in refining the methodology naturally comes from re-visiting those initial five scenarios to identify their weak points and address them in the final methodology. The main extracted insights from feedback on the scenarios are as follows:

1. Big data analysis is becoming increasingly relevant to various stakeholders in the PSPS domains, therefore a diverse audience may be interested in utilising the AEGIS offerings. It became evident that assumptions regarding the expected technical knowledge of the users should be clearly defined and evaluated in order to ensure that appropriate interfaces are provided for the various processes involved in all big data value chain steps, at least for the users whose needs AEGIS primarily aims to address.
2. The concept of a big data-enabled service is very broad and concrete examples should be provided to help users understand what can be implemented through AEGIS, what can be seen as an exploitable asset, what can be shared externally to AEGIS and what can only be seen as a value-added service tightly connected to and provided through the platform.
3. The provision of datasets that can be directly used in analysis is not extensively discussed, therefore it is unclear whether there is any additional support in exploring and processing them or whether a catalogue of open data is envisioned. The expectations from the semantic annotations should be described, as well as the way the users will be able to contribute with their own data in this repository/ marketplace.

4.2. Updated integrated methodology

AEGIS addresses the needs and requirements of a diverse audience involved in big data analysis in the PSPS domains. This diversity manifests itself in terms of

(a) the role of a user in the core AEGIS platform, which is mainly related to the user's technical background. From this perspective, AEGIS addresses four main roles: the business user (e.g. a manager or in general a user who will not perform any development/ data analysis work, but consume the results of an analysis), the developer and the data analyst who, as also stated in Section 3, may belong to the coding or non-coding group. The distinction here serves to understand the different ways in which members of the two groups may approach the same

underlying data analysis workflow, hence the different ways that the AEGIS system should support them. Hereinafter the two groups, i.e. the coding and non-coding data analysts, will be addressed as different AEGIS users in order to better explain the AEGIS dual approach in handling their needs.

(b) the role of the user that is directly stemming from the motive for using AEGIS in a given workflow. From this perspective, AEGIS addresses the roles of data providers, data curators, data consumers, service providers and service consumers.

Other roles related to the technical requirements of the AEGIS platform, e.g. the system administrator, are not examined here.

The first category of roles is inherently more stable: although a user's technical background may progress, it is not expected to change frequently. On the other hand, as AEGIS offers a range of functionalities that cover the complete data value chain, users are expected to "traverse" the available workflows holding different roles from time to time. As such, a data analyst may use the platform to create a visualisation for an analysis to be provided to the decision-maker of the company as supporting information or may choose to leverage the platform's data processing and filtering functionalities in order to provide a more high-quality dataset to other users etc. Importantly, user roles in the second category may not correspond to individual users, but to teams of users, possibly working in the same company/organisation. As an indicative example, in the case of a service provider it may be a user with business background that sets the goal and the requirements for a data-enabled service that needs to be created (to be provided to a customer or kept for internal usage), a data analyst selects and combes the appropriate datasets and provides the algorithm to be applied and a developer that handles the core coding part of the solution.

The realisation of this dual diversity is among the main conclusions from the feedback on the initial methodology definition and as such, the final AEGIS methodology foresees not only various alternative workflows but also alternatives in the way a specific workflow is realised, depending on the background and motivation of the current user. Figure 4-1 provides a high-level overview of the final integrated AEGIS methodology, which is in fact a synthesis of all the envisioned workflows.

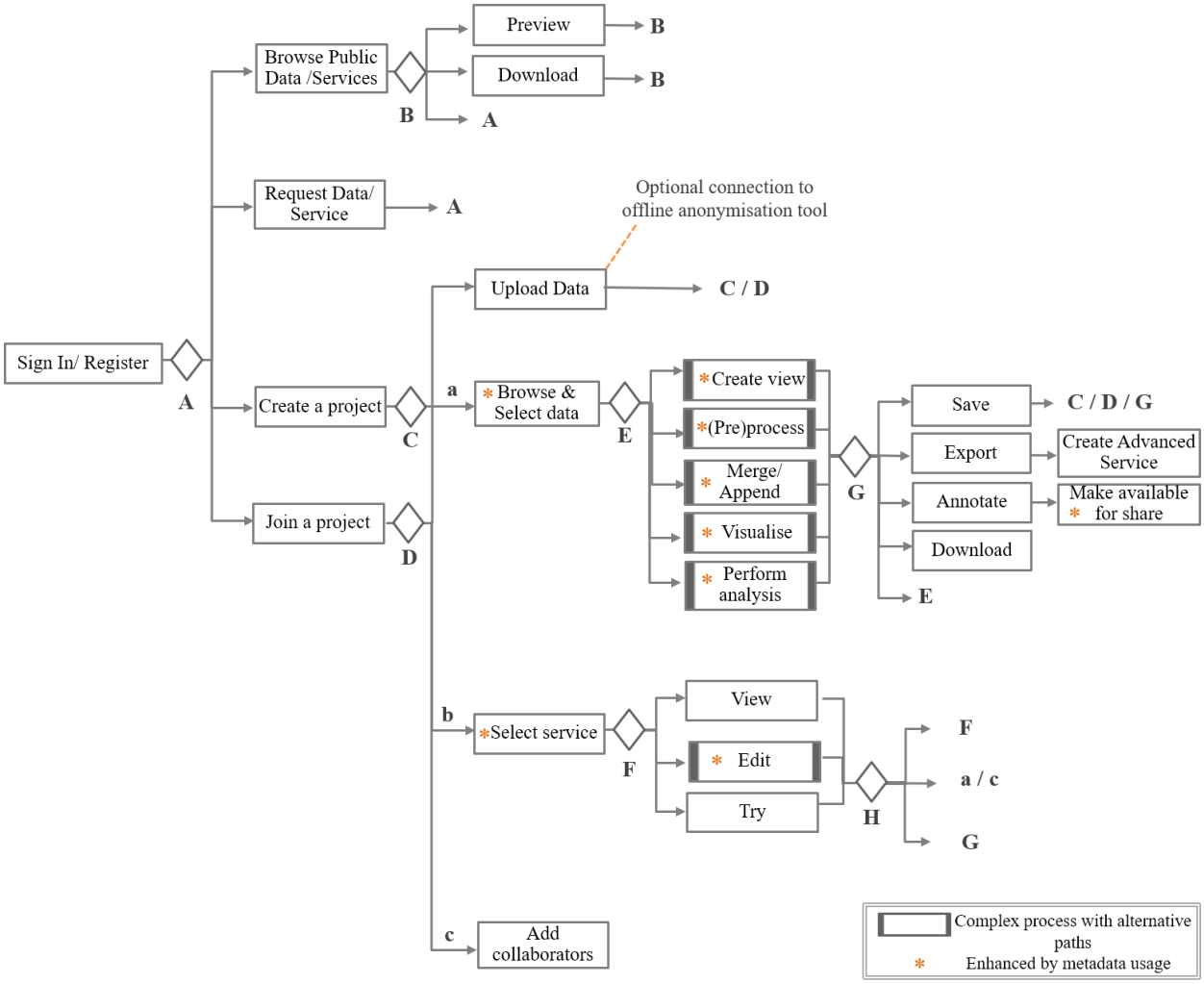

Figure 4-1: Final AEGIS methodology - overview

As AEGIS aims to facilitate big data analysis through iterative exploration and experimentation of data and data-enabled services, the scope of its functionalities is too broad to allow for an exhaustive depiction of all possible workflows. At the same time, it was decided that the true contribution of the current methodology in realising the AEGIS vision should be to highlight the common steps needed to accomplish possibly very diverse big data analysis tasks by very diverse users.

As a means of validation of its correctness and completeness, two core workflows are selected to showcase how the methodology is instantiated. The two workflows are described in detail in Sections 4.2.1 and 4.2.2.

Prior to that, each of the high-level methodological steps will be briefly explained for reasons of completeness.

Table 20: High-level methodology steps explanation

Step	Brief overview
Sign in/ Register	AEGIS is offered to non-anonymous users in order to facilitate the management of resources and because it aspires to be used for its collaborative big data analysis functionalities and not just as a catalogue for its free/public datasets and services. Therefore, all workflows refer to registered users.
Browse Public Data/Services	Some of the AEGIS assets are available to all registered users, e.g. datasets from open data portals, visualisations on them and code snippets for indicative analysis tasks. These assets could be also previewed and/or downloaded. For further experimentation with them or for enhanced exploration, the user is required to proceed with creating a project.
Download	Public datasets are directly available for download. Assets within private projects can also be downloaded if the user has the appropriate permissions.
Request data/service	As AEGIS aspires to serve also as a marketplace for PSPS data and data-enabled services, it may be the case that a user is interested in posting a request for an asset not currently available in the system. This will also promote the creation of a community inside AEGIS, especially in the case of requests for services, which could foster new collaborations and attract data analysts.
Create a project, Join a project	Projects are the main spaces for data collection, data analysis and collaboration. It should be noted that after a request for data/service (step described above) is addressed, the user making the request will be able to access it through a new project shared with him/her by the person/organisation that responded. Each project conceptually encloses all activities related to all steps of a specific analysis and helps clarify the user roles and data permissions in its context.
Upload data	Uploading here is a simplified term to describe the process of making data available in AEGIS. For static data files, this indeed refers to a simple data uploading, whereas for example in the case of scheduled bulk uploads of large volumes of data it will be a more complex process.
Browse & Select data	Selection of datasets may be performed not only through traditional keyword search and browsing, but also through a smart metadata-enabled semantic search and data exploration.
Create view (on dataset)	This step mainly entails the application of filters that allow the user to create the desired subset (view) on a given dataset or on the combination of multiple datasets. If semantic information is available for the specific dataset(s), the process is simplified through the automated enabling/disabling of certain

	options. The step follows a dual approach to support both coding and non-coding users.
(Pre)process (dataset)	This step includes a wide variety of data processing tasks and follows a dual approach to support both coding and non-coding users. If semantic information is available for the specific dataset(s), the process is simplified through the automated enabling/disabling of certain options.
Merge/Append (datasets)	Being able to combine and integrate various and diverse datasets is central to the AEGIS data value chain. The step follows a dual approach to support both coding and non-coding users. If semantic information is available for the specific datasets, the process is simplified through the automated enabling/disabling of certain options.
Visualise (dataset)	A variety of visualisations, most of which are configurable and/or interactive, supports the user through the visual exploration of the raw or processed data. The step follows a dual approach to support both coding and non-coding users. If semantic information is available for the specific dataset(s), the process is simplified through the automated enabling/disabling of certain options.
Apply analyses	The step entails selection, configuration and application of algorithms from a rich predefined list, but also the ability to implement advanced custom analyses. It follows a dual approach to support both coding and non-coding users. If semantic information is available for the specific dataset(s), the process is simplified through the automated enabling/disabling of certain options.
Select service	AEGIS adopts a broad definition of what may constitute a service, in order to address the needs of its diverse users. Indicative service examples are visualisations, code snippets that create a visualisation or run a data processing/analysis task, prefilled reports that showcase how a hypothesis can be proven etc. Public services are also searchable and selectable here with the addition of the semantically-enhanced functionalities which are not available externally to the projects.
View/ Try a service	Where applicable, a user may view a service and how it is created and, in some cases, may directly experiment with it. In the case of a code snippet service, this would mean actually running the code inside AEGIS and reviewing its results.
Edit a service	The step follows a dual approach to support both coding and non-coding users. Where applicable and available, metadata are used to support the user in making more informed decisions on how to alter/extend/update the existing service. Depending on the nature of the service, the editing process will be different. In the case of code snippets and interactive reports, if

	semantic information is available, the process is simplified through the automated enabling/disabling of certain options.
Save data/service	All assets created in AEGIS can be saved for future usage.
Export and Create Advanced Service	These two steps address the needs of the users that will progress from big data experimentation to providing production-level big data enabled services to other stakeholders, leveraging the complete power of AEGIS. They require an experienced user with technical background to assist in deploying the services.
Annotate	Users may optionally annotate datasets and services in order to make them searchable by others and enable their usage in all advanced metadata-enabled functionalities.
Make available for share	Sharing datasets and services here refers to the act of making a dataset/service searchable by others through the advanced metadata-enabled exploration functionality. The act of providing access to a dataset/service to collaborators is performed in the "Add collaborators" step described later on. Depending on the configured permissions and pricing, a searchable asset may be made directly available to other AEGIS members or under specific terms. The step therefore entails the business brokerage functionalities of AEGIS, i.e. the mechanisms to monetise or otherwise exploit the created assets. It should be stressed that only annotated assets can be shared in this way in order to ensure measurable quality and usefulness.
Add collaborators	Provide access to other AEGIS members in a project or for a given dataset. Permissions granted can be for read-only or read-write access in a project level but also per dataset (the term dataset is used here to denote a set of files).

4.2.1. Methodology instantiation: Creation of interactive report

The first instantiation example for the AEGIS methodology showcases how an idea for a data-enabled service related to PSPS can be drafted, enriched, implemented, refined, perfected and finally offered to 3rd parties through AEGIS as an interactive report. The workflow essentially encompasses all the phases that a group of colleagues with complementary responsibilities and competencies should go through in order to create a data analysis of high value with business potential.

In order to provide a more intuitive description of the workflow, the following simple scenario is assumed:

B is a product manager (i.e. a business user) in an SME in the domain of smart automotive services and is interested in exploring the potential usage of some car sensor data that the company has been collecting since the sensors came to its possession and got installed some months ago, without having a clear view on their exploitation. B is working very closely with D, a data analyst with strong mathematical background and analytical thinking who is not very competent in coding (i.e. non-coding data analyst) and E, a developer who, among other things, assists D when programming is required.

The scenario is executed in five distinct phases, starting from familiarisation with AEGIS and concluding with the creation of a ready to be consumed data-enabled service. For each phase, a list of methodology steps to follow is presented. It should be noted that the example is simplified for demonstrative purposes.

Phase 1: M defines the high-level service requirements

1. M registers in AEGIS
2. M explores public data and services to assess AEGIS functionalities/potential
3. Based on the available previews, M decides that there are some datasets that could be useful.
4. M creates a project and uploads there a large subset of the available sensor data.
5. M adds D and E as collaborators to the created project.
6. M explains to D and E the idea for making something useful out of the combination of internal data with some public AEGIS data.

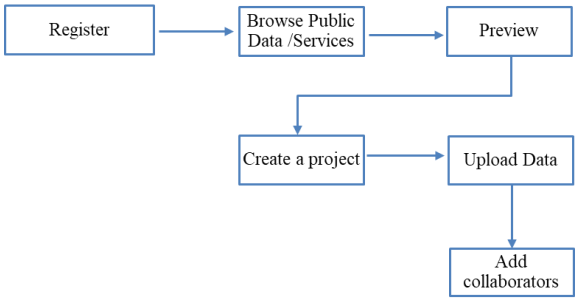

Figure 4-2: Business user workflow – 1st Methodology Instantiation (Phase I)

Phase 2: D performs the core data experimentation and analysis

1. D registers in AEGIS
2. D joins the project created by M
3. D uses the semantically enriched (metadata-enabled) browser and selects various public datasets, including one with historic weather data, which he adds to the project. In the background, snapshots of the selected datasets are created and added in the project.
4. D chooses the weather dataset and iteratively applies some pre-processing methods in order to (a) bring the datetime field in the desired granularity, (b) remove certain

unnecessary fields and (c) replace some city names with the values used in the internal dataset (UI-based)

5. D creates a view (subset) on the selected dataset (UI-based)
6. D merges the created view with internal data available inside the project (UI-based)
7. D selects and iteratively configures the data analysis algorithm to apply to explore driving patterns related to weather conditions in several areas (UI-based)
8. D chooses the appropriate (map-based) visualization for the result (UI-based)
9. D saves the created report-like service

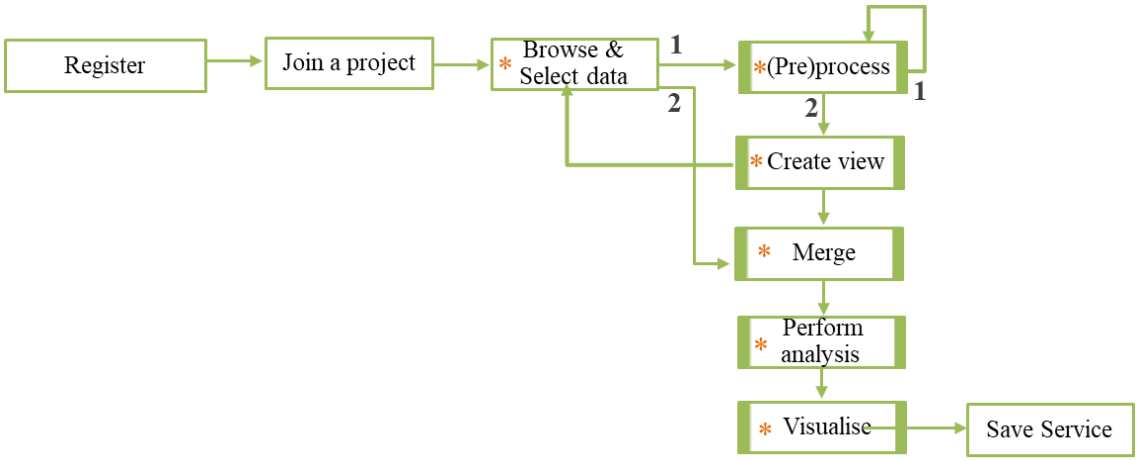

Figure 4-3: Data analyst workflow – 1st Methodology Instantiation (Phase II)

Phase 3: E refines the created result

1. E registers in AEGIS
2. E joins the project created by M
3. E selects and edits (through coding) the service created by D. Specifically, E implements a custom visualization that combines two layers of information and exposes certain parameters for the user to interact with and dynamically change the visualization result.
4. E saves the result as an AEGIS service (report)

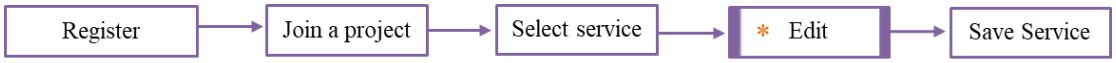

Figure 4-4: Developer workflow – 1st Methodology Instantiation (Phase III)

Phase 4: M reviews the created interactive report

1. M signs in AEGIS and enters the project
2. M selects the created report and tries it (runs it)
3. M changes the values for the exposed parameters and identifies some interesting cases.

4. M decides that with few changes the interactive report could be useful as a safe driving application. M plans to show it to a prospective customer, but also wants to see if other AEGIS users would be interested and therefore asks D and E to provide this sample as a free service available for others.

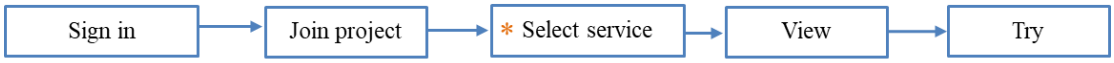

Figure 4-5: Business user workflow – 1st Methodology Instantiation (Phase IV)

Phase 5: D creates a searchable AEGIS service

As there is no need for advanced service creation that would require coding, E can complete the last phase without help.

1. D signs in AEGIS and enters the project
2. D selects to view the created report (service)
3. D provides a set of annotations for the service, i.e. completes the required information in a form, which will make the report searchable inside AEGIS in a semantically-aware manner. D marks the service as free of charge. Thanks to the smart browsing/search, the service will be discovered by people working on relevant datasets and analyses, therefore requests for more information or collaboration are expected to be more targeted.
4. D exports the report as a searchable AEGIS service.

Figure 4-6: Data analyst workflow – 1st Methodology Instantiation (Phase V)

4.2.2. Methodology instantiation: Exploration and experimentation with PSPS-related datasets and services

The second methodology instantiation example showcases how AEGIS can be used as an experimentation playground that impels the exploration of innovative ideas on ways to combine datasets and extract meaning. The underlying workflow spans across the complete big data value chain and could be seen as part of various larger workflows that include collaborations among colleagues (like the previous example) or interactions between service providers and customers leading to iterative enhancements of the initial offering. However, the focus here is on how data analysts can benefit from AEGIS to discover datasets and ways to use them towards building their own enhanced solutions.

Like before, a scenario is provided to make the description more intuitive: B is a junior data analyst in a large insurance company, tasked with performing specific analyses on certain predefined datasets and reporting back when irregularities appear that could imply misuse of the

company's insurance provisions. B believes that in many cases publicly available data could provide very useful insights if properly used and combined. However, such datasets are often noisy, i.e. contain inaccuracies or are missing information, and require a lot of effort to cleanse, hence the company does not consider them useful. B is competent in coding and would like to experiment with other types of data as well, even independently of work, however does not have the time to devote in creating something from scratch. B will leverage AEGIS to speed up the required processes and explore some of her ideas without investing disproportionally significant time and effort.

The scenario is described in one phase (since there is only one actor) comprising 12 steps. Again, the example is simplified for demonstrative purposes.

1. B registers in AEGIS and creates a new project.
2. B opens the semantically-enabled data browser and iteratively loads several public datasets and creates views on them to get a better understanding of their content.
3. B has some ideas but decides to have a look on the available services first in case one of them can be used.
4. B browses the code snippet services and selects one that, based on the available metadata, combines fabricated data for antibiotic prescriptions, public weather data and flu-related mentions extracted from social media. B likes the fact that the fabricated dataset is created to statistically match the properties of a real dataset which could not be provided due to data sensitivity. The provider of the dataset is a well-known pharmaceutical company, so B trusts that the data will be reliable for analysis purposes.
5. The service is linked to the underlying data, which in this case are publicly available, so B can run the code. The final output is a correlation between flu mentions and weather, shown in a heatmap. When hovering over the various regions, the number of antibiotic prescriptions is also shown. B does not see much value in the analysis itself, however a lot of work has been done in terms of processing the unstructured data with the flu-related mentions and in cleansing the public weather data. B decides to edit the service (in the background a copy of the service is created and added in B's project).
6. B first wants to explore what other datasets could be combined with the ones already included. From within the service editing interface, B opens the semantically-enhanced dataset browser which now provides more targeted results that are relevant to the data types of the existing service and it is easier to identify which are more promising. B discovers a dataset containing locations and times of smart inhaler devices' usage for a period of three years. The dataset is searchable but not publicly available, so B has to wait for the owner's approval in order to get access to the actual data.
7. The owner knows that the dataset does not contain sensitive data but prefers to monitor who requires access and is also interested in the way it can be utilised. The owner updates the dataset's permissions to share it with B.
8. B is not very familiar with time series data but uses the enhanced data processing and merging AEGIS functionalities for guidance in transforming the new dataset in a way that can be integrated with the other datasets.

9. B selects one of the available correlation algorithms and applies it first to explore the relation between weather and inhaler usage. B can easily include the flu-related mentions in the analysis, all through the provided UI, since the dataset had been already pre-processed for the original service.
10. B wants to change the way the antibiotic prescription dataset is used, since B is aware of some patterns to look for in order to exclude some records. For this part B prefers to write a custom code which is integrated easily in the rest of the service. B also saves the improved dataset as a private dataset for future usage.
11. Finally, B can create a new visualisation to showcase the findings.
12. B annotates the visualisation and only makes this searchable, instead of the complete code snippet or the report. B might propose to the manager to use this internally combined with real customer data, but also plans to revisit it and validate whether the same conclusions could be reached with different dataset combinations and/or different time periods and regions.

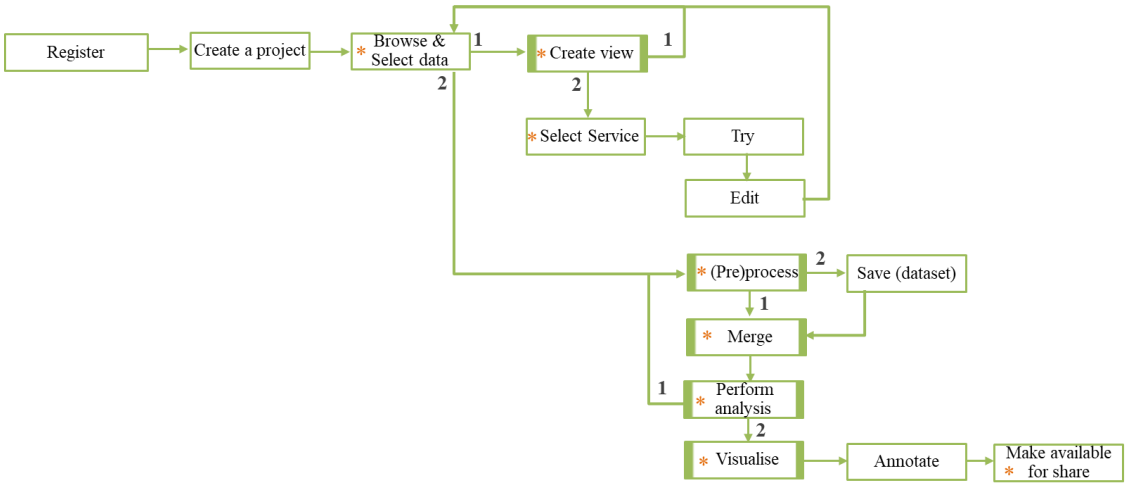

Figure 4-7: Data analyst workflow - 2nd Methodology Instantiation

4.3. MVP definition

The Deliverable D1.2 provided a description of a set of features suitable to make up an early definition of the AEGIS Minimum Viable Product (MVP) based on the requirements coming from high-level AEGIS scenarios. Most of the features used to belong to the Core Big Data Value Chain processes, further clustered in four internal groups, conceptually described in the D1.2 as:

- Related to data and results visualisation
- Related to multi-source and multi-format configurable big data analysis
- Related to data-as-a-service discovery, exploration and acquiring
- Related to more advanced experimentation and configuration of the provided by AEGIS building blocks that address the needs of more tech-savvy and data-savvy users

Furthermore, an initial hypothesis of business models for AEGIS has been outlined, describing it as “a service marketplace and big data-enabled business intelligence creation space for all

stakeholders across the PSPS value chain", in principle a marketplace and subscription business models. It is worth noting that in a marketplace business model it is relevant to recruit vendors (in this case, users as data providers) and monetization is based on a commission per sale; whereas, subscription asks for a focus on customization and maintenance of the services (in this case, business intelligence solution), and monetization is based on the time of access and features used. Thus, providing users with the following features:

a) *flexibility* in terms of data formats the platform can handle (*Marketplace feature*);
b) *discoverability*, acquiring and consumption of interesting data services and seamless combination under a PSPS semantic context (*Marketplace feature*);
c) *selection* from predefined options and the application of various algorithms on the cloud targeting both generic and more specific domain needs (*Subscription feature*);
d) *intuitive easy to create visualisations*, through a set of available visualisation options, configurable to an extent and easy to combine in user created dashboards (*Marketplace and Subscription feature*).
e) *export the visualisations* and analysis results for easier consumption and sharing with others (*Marketplace and Subscription feature*).

Yet, it is worth noting that in the Deliverable D5.1 the following early elaborations of the AEGIS mission and value proposition have been proposed

- *AEGIS aims to drive a **data-driven innovation** that expands over multiple business sectors and takes into consideration structured, unstructured, and multilingual datasets, rejuvenate the existing models and facilitate all companies and organisations in the PSPS linked sectors to provide better and personalized services to their users. (**AEGIS Mission**)*
- *Hence AEGIS aims to develop a curated, semantically enhanced, interlinked and multilingual data platform for PSPS - to allow businesses and developers to provide better and personalized services to users. (**Preliminary AEGIS Value Proposition**)*

Compared to current players in the data platforms competitive environment, the focus on "data-driven innovation" in Public Safety and Personal Security (PSPS) sectors is relevant to differentiate the AEGIS proposals. Furthermore, we argue that data-driven innovation should be accessible and not bounded by technical (advanced knowledge of data management, statistics, etc.) or technological issues (advanced knowledge of big data infrastructure components), especially in PSPS related businesses or organization, where also lay users or managers with no data scientists background are called to take action in decision-making or service proposals/design. Thus, considering these issues, and the set of target users provided in the Deliverable D7.1, a specific choice has been made to identify the AEGIS MVP, that is elaborated as follows:

- *AEGIS Main Target: Easier Transition to Big Data Analysis in the Public Safety and Personal Security domains for **tech-wise non-advanced users***

The focus on tech-wise non-advanced users, which is also evident in the methodology description provided in the previous section, does not prevent the use of the platform by advanced users. Actually, considering the evolution of the business model of the AEGIS platform as first a "Two-

sided platform" to further develop it as a "Multi-sided platforms" (Bharosa, Janssen, Klievink, & Tan, 2013; Eisenmann, Parker, & Van Alstyne, 2006; Hagiu & Wright, 2015), due to the service side of AEGIS, we can see an opportunity, for example, for advanced data scientists (acting as suppliers) to provide their datasets elaborations (e.g., views) for a fee to tech-wise non-advanced users (acting as customers) or vice-versa these latters providing their datasets for advanced elaboration to advanced data scientists, under a collaboration agreement enforced by the platform itself. In any case, the MVP should support network effects to increase the number of datasets offered as well as the number of users demanding for them.

As to these issues the above-mentioned features identified in D1.2 and the ones related to the methodology discussed in this deliverable, they all should be oriented towards the building of *accessible* (also in terms of channels: e.g., via mobile) and *connectable* ***personal project spaces*** (i.e. with "must have" security and privacy features that allow easily to decide when, how, and what to connect within each space), which should enable the platform dynamics just exemplified. As for the tools, it is relevant to provide visualization of results for data analysis concerning first descriptive statistics and a guided way to perform basic inferential statistics, such as, e.g., t-test, Analysis of Variance (ANOVA), Analysis of Covariance (ANCOVA), regression analysis, factor analysis, multidimensional scaling, cluster analysis. It is worth noting that from discussions reported in D1.2 and D5.2 it emerges that part of the MVP could be not only the core big data analysis technical enablers that constitute the AEGIS platform, but also advanced services implemented through them and made available for usage/ consumption to the project's community. A potential candidate in this context could be the Event Detection tool that is configured to provide insights on PSPS-related events for specific regions.

5. AEGIS ETHICAL, PRIVACY, DATA PROTECTION AND IPR STRATEGY – FINAL

5.1. Objectives

AEGIS Ethical, Privacy, Data Protection and IPR Strategy (in brief "EP Strategy"), outlined in this section, will serve:

i) to define the regulatory framework for data protection, IPR and Ethical Issues that will drive the Data Policy framework of the AEGIS platform and comply with EU directives on data safety and privacy;
ii) to illustrate an overview of AEGIS platform and components, focusing on portions of the system processing personal data, as well as representing the purpose of the processing of personal data and describing the origin of personal data and its collection method;
iii) to elicit the legal, data protection and ethical requirements (legal, technical, organisational, personnel and material requirements), providing input to the use cases, the architecture and specification task and specifying the measures to cover these requirements for data protection, and
iv) to assess to what extent they have been taken into account during project implementation and within the final AEGIS system.
v) to define ethics roles, procedures and roadmap.

5.2. Relations to internal AEGIS environment

AEGIS EP Strategy is strictly interrelated to the overall project implementation and final achievements, being aimed at providing the basis for the main guidelines that AEGIS Consortium will have to respect towards ethics, privacy and data protection, to be constantly updated during project's lifecycle. Its final release, notably regarding the Data Policy framework, will be delivered in D1.3 "Final AEGIS Methodology".

Given this, AEGIS EP overall Policy is particularly interconnected with:

- T2.2 "Data Policy and Business Brokerage Frameworks", because this is devoted to the design of the core methods for powering both the Data Policy Framework and the Business Brokerage Framework, including categories and predefined lists to describe data IPR, security, trust and quality features, as well as extra tag for the classification of personal and sensitive data, IPR annotations, and methods to cross-check IPRs and allow a semi-automatic negotiation;
- WP4 "AEGIS Infrastructure Implementation and Rollout" and WP5 "AEGIS Data Value Chain Early Community Demonstrators", because AEGIS EP Strategy supplies key input to the use cases, the architecture and the specification task, thus representing the reference point for assessing to what extent the legal and ethical requirements have been taken into account;
- T6.4 "Project Data Management Handling", since this task is expected to work in synergy with T1.4 and with WP9, though from different and complementary perspectives, in view of continuously monitoring the data protection and ethical issues of the project, as well as the IPR issues of the data to be contributed to the platform;

- T7.1 "Project and Demonstrators Exploitation Planning and Data Sharing IPR Definition", where a special focus should be given to the IPRs not only of the technology but also of the data to be exchanged over the AEGIS platform;
- WP9 "Ethics Requirements", pursuing the compliance with the listed set of "ethics requirements" that the project must comply with. A close connection is established particularly with D9.1 "OEI – Requirement N° 1", where the EP Strategy described in this deliverable will completed with the overall Data Protection Impact Assessment methodology, outlining how to assess EP Strategy's implementation, both as regards the demonstrators' operations and the overall design and development of AEGIS architecture. Also, the other requirements set out in WP9 pertain to the ethical and legal concept are tackled in this document:
    - opinion or confirmation by the competent Institutional Data Protection Officer and/or authorisation or notification by the National Data Protection Authority, to be submitted where applicable (D9.2);
    - Ethics Advisory Board's periodic reports to the Commission on the implementation of the ethical concerns (issues) in project and on compliance with applicable national and EU regulations (D9.3). These reports will refer to both this document and to the assessment methodology depicted in D9.1

5.3. Regulatory Framework

5.3.1. Introduction

AEGIS' use of technologies could potentially interfere with the right to privacy and the protection of personal data. It is therefore important to analyse the regulatory framework concerned and thus providing safeguards against the potential pervasiveness of AEGIS solutions, in order to design and develop them in a privacy-friendly fashion.

The main aim of the regulatory framework is to guarantee the individuals' sphere of autonomy within which to operate. The main legal instruments relevant to AEGIS pertain to privacy and data protection and contain a set of substantial safeguards and countermeasures against the spread of technologies resulting in an unfettered surveillance: the following chapters outline the key aspects of such regulations, relevant to project's progress and results.

Furthermore, in addition to legal provisions and principles, we will refer also to ethical, social and political oriented values applicable to AEGIS results and activities, being the "privacy in law" concept strongly interconnected not only with a number of legal values and principles -foreseeability, accountability, legality, necessity, proportionality and transparency, etc.-, but also with principles and values of ethical, cultural, social and political nature.

Within AEGIS EP Framework, and in AEGIS requirements' definition, it is therefore imperative to take all this set of variables into account in a systematic way.

The consideration of these principles will let us answer the questions why privacy matters in AEGIS R&D implementation and final system and how it should be safeguarded.

Before starting the overview, a remark has to be borne in mind: **this chapter has the ambition to look at the AEGIS project from a legal perspective**, and not to present a comprehensive analysis of the European regulatory framework of privacy and data protection - that would fall outside the scope of this deliverable.

The main documents that will be addressed are:

- European Convention of Human Rights
- Charter of Fundamental Rights of the European Union
- Regulation 2016/679/EU (GDPR), repealing Directive 95/46/EC ("Data Protection Directive")
- Directive 2002/58/EC "ePrivacy Directive"
- Regulatory Framework in the selected jurisdictions, stating how privacy and data protection norms and principles are implemented in each country where the demonstrators will operate.

This composite regulatory system applicable to AEGIS is completed by European Courts' case law: though the legal system may appear somehow vague and fragmented, such a jurisprudence is very helpful for partially filling the gaps and pitfalls that can be found in legislation.

This overall framing represents the basis for setting the AEGIS ethical, privacy and data protection requirements, which will emphasise existing legal and ethical safeguards, boundaries and obligations to ensure the legitimacy and fairness of AEGIS final solutions and actions.

5.3.2. Privacy Concept and Data Protection Concept within the European regulatory system

As a starting point, it is useful to briefly examine the right to privacy and right to data protection concepts:

1. **Privacy concept**. Privacy is an ambiguous and contentious concept, varying according to time, space and peoples. It shifted from the "right to be let alone", referring to the realm of intimacy and wish for solitude, as a concept hinging on physical privacy, to a broader notion of privacy, referring to the relationship between the individual and other individuals, based on "the claim of individuals, groups, or institutions to determine for themselves, when, how and to what extent information about them is communicated to others". In this renovated meaning, the privacy concept encompassed several other aspects and embraces several rights, ranging from the right to be left alone and to enjoy solitude, to the right to individual autonomy, the right to control information about oneself, the right to a private life, the right to limit accessibility, the right to minimise intrusiveness, the right to exclusive control of access to private realms, the right to expect confidentiality, to the right to enjoy intimacy, reserve and anonymity and the right to secrecy.

Both the European Convention of Human Rights and the European Court of Human Rights' consolidated jurisprudence recognise the right to privacy, promoting a living interpretation of the same, in the light of existing conditions.

2. **Data protection concept**. It was considered for a long time as a corollary of the right to privacy

and is a relatively new autonomous human right in European legislation. This right had a new legal source of legitimacy in European legislation since the entry into force of the Lisbon Treaty. As recognised by the jurisprudence of the European Court of Human Rights and of the European Court of Justice, there is a tight relationship between privacy and data protection: the protection of personal data is functional to the enforcement of the right to privacy and, subsequently, the infringement of the individual's right to data protection leads to a violation of the right to privacy. Even so, these two rights don't totally correspond: not every privacy infringement results in a violation of the right to data protection. Data protection is more specific than privacy and is applicable every time personal data are processed.

5.3.3. European Convention of Human Rights and Charter of Fundamental Rights of the European Union

The recognition of privacy and data protection as fundamental human rights in Europe relies on the European Convention of Human Rights and the Charter of Fundamental Rights of the European Union, whilst at an international level, the Universal Declaration of Human Rights (1948) recognised the privacy as a fundamental human right by protecting territorial and communications privacy.

I. European Convention for the protection of human rights and fundamental freedoms

The European Convention of Human Rights (1950), in particular its Article 8, deals with private and family life, home and correspondence of the citizen. Since then, more enforceable European tools surpassed its value in the field of data privacy.

Article 8 recognises the privacy as one of the human rights and fundamental freedoms. It states as follows:

"1. Everyone has the right to respect for his private and family life, his home and his correspondence.

2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others."

The European Court of Human Rights' jurisprudence pointed out that private life concept extends to aspects relating to personal identity (e.g. an individual's name or picture, but also other means of personal identification and of linking to a family) and that therefore, the right to privacy established by this provision refers also to identity and personal development, also within interaction with other individuals, even in a public space, as well as to the right to establish, maintain and develop relationships with other human beings in general. This Court's case law also confronted with situations involving new technologies and its interpretation was taken in account by the Consortium and will be further taken into account in future AEGIS progress.

Article 8.2 states the lawfulness criterion, in the meaning of rule of law: it states a negative obligation for public authorities whilst allowing exceptions for interferences that are "in accordance with the law". Such rule of law is very important to ascertain the boundaries between the use of technologies (like AEGIS solutions) and democracy. The lawfulness criterion is the first step in assessing whether technological solutions are in line with Article 8.1, and it has to be applied on a case-by-case basis.

Essential legal principles for data privacy:

Data sovereignty

Sovereignty as constitutional term means that the people have an effective influence of the exercise of public power.[1] What is meant for the constitutional construction of a state to clarify the relation between the government and its citizen can be applied to the relation of a company to its customers respectively. In both cases, the power of the superior person derives from the rights of all inferior persons as individual. Essential for this relation are regulations determine the scope and borders of rights and obligations.

From the term *sovereignty*, two aspects can be derived. As already mentioned, the person concerned is capable of controlling a specific action guaranteed by law. But more import, to have an effective use of its own rights, the person implicitly has to be aware in which way his power can be exercised and what the exercise of this power effects. Being holder of specific rights alone does not constitute an effective legal status as long as the person does not know how to make effective use of it. Especially in situations where single consumers stand against huge companies operating Big Data applications, the obvious gap – imbalance of power - needs to be bridged. To avoid this kind of overpower and instable balance (of power), a legal framework must determine particular rights and obligations that intend to equate both roles.

Self determination

One fundamental characteristic of privacy is that its understanding varies from person to person. The subjective perception of privacy and the different handling of privacy concerns represent its ambiguousness. Due to the lack of objectiveness, an uniform and clear definition of privacy does not exist.[2] Consequently, privacy depends on a subjective understanding that is affected by technical, social and economic conditions.[3] One the one hand privacy can mean the need for delimitation of publicity which guarantees a private area, in which the individual is capable of dispose its own material and immaterial resources without influence of external factors.[4] As privacy is expression of the personal development, an essential part are all information concerning the personality, character and identity as well as particular circumstances. To quote the German Federal Constitutional Court in the process constituting the basic right of

[1] Maunz/Dürig/Grzeszick, 79. EL Dezember 2016, II. Rn. 61
[2] ZD 2015, 517
[3] ZD 2015, 517
[4] ZD 2015, 517

informational self-determination in German, "*the individual itself must fundamentally have the right to disclose all information concerning personal facts in self-chosen limits at any time and any place*".[5] Subject of this basic right and fundamental for the data privacy law are personal data, which are all information relating to identifiable natural person[6]. Background of this judgement are concerns regarding the procedure and the indeterminacy of the national census.[7] One feared that public authorities excessively collect and store data about citizens not necessary for the purpose of the national census[8].

According the judgement, the protection of informational self-determination grants the individual to decide about the disclosure and use of his personal data that explicitly covers the collection, storage, process and disclosure of personal.[9] The result of the protection of personal data and informational self-determination has been realized with data privacy laws[10] which guarantees power to control the handling of personal data with others in order to avoid unlawful interferences. On level of the European Union, a basis right for informational self-determination can be derived from article 8 of the European convention on human rights.[11] With regard to automated profiling and decision finding, where the individuals cannot overview the collected data source of which a profile has been created, automated decisions can lead to arbitrary results, which have an important influence in cases of granting a credit for a private or commercial purpose or the creditworthiness necessary for rental contract.

The key principles of data privacy in the field of European legislation have been adopted with the Convention on Data Protection[12], constituting common guidelines with the intention to give the member states an orientation for national regulations promoting an European wide integration of data protection principles.[13] According to Article 1 of this Convention is to ensure that every person regardless of their citizenship is protected in their basic rights especially in their general right of privacy protecting the individual in automated processing of personal data[14]. With regard to Article 8 of the Charter of fundamental rights of the European Union, the European legislation guarantees a certain level of protection and furthermore requires the process of personal data is basically prohibited unless consent has been given. Therewith, all European member-states are bound to this fundamental data protection level.

[5] BVerfG Urt. v. 15.12.1983
[6] BVerfG Urt. v. 15.12.1983
[7] BVerfG Urt. v. 15.12.1983
[8] BVerfG Urt. v. 15.12.1983
[9] BVerfGE 65, BVERFGE Jahr 65 Seite 1 (BVERFGE Jahr 65 43
[10] Respectively its improvements – data protection law (in Germany) already existed before this judgement.
[11] Calliess/Ruffert/Kingreen EU-GRCharta Article 8 Rn. 4
[12] 23.01.1981
[13] Stern/Sachs GRCh p. 213
[14] Stern/Sachs GRCh p. 213

Autonomy – (the way of handling personal data within the rights granted)

Data Autonomy means to realize and control the own perception of privacy. Due to its dynamic adaption and different perception depending of individual needs and desires, the legal frame of privacy should thereby satisfy two functions. It should enable those person to extent their privacy and protect those person against interferences. Generally, the individual has the free disposal to allow interference of specific basic rights.[15] The interference has to conform to the individual´s will constituting the basis on which the interference of the basic right is granted.[16] The legitimation of the data controller to collect and process data arises from the individual autonomy and self-responsibility, which has, in return, to be regulated in the legal framework overall and recognized in the legal relationship between the data controller and the individual.[17] The realization of self- responsibility and autonomy requires the conscious decision including an appropriate knowledge basis in order to estimate evaluate the scope and degree of (first function) and to control and regulate the interference (second function).

II. Charter of Fundamental Rights of the European Union

The Charter of Fundamental Rights of the European Union was proclaimed and published in December 2000 and then became legally binding in the EU Member States since the adoption of the Treaty of Lisbon on 1 December 2009.

The Charter refers to both the right to privacy and the right to data protection, containing an explicit right to respect for privacy (Article 7), as well as an explicit right to protection in case of personal data processing (Article 8). Both of these provisions have to be applied in coherence with European Court of Human Rights' interpretation of Article 8 of the European Convention on Human Rights.

Article 7 reads as follows:

"Everyone has the right to respect for his or her private and family life, home and communications".

Article 8 reads as follows:

1. "Everyone has the right to the protection of personal data concerning him or her.
2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3. Compliance with these rules shall be subject to control by an independent authority".

[15] Freedom of disposal is restricted to constitutional order - Kindhäuser/Neumann/Paeffgen, Strafgesetzbuch, StGB § 228 Rn. 3.
[16]Kindhäuser/Neumann/Paeffgen, Strafgesetzbuch, StGB § 228 Rn. 4
[17]Kindhäuser/Neumann/Paeffgen, Strafgesetzbuch, StGB § 228 Rn. 4

5.3.4. GDPR as legal reference

In 2016 the European parliament passed the European data protection regulation constituting an uniform legal framework within the European member-states. One improvement compared to the data protection directive 95/46/EC is the regulatory nature. Compared to a directive, a regulation, passed by the European institutions, is directly applicable pursuant to article 288 sec. 2 TFEU.[18] The scope of this assessment refers to European wide regulations, taking into account the European data protection regulation and other regulations for the legal assessment.

5.3.4.1. Material scope

According to Article 2 GDPR this regulation applies for the process of personal data. In contrast to anonymous data, personal data contain personal identifier that can be assigned to a person. Consequently, it is necessary to define and determine the term personal data, as all articles listed within this regulation require the presence of personal data. This comprises the conditions for lawful data process, the rights of the person concerned as well as the obligations of the data controller including the technical and organizational safeguards. The term personal data is fundamental for this methodology and therefore needs to be appropriately defined.

5.3.4.2. Definitions

Hereinafter, we will outline the concepts and definitions of personal data and data processing relevant to AEGIS, which remained substantially unchanged.

I. Personal Data: definition and classification

According to article 4 sec. 1 GDPR, personal information means any information relating to an identified or identifiable natural person. Generally, this includes obvious information like the name, an address, telephone number, appearance of a person or the license plate.

Notably, Article 4 provides the following broad definitions of "personal data" and of "processing of personal data":

- "**Personal data"** means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". Therefore, the Regulation is applicable only to data subjects as natural persons, notably as human beings. According to this definition, personal data may be:
    - Identification data, which directly identifies the data subject, being pieces of information acting as identifying factors and able to distinguish a data subject from all the others;
    - Indirect identification data, which makes possible only an indirect identification of the data subject, through an association with other available information. The

[18] ZD 2017, 556

> wording "other information available" refers both to other information available to the data controller (entity primarily in charge of the data processing) and to any information that may be possessed by any third party. It is considered sufficient, in view of the application of the regulation (as well as of the Directive), the potentiality of identification. Anonymous data, though not directly referring to a specific data subject, may keep this potentiality of identification.

In order to determine the type of present information and other kinds of information, there are two theories prevailing for determining if they include personal data.

- Absolute theory: The assignment of information to a natural person is independent of the actual informational basis of the data controller with considering the relation of effort necessary to gain the information needed to identify a person[19]. This is assumed, when any third person can relate information to a person. In short: With the theoretical possibility to identify a person with extern available information and on basis of the data controller´s dataset available, personal related data is present.
- Relative theory: In contrast to the absolute theory, decisive are the actual knowledge, means and possibilities of the data controller available[20].

In the judgement of the European Court of Justice concerning the question of the presence of personal data, relevant for determining its presence are the role of the data processor and the actual and potential legal means available to determine a person - a combination of absolute and relative theory.[21] The possibility to identify a person on basis of a particular information or bunch of information can differently interpreted by different persons as data controller[22]. For example an internet provider can interpret IP addresses differently than an average internet user or a website owner[23].

According to the European Court of Justice, excluded are those means that require a not proportional effort of time costs and human performance/workforce[24]. By determining the probability of identifying a natural person, an objective standard applies, which means that the motivation and intention of the data processor are not relevant for this question. Primarily objective factors have to be considered e.g. the possible accessible means on the market at the time of processing data. Consequently, the time of identification is not defined when the natural person is actually identified, but earlier in the moment of processing data when with appropriate means the identification of a person is reasonable certain[25].

Important is the relative meaning of personal data in context of the term *identifiable*. Article 4 sec. 1 GDPR differs between information referring directly and indirectly to an identifiable

[19] ZD 2017, 223
[20] CR 2016, 234
[21] MMR-Aktuell 2016, 382533
[22] Auer-Reinsdorff/Conrad, IT- und Datenschutzrecht, § 36 Datenschutz der Telemedien Rn. 68
[23] MMR-Aktuell 2016, 382533
[24] CR 2016, 235
[25] Kühling/Buchner Art. 4 Rn. 22

person. Basically relevant are two categories of data: Personal data and non-personal data. As the first term has already been mentioned, the second category can be divided into two further kind of data:

- Anonymous data, consisting of data that do not allow neither directly, nor indirectly, the identification of the data subject. Anonymous data are information that do not contain any information about a natural person. The person-related information have been cleared from the datasets. Such data do not fall under the scope of the GDPR and are not relevant in the legal assessment. As specified by Article 29 Data Protection Working Party, it is "any information relating to a natural person where the person cannot be identified, whether by the data controller or by any other person, taking account of all the means likely reasonably to be used either by the controller or by any other person to identify that individual. Anonymised data is anonymous data, which previously referred to an identifiable person, in case such an identification is no longer possible, usually thanks to processing and elaboration activities. This data does not fall within the EU data protection legislation. However, its first gathering, elaboration and processing were performed on personal data: therefore, data protection legislation has to be applied in such activities, until data is made anonymous. It may also happen that, under certain circumstances, anonymised data receives protection in European member states' national data protection legislations or through Article 8 of the European Convention on Human Rights.
- Pseudonymous data, consisting in personal data that, after its processing, become quasi-anonymous data: after such a processing, though there is the possibility to identify the data subject, the data Controller, according to the lawful data processing and data quality principles, makes the identification more difficult after their collection. Pseudonymization of personal information is a procedure where person-related information is replaced by non-identifiers in order to ensure that these informational cannot be assigned to natural persons anymore. In particular, the Regulation states that "pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person". Therefore, though the use of pseudonymised and key-coded data is fostered by the European legislation to protect personal data (since it lowers the possible risks for the data subject), in case the data subject remains indirectly identifiable, this kind of data too is subject to application of the European regulatory instruments (Data Protection Directive and then Regulation). It should be noted, in fact, that in relation to key-coded data, Article 29 Data Protection Working Party followed the rule that, if the data subjects may be identified starting from the such data, "taking into account all the means likely reasonably to be used by the controller or any other person", it is personal data and therefore Data Protection Directive is applicable. The assessment has to be done on a case-by-case basis, considering all the specific circumstances concerned.

II. Processing of personal data

According to Article 4, "processing of personal data" ("processing") means "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by

automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction".

III. Implications in Big Data applications

One main characteristic of Big Data is the ability the decentralized storage capable to handle and process an enormous amount of data. The classification of data is domain dependent, which means that a data scientist will differently categorize datasets by means of different characteristics, like what are the datatypes, are structured or unstructured data present, are metadata available and so on. A data scientist intends to maximize the result of the data process pursuant to the expected intention. Whereas the legal classification examines data by its assignment to a person or an object. The presence of persona related information can lead to the application of data protection regulations intending to guarantee informational self-determination whereas objective-related information like copyright related information or business and trade secrets lead to the application of regulations intending to protect material and immaterial related goods.

Personal data are subject of data protection regulations, so that the determination of its presence is highly relevant for Big Data applications. The necessity to classify between personal and non-personal information derives from the presence of information that alone are no personal information, so called "reference data" but linked together, e.g. in data mining process, provide represent personal data with identifying a natural personal. The relative theory assumes that the data controller will only use technologies reasonable for the pursuit purpose and no or illegal complicated means. Regarding the question of *identifiable*, the recitals of the GPDR propose to take into account "*of all objective factors, such as the costs of and the amount of time required for identification, the available technology at the time of the processing and technological developments*"[26]. The technology-neutral approach in context of "*technology at the time"* should consider those technologies that are efficient, suitable and available for everyone and based on scientific knowledge in theory and praxis[27].

The task for current data processing technologies is to develop tools or implement useful creating an environment so that data and data results ensure individual privacy. Basically, anonymization intends to hide identifiable or sensitive data of an owner by removing explicit identifiers[28].

Concerning personal data, there are two main problems representing a source of danger:

1. Data reusability

The identified problem derives from the neutrality of data. Data is nothing more than a value, which has no particular meaning. The scalability and static growth of the data storage enables the possibility to combine various values of data to a potential infinity space, to allow variable insights of new information. As result, data in combination with a particular purpose can be

[26] Recital 26 of the GDPR - https://gdpr-info.eu/recitals/no-26
[27] ZD 2017, 224
[28] Privacy preserving p. 13

interpreted differently than in combination with another purpose. In short: The purpose of the data process allows to re-use data to various conclusions. For example (taken from ENISE - Privacy by design in Big Data):

"Mobile apps providers collect personal data in order to provide users with information about their fitness or health status. These data can be valuable to insurance companies and/or other providers who may target specific users"[29].

2. Data re-identification

Anonymizing personal data is not the goal, as re-identification becomes more problematic with stronger algorithm and extremely wider data storage. With combining various non - personal respectively anonymized data with advanced analytics, there is the possibility to infer information related to a person or a group[30]. A person is identified when information combined will allow the individuals to be distinguished from others and therewith create a context to a natural person.

Important problem resulting from de-anonymization and the wider presence of personal data is the extent of the application of the general data protection regulation. Some expertise proposes that with the increasing enhancement of Big Data applications, every data represents personal data with the consequence of the "*end of anonymity*"[31]. Legally problematic is the assignment of data as anonymized data, as data is held as anonymized, if due to the personal, temporal, technological and financial cost are that much that de-anonymization can no longer be anticipated[32]. Insofar, the aforementioned criteria are, due to the rapid technological progress, decreasingly devalued, that de-anonymization is likely to be expected and realized procedures like profiling and automated decision-making.[33] Consequently, relevant criteria for the determination of the presence of personal data must be the potential circumstance of de-anonymization in Big Data. In particular, it is not appropriate to consider the effects of de-anonymization only with the presence of the harm. A practical solution is given in Privacy by design.

AEGIS will provide an anonymization tool supporting data publishers in removing all personal identifiers in the datasets. Highly problematic is the circumstance when the re-identification is possible after AEGIS re-uses datasets and additionally datasets generated from various data mining processes from several stakeholders[34], even if all identifiers have been removed.

[29] Enisa – Privacy in Big Data p.13
[30] Enisa – Privacy in Big Data p.13
[31] DuD, 2016, 422
[32] DuD, 2016, 422
[33] DuD, 2016, 422
[34] Which means more types of data is present in the data storage.

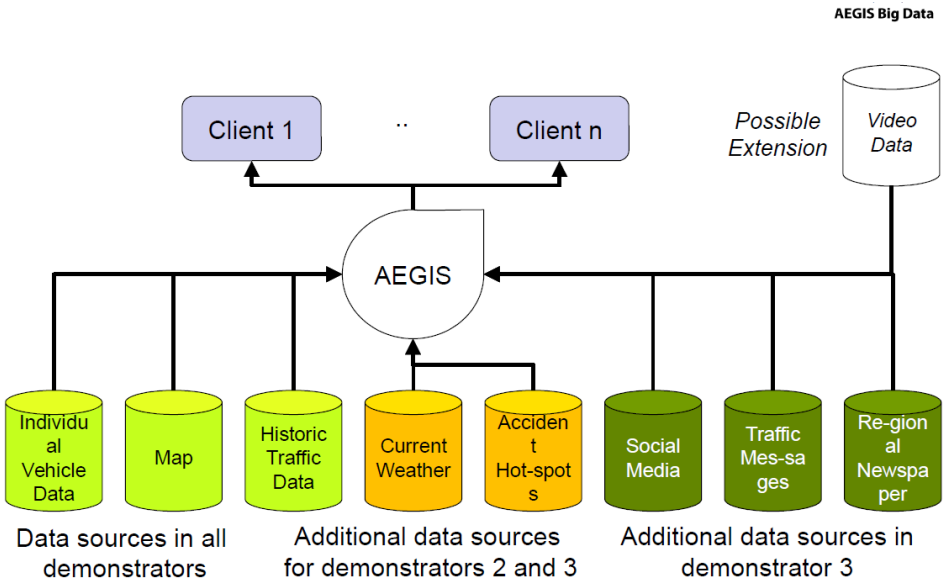

Figure 5-1: Data sources - from D1.2 p. 69

Sweeney´s study, already conducted in the year 2002 holding the observation that "*87% of the U.S. population is uniquely identified by date of birth, gender, postal code*"[35] and data protections affairs like *AOL*[36] are reason to identify two main problems:

1. The disclosure of personal information within the massive amount of anonymized data.
2. The application of data privacy regulations.

5.3.4.3. Relevant key principles for data protection

The application of the data protection principles has to be complied when processing personal data. The data protection principles constitute the fundament of data privacy by preventing arbitrariness and irresponsibility and granting transparency and flexibility and accountability by binding the data controller on law and every data process to particular requirements. The necessity to comply with these principles in AEGIS should already be considered during the establishing process, concretely in the design process respectively during and after the data process.

I. Lawful, fairly and transparent data process

The data process is bound to the data protection principles, which represent the fundament on which the data process has to be executed. In cases of violating or insufficiently applying these

[35] https://en.wikipedia.org/wiki/Latanya_Sweeney
[36] Unintentional disclosure of costumer data, see: https://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/

principles, pursuant to article 83 sec. 5 lit. (a) fines in the amount of 4 percent of the global revenue can be imposed.

The data protection principles can be found in article 5 GDPR and consist of the following integral parts:

Lawful

The principle of lawful that processing data is only permitted with presence of a sufficient legal basis.[37] In case of the European legislation, according to Article 6 sec. 1 lit. a GDPR legal basis can arise from gaining consent or from other reasons of article 6. GDPR[38].

Fairly

A fair data process requires that the data subject is aware of the data process and capable of understanding the conditions and procedures of the data process. Processing personal data needs to be required for the realization of a legitimate purpose[39]. Along with the principle of data minimization, the way of processing data has to be proportionally chosen to realize the determined purpose, which means that the single steps of the data process with the amount of data are really needed[40]. The data processor obligated of fairly weighing his own interests with the interest of the person concerned, is required to be aware of his own responsibility to protect and secure personal data in complex analysis procedures in Big Data.

Especially in Big Data technologies, the aforementioned conditions are highly considerable. As the data subject needs to be aware of the data process, goes along with the principle of transparency. Beyond, there are intersections with gaining an appropriate consent, as it is required that the consent declaration which comprises which data are needed for which steps in the data process, has to be fairly written. This means that formal design of this declaration complies an easy and transparent language, so that, in turn, all data required for the data process that are not included in the respective consent declaration are excluded from a lawful data process as these data are not are in included in the data subject´s interest, respectively an interest of processing these personal data has not been expressed[41].

Weighing the own interest with those of data person concerned, especially intends to avoid misuse of data, intentionally and not intentionally. As Big Data analysis predict certain actions and circumstances, like the preferences of someone´s buying behavior on online markets (e.g. amazon), not intentional and privacy interfering outcomes of Big Data analysis could be revealed.

[37] Kühling/Büchner Art. 4 Rn. 8
[38] Article 6 sec. 1 - necessary for: (lit. b) the performance of a contract, (lit. c) for compliance with a legal obligation, (lit. d) in order to protect vital interests, (lit. e) the performance of a task in public interest, (lit. f) legitimate interest.
[39] Wybitul, Handbuch DS-GVO Introduction Rn. 67
[40] Wybitul, Handbuch DS-GVO Introduction Rn. 67
[41] Wybitul, Handbuch DS-GVO Introduction Rn. 67

Exemplary[42], a father received Target's marketing offers based on the daughter´s consumption patterns in that store. The father assuming his daughter not be pregnant, complained about those advertisements at the supermarket. Afterwards he talked to his daughter, who revealed her pregnancy. As result, the father knew about his daughter´s pregnancy before the daughter could have told him. Although this is not an unlawful conducted data analysis, but an analysis based on a legal basis, there are several questions arising from this story. Does the father resp. family to anticipate these kinds of advertisements or was there a lack of information about the data process and the existence of own rights? Did the supermarket assume, based weighing own interest with expected interest of the family, what personal data shall be processed?

Transparency and responsibility

With respect to the principle of fairness, transparency comprises that information are formulated in such a way, that the data subject is capable of comprehensively understanding the what amount and in which way his personal information is processed, from whom and for what purpose.[43] This enables him to understand and recognize at which stage personal data is collected or in case of interacting with a online service od mobile app, which function triggers the collection of personal data[44]. Transparency is seen as important condition in order to autonomously control the handling of own personal data, which is necessary for executing the rights given in this regulation realizing informational self-determination[45]. The way of formulating the data process is highly important for the knowledge basis of the data subject, as he needs to be capable of informing about the processing of his personal data and therewith to agree with the data process[46]. Concrete obligations derived from the transparency requirement are found in the information requirements of the data controller, Article 12 – 14 GDPR, as well as in the information rights of the data subject, Article 15 DSGVO[47]. In order to properly inform the data subject and ensure his awareness about the data process is a comprehensively and easy written consent declaration whose approval is requirement of the data process[48]. The requirements of gaining consent is explicitly formulated in Article 7 sec. 2 as "*the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language*." The main criteria are availability, readability, comprehensiveness and clarity. Illustrative problems making this principle necessary are cases where the data subject is overloaded with too large consent declaration filling multiple PDF-pages, which results that the data subject accepts the data process without reading it. Another problem are too complicated formulations, mainly in high-flown juristic language, not

[42] See: http://www.wiwo.de/unternehmen/it/digitale-revolution-der-wirtschaft/algorithmen-was-heute-schon-geht/7865208-2.html
[43] Wybitul, Handbuch DS-GVO Introduction Rn. 67
[44] Wybitul, Handbuch DS-GVO Introduction Rn. 67
[45] Kühling/Büchner Art. 5 Rn. 1
[46] Kühling/Büchner Art. 5 Rn. 1
[47] Wybitul, Handbuch DS-GVO Art. 5 Rn. 9
[48] Wybitul, Handbuch DS-GVO Art. 5 Rn. 9

understandable for the majority of people. The issue of consent requirements a more detailed discussed in the consent requirement section.

The term *responsible person* is defined in Article 4 Nb. 7 GDPR as "*the natural or legal person, […] which, alone or jointly with others, determines the purposes and means of the processing of personal data*". The assignment of responsibility within data protection is important, as the subject of this reference is confronted with a bunch of rules within this regulation, for those compliance he is responsible for[49]. The data controller is reference point and for the rights and obligations granted by this regulation, receiver public measures and fines according to Article 82 and 83, as well as the technical and organisation measures and procedures have to be met in order to properly process personal data pursuant this regulation[50]. Beyond that, the data controller bears the burden of proof in legal disputes[51].

With regard to AEGIS, it should be distinguished between two groups. There are the developer group providing AEGIS and the stakeholder group using AEGIS. The different roles of developers and stakeholders imply a different degree of responsibility and therewith a different obligation for assessing particular risks. The stakeholder-groups collect and process data, they are liable for infringements against the general data protection regulation, whereas AEGIS developers only provide the Big Data application which lead to an overall risk assessment of Big Data Linked Data applications. As result, AEGIS developer must recognize an indirect responsibility deriving from Article 25 sec. I GDPR to ensure technological and organizational measures regarding the architecture of AEGIS, to increase the privacy efforts of the AEGIS stakeholders[52].

Also corresponding with this principle are documentation issues about processing procedures. The necessity arises from the fact, that the data controller has to guarantee and prove suitable data protection measures, documentation obligations relevant for the burden of proof[53]. Suitable data protection measures aim to provide an appropriate level of security and protection of personal data within the data process including the prevention of loss and harm of personal data as well as guaranteeing a lawful data process[54]. For ensuring compliance in this context and enhancing the relationship of trust, the responsible person should evaluate the kind, amount, conditions and purpose of the data process to estimate possible impacts for the rights and interest of the data subject[55]. The function of a risk assessment is not only to document and prove the compliance of obligations pursuant to this regulation, but to implement appropriate technical measures necessary for Article 25 GDPR in advance[56]. What criteria an assessment comprises, has to be considered by the data protection officer of the respective organisation. In the section

[49] Kühling/Büchner Art. 13 Rn. 93
[50] Kühling/Büchner Art. 13 Rn. 94
[51] Kühling/Büchner Art. 13 Rn. 94
[52] Kühling/Buchner DS-GVO Art. 25 Rn. 13
[53] Wybitul, Handbuch DS-GVO Art 24 Rn. 7
[54] Wybitul, Handbuch DS-GVO Art 24 Rn. 8
[55] Wybitul, Handbuch DS-GVO Art 24 Rn. 8
[56] Wybitul, Handbuch DS-GVO Art 24 Rn. 14

Technical and organisation measures, offers particular criteria and implementation strategies necessary for a proper risk assessment and methods to promote the creation of solutions which can promote compliance strategies[57].

II. Data minimization

According to Article 5 I lit. b S. 1 GDPR, data minimization requires that data are only processed due to a legitimate purpose at the time of collection. The data process shall be restricted to the minimum amount necessary to fulfil the pursuit purpose of the data process. According to this definition the data process needs to be appropriate, substantial and restricted to the minimum amount necessary.

Pursuant to the common Big Data definition, the so called "V- characteristics" abstractly reduce Big Data its capability of *Volume, Variety and Velocity*[58]. One aspect why Big Data is so attractive, is the capability of storing and processing a vast amount of data within a decentralized storage system. This attractiveness leads to the motivation, the more data can be stored and processed in Big Data, the more insights and solutions can be given to a particular question. This outlines the contrast of Big Data and data protection insofar, that Big Data intends *data maximization*, whereas data protection regulations demand the compliance of *data minimization.* The scale of data minimization is strictly dependent on the respective purpose.

III. Purpose limitation principle

Along with Article 8 CFR, the use of personal data must be determined for a purpose[59]. The purpose limitation principle states that processing personal data is only allowed/lawful with proof of an explicitly determined and lawful purpose adequate and relevant for one particular data process[60]. Important is a relation between the means and the purpose based on the weighing of the contrary interests of both parties[61]. The purpose must be explicit determined as well as communicated to the person concerned[62]. Not forbidden is the use of a purpose acting as umbrella purpose under which a number of separate processing operations are summarized[63]. This enables the data controller to collect data for multiple purposes. Nevertheless, all the other "sub-purposes" related to the main purpose have to be separately specified enough and appropriately described and the person concerned to be informed, additionally to decide suitable safe-guard measures and to ensure a certain compliance level[64].

Legitimate purpose

[57] WP-29 DPIA p. 13
[58] ZD 2017, 226
[59] Ehmann/Selmayr, DS-GVO Art. 5 11
[60] ZD 2017, 226
[61] 21Wybitul, Handbuch DS-GVO Introduction Rn. 70
[62] 21Wybitul, Handbuch DS-GVO Introduction Rn. 70
[63] WP-29 purpose limitation p 16
[64] WP-29 purpose limitation p.16

Equivalent to Article 6, the purpose has to be legitimate, which means that the data process demands a legal basis and the purpose may not violate against any legal norm. This has to be ensured during all stages of the data process at any time, and additionally in accordance with other all forms of applicable laws[65]. Besides the pure legislation, it is advisable to apply intern police regulations or codes of ethics or even contractual agreements in the purpose determination[66].

Specific purpose

The purpose must be explicitly defined. The data controller has to carefully consider what purpose is suitable and are actually needed[67]. Helpful therefore should be an internal assessment of the data controller, particularly with help of the data protection officer, what purposes can be identified, which is a necessary condition for his accountability[68]. For Big Data applications relevant is the question of the granularity of the purpose. Required is a formulation providing the kind of processing respectively the methods of the data process, so that along with the principle of transparency the person concerned has the chance to understand the use of his personal data[69]. Vague wording purpose formulations would thwart the sense of data minimization[70]. The obligation to avoid unprecise purposes and too short or overloading descriptions intend to avoid that, the subject is reduced to a pure object of the controller´s action as result of an arbitrariness data processing. Insofar the purpose must be explaining and the formulation transparent enough for the person concerned to understand and control the data process[71].

IV. Compatible further processing

Nevertheless, it is not excluded to process data on the basis of different purposes. The most straightforward possibility is to gain the respective consent for further processing personal data[72].Besides, according to 6 sec. 4 GDPR, processing data for a purpose other than the original purpose requires to be compatible, and vice versa not incompatible, with the original purpose. With regard to the wording of article 6 sec. 4 GDPR, "further processing implies that subject is the extension of the current data process and not a new data process independent of the previous data processes[73]. A new data process requires a new legal basis, whereas the extension of the current data process does not require a new legal basis, but a reasonable justification according to article 6 sec. 4 GDPR[74]. Decisive criterion is compatibility. Insofar, the further processing of

[65] WP-29 purpose limitation p. 20
[66] WP-29 purpose limitation p. 20
[67] WP-29 purpose limitation p.15
[68] WP-29 purpose limitation p.15
[69] WP-29 purpose limitation p. 15
[70] E.g. Future research, marketing purpose, improving user experience etc.
[71] ZD 2017, 226
[72] ZD 2016, 507.
[73] ZD 2016, 510.
[74] Recital 50 of the GDPR: „[…] no legal basis separate from that which allowed the collection of the personal data is required.": https://gdpr-info.eu/recitals/no-50/

data is not restricted to the pure compatibility, but the decisive criterion whether data process derives from the original purpose[75].

Processing personal data on basis of article 6 sec. 4 GDPR requires the data controller to inform the data subject appropriately with respect to article 13 sec. 3 GDPR.

V. Accuracy

Data quality is an interest of the industry in the first place. This includes the process of valid, latest and correct datasets.

Accuracy according to article 5 sec. 1 lit. (d) GDPR means that personal data have to be objectively correct and if necessary to be updated. Thereby objectively correct means that all information about a person have to match with reality. This principle only refers to objective circumstances and can consequently only to provable facts but not for subject value judgments[76].

The accuracy of data has to be guaranteed with regard to the respective purpose, which means that if the purpose of the data process does not require data to be updated, this principle does not apply in this manner. For example, when using health data for determine insurance fees or using datasets for scoring purposes in credit checks, updated and accurate datasets are necessarily required for the data process[77]. The relevance of this principle arises from the fact that violations against data protection principles result in *"administrative fines up to 20 000 000 EUR, or in the case of an undertaking, up to 4 % of the total worldwide annual"* pursuant to article 83 sec. V lit. a GDPR.

For those data processes not necessarily requiring accurate datasets, the data subject needs to be aware of this fact. In case of inaccurate datasets, the data subject has the right of correction or erasure of wrong datasets, according to article 16 GDPR. Furthermore, with claiming the incorrectness of data pursuant to article 18 GDPR, the data process has to be restricted.

The data accuracy principle is interrelated with the Data relevancy principle, stating that personal data processed have to be adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed.

VI. Restriction of storing data

In order to avoid long term storing, personal data have to collected and stored as long as it is necessary for the respective purpose. This constitutes a time limit for storing personal data. Whenever the purpose is achieved, all personal data have to be erased from the data storage under particular circumstances[78]. This is necessary as binding (personal) data to a particular purpose guarantees transparency for the data subject and avoids arbitrariness of the data controller to use data afterwards to a not specified purpose. Another possibility is to clean the data from all

[75] Kühling/Buchner Art. 5 Rn. 38
[76] Kühling/Büchner Art. 5 Rn. 57
[77] ZD 2016, 459.
[78] With respect to the data subject´s desire and concerns - Kühling/Büchner DS-GVO Art. 5 Rn. 61

personal relations. If datasets have been anonymized and cannot be re-identified, erasing is not necessary.

VII. Integrity and confidentiality

The data controller has to guarantee the safety and security of personal data during the data process. Thereby, he is fully responsible and so, according to this principle, obligated to implement suitable technical and organizational measures in order to prevent unintentional harm of personal data. There are two kinds of harm relevant for complying with this article. The first meaning refers to unlawful processing of personal data by a third person that is not attributed as responsible person of the data process or does not have any other legal basis for this kind of action[79]. Another meaning of harm occurs when data is destroyed or unintentionally damaged with the consequence that the damaged datasets are not usable enough in order to suffice the pursuit purpose[80]. This is a critical aspect whenever a contractual relation exists or the data subject is legally dependent on the correctness of the result of the process.

The data controller has to consider the decentralized storage in Big Data. With regard to the CAP theorem, when working with different storage nodes, working with "updated" datasets resulting in consistency should be combined with the availability of the datasets in the single nodes.

VIII. Automated processing - Profiling and automated decision-making

According to the definition of Article 4 number 4 GDPR, profiling "*means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements*".

The considerations concern the aspect of digital profiles of individuals that are used in an increasing number of sector, private and public.[81] The obvious advantages[82] of automated decisions of digital profiles are obvious, but significantly affects the individuals' rights and freedoms[83]. The general data protection regulation introduces explicit regulations that address the risk of Profiling and automated decision-making in Article 22 and Article 15 sec. 1 lit. (h) GDPR.

<u>Admissibility according to article 22 GDPR</u>

Subject of article 22 GDPR concerns the result of automated processing, whereas the process itself has to comply with the data protection principles in general.[84] Analogues to the process of

[79] Kühling/Buchner Art. 5 Rn. 74
[80] Kühling/Buchner Art. 5 Rn. 75
[81] E.g. Banking and finance, healthcare, taxation, insurance, marketing and advertising - WP-29 decision making p. 5
[82] Faster with less effort - WP-29 decision making p. 5
[83] WP-29 decision making p. 5
[84] Kühling/Buchner Art. 22 Rn. 11

person data, it is generally prohibited to make decisions solely based on automated processing. Insofar, article 22 sec. 1 requires that decisions may not be based only on automated processing. The restriction to decisions based *"solely"* on automated processing means that the decision does not contain any human valuation or action decisive for the decision[85]. The key argument of the general exclusion intends to avoid situations, where the individual is degraded to the pure object of algorithm and relevant decisions are no longer in the hand of human action and valuation[86]. In contrast to that, excluded are those cases, where automated processing supports the decision making and is not the exclusive basis of the decision. A legal relevance occurs whenever the legal position of an individual changes in a way, constituting or revoking a right or legal relationship[87]. The assignment of legal relevance is thereby independent of a positive or negative result[88]. A significant effect can be assumed, when the individual is sustainably disrupted in his economic or personal development[89]. Pursuant to article 22 sec. 2 GDPR, there exist three exceptions relativizing the general prohibition.

Automated decision-making is permitted, when it is necessary for entering into, or performance of, a contract between the data subject and a data controller. The use of automated processing has to be proportional and appropriate safeguards in favour of the data subject should exist[90].

(*Necessary)* Whether the decision is necessary for entering and performing a contract, depends on presence of a direct context between automated processing the purpose of the contractual obligation[91]. Consider the case of using scoring procedures to calculate the creditworthiness of the individual for the purpose of granting a credit agreement, there exists a context between the basis of a calculation and a contractual obligation[92].

(Adequate) Assuming the presence of a necessary automated processing, the admissibility depends pursuant to section 3 of article 22 GDPR on the condition, whether appropriate measures have been implemented guaranteeing the data subject rights and interest[93]. As result, the data subject shall be able to interfere with the automated processing by requesting information and challenge the decision[94]. Relevant for implementation of appropriate safeguards is obligation to inform according to article 13 and 14 GDPR.

[85] The human action did not affect the decision in any way - Kühling/Buchner Art. 22 Rn. 15
[86] Kühling/Buchner Art. 22 Rn. 11
[87] Kühling/Buchner Art. 22 Rn. 24
[88] Kühling/Buchner Art. 22 Rn. 25
[89] E.g. Termination of a credit, increasing rates, refusal of a public approval - Kühling/Buchner Art. 22 Rn. 26
[90] Kühling/Buchner Art. 22 Rn. 29
[91] Kühling/Buchner Art. 22 Rn. 31
[92] Kühling/Buchner Art. 22 Rn. 31
[93] Kühling/Buchner Art. 22 Rn. 31
[94] Kühling/Buchner Art. 22 Rn. 31

According to the second exception, automated processing is allowed by Union or Member State law[95]. Thereby[96] addressed are purposes regarding the public interest like fraud and tax-evasion monitoring and prevention.

At last, automated processing is allowed with gaining consent. The conditions necessary correspond to the general conditions of article 6 GDPR – outlined in the section below. The data controller has to the requirements of consent and to provide appropriate safeguards and to ensure a fair and transparent (automated) data process, granting the chance of interfere with the result of and to force the data controller to take position to the data process[97].

IX. Privacy by design and by default

Article 25 of the Regulation refers to the principles of "Data protection by design and by default", expressly stating that, considering the set of circumstances, the controller shall implement appropriate technical and organisational measures:

- "such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects";
- "for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility".

Further considerations on Privacy by Design and Privacy by Depafult paradigms are in Section 6.5.1.1.

X. Accountability

Finally, it is important to mention here also the principle of accountability, which, besides requiring the active implementation of measures by controllers to promote and safeguard data protection in their processing activities, requires that the data controllers should be able at any time to demonstrate compliance with data protection provisions to data subjects, to the general public and to supervisory authorities.

5.3.4.4. Key figures of data processing: the data Controller and the Processor

Article 2 of the Directive lingers over two key figures of data processing:

- Data Controller: "natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of

[95] Only by providing *"suitable measures to safeguard the data subject's rights and freedoms and legitimate interests"*.
[96] Recital 71 of the GDPR - https://gdpr-info.eu/recitals/no-71/
[97] Kühling/Buchner Art. 22 Rn. 43

personal data…". Therefore the data Controller may be a natural person or a legal entity, of both public and private nature. With regard to the same data processing, it is possible to have one or more data controllers.
- Processor: "natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller".

The whole chapter IV of the Regulation pertains to "Controller and processor" (Article 24 ss.), regulating:

- general obligations concerning, among other, the responsibility of the controller and the role of the processor, data protection by design and by default principle and related controller's duty, the case of joint controllers, the authority of the controller and of the processor, the record of processing and the cooperation with the supervisory authority (Section 1);
- the security of personal data, including the security of the processing, the notification of breach to the supervising authority and the communication of the same to the data subject (Section 2);
- data protection impact assessment and prior consultation (Section 3);
- the figure of the data protection officer, including his designation, position and tasks (Section 4);
- the codes of conduct and certification (Section 5).

5.3.4.5. National law applicable

The Regulation, unlike the Directive, "shall be binding in its entirety and directly applicable in all Member States".

Par. 6.3.6 provides an overview of the regulatory framework implementing European privacy and data protection legislations respectively in Italy, Austria and Greece, where the use cases and the demonstrator will be located. These national provisions were adopted pursuant to the Directive repealed by GDPR. Therefore, though they can still be considered in the regulatory framework relevant to AEGIS, but, in case of conflict, as a general rule (with exceptions for instance as regards constitutional laws), the Regulation prevails.

5.3.4.6. The notification to the National Data Protection Body (NDPB) and the procedure of "prior checking"

Recital 89 of GDPR states that the indiscriminate general notification obligations, provided by the Directive 95/46/EC, should be "replaced by effective procedures and mechanisms which focus instead on those types of processing operations which are likely to result in a high risk to the rights and freedoms of natural persons by virtue of their nature, scope, context and purposes. Such types of processing operations may be those which in, particular, involve using new technologies, or are of a new kind and where no data protection impact assessment has been carried out before by the controller, or where they become necessary in the light of the time that has elapsed since the initial processing".

The implementation modalities of the notification/authorisation/declaration of compliance

requirement vary from country to country, for instance as to the means through which it has to be filed, the cases in which it is due and the amount of information to be provided. As regard AEGIS demonstrators, details on the notification procedures and bodies are detailed in par. 5.4.2.

In accordance with the aim of suppressing the "indiscriminate general notification obligations" (Recitals 89 mentioned hereabove), the Regulation introduced Article 36: "the controller shall consult the supervisory authority prior to processing where a data protection impact assessment… indicates that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk".

5.3.4.7. Information to be provided by the Controller to the data subject

Transparency of the data processing towards the data subject is one of the most important principle to be fulfilled when collecting and processing personal data, in AEGIS too. The corresponding obligation to inform the data subjects cannot be exempted under national legislation, save for very limited circumstances, including the case that compliance with this information obligation results is impossible or requires a disproportionate effort for the Controller.

There is a list of minimum mandatory information to be given to the data subject, including: the purposes of the data processing; the categories of the data involved in the processing; the list of recipients (or of the categories of recipients) of data communications; data subject's right to access his/her personal data and to rectify them; the identity of the Controller and, if applicable, of his representative. In case the Controller intends to share personal data with third parties, the mandatory information must be given to the data subject no later than when such communication occurs.

GDPR dedicates Chapter III to the "Rights of the data subjects", describing in detail transparency and its modalities, information and access to personal data. Article 13 lists the information to be provided where personal data are collected from the data subject.

5.3.4.8. Special categories of processing: sensitive data and judicial data

The "special categories" of data, earning a higher degree of protection are:

- **Sensitive data**: "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life";
- **Judicial data**: data related to "offences, criminal convictions or security measures".

The lists are mandatory and closed: a personal data may not be considered as sensitive or judicial if it is not comprised within them. In case data processing performed on this kind of data, stricter requirements have to be fulfilled by the Controller and specific precautions are established. The analysis of them is outside the scope of this deliverable.

The issue is addressed also by Articles 9 and 10.

5.3.4.9. Criteria for data processing legitimacy

The frame for lawful data process is constituted in Article 8 sec. 2 of the Charter of Fundamental Rights of the European Union, which states that "*data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law".* Bounded to a fair and purpose limited data process, the data protection regulations requires to adopt the principle of process data only with gaining consent or another legitimate basis. Simultaneously, gaining consent has been highlighted in Article 8 CFR as it is seen as most important expression for informational self-determination[98]. The concerned person is capable to determine whether his own data may be proceeded and in which way[99]. Consequently, Art 6 GDPR contains the possibilities for the lawfulness of processing, whereby systematically given consent to the processing is mentioned at the top in lit. (a). Moreover, data is lawfully processed due to contractual obligations, whereby the data process is necessary for the performance of the contract.

According to article 6 GDPR sec. 1 other reasons for lawful process data are:

- Compliance issues for legal obligations or in order to protect the vital interests of the data subject, lit. (c), (d)
- For the performance of a task carried out in the public interest or in the exercise of official authority, lit. (e)
- Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, lit. (f)

I. Consent and its requirements

Gaining consent is seen as the realization of self-determination, as the individual autonomously decides about what personal data can be collected[100]. Insofar, AEGIS Stakeholder shall primarily attempt to gain consent when working with personal data.

The conditions deriving from Article 8 CFR and underlining its importance for self-determination are required for gaining valid consent. The conditions of the consent requirements also comprise several principles of data processing like the purpose limitation or transparency issues. These principles are laid down in Article 6 and 7 GDPR. Article 6 states that "*the data subject has given consent to the processing of his or her personal data for one or more specific purposes*". On basis of the consent requirements, Article 7 GDPR lists certain conditions that have to be fulfilled within the framework of consent. The conditions are listed in Article 7 sec. 1 GDPR which state that the responsible data processor is in charge of the burden of proof, specifically he "*shall be able to demonstrate that the data subject has consented to processing of his or her personal data*". This shall ensure that no doubts exist by the person concerned.

[98] Ehmann/Selmayr, DS-GVO Art. 6 Rn. 5
[99] Kühling/Buchner Art. 6 Rn. 18
[100] Wybitul, Handbuch DS-GVO Art. 7/8 Rn. 3

The consequence of not fulfilling these conditions is that the declaration of the person concerned cannot be considered with legal effect for a lawful data process[101]. The following conditions should be considered by AEGIS stakeholders when requesting consent.

Transparency and formal requirements

The formal and textual requirements are summarized in one sentences stating that *"the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language."* Transparent requires that the declaration shall be perceptible enough and highlighted to be distinguishable from other matters[102]. The declaration shall not be overloaded with unnecessary and verbose explanations intending to overstrain the data subject[103]. Especially domain specific terms should be avoided. The challenge is to identify an appropriate linguistic mark at which it is guaranteed that the declaration is understandable for a broad mass of people. It often practically happens that people agree with a declaration of consent without having read the content before[104]. Self-determination requires the person to be informed which is obstructive when the respective declaration cannot be understood. Particularly in Big Data applications including high complex algorithm procedures, the tension between a clear and easy understandable language and the informational content is extremely high. Moreover, the technology-neutral compromise sets no scale for an appropriate declaration that suffices this linguistic requirement. Presumably with having the burden of proof, it is task of the data controller to solve this gap by finding an appropriate language for the declaration.

Voluntariness

As there is no unique definition of voluntariness, Article 7 sec. 4 GDPR names four criteria – additionally to the already mentioned criteria – that can be used to determine whether the declaration of consent has voluntarily been given. First, it is necessary to determine whether an unequal situation exists between the data controller and the person concerned[105]. This is especially the case when the data controller represents a public authority or a powerful company. It has to be avoided that this imbalance is misused with the consequence that the declaration of consent has not willingly be given or is coupled by with contractual obligation. A certain imbalance is not reason for an invalid consent, but a hint for the need of protection in favour of the consumer/single person together with situationally circumstances[106]. To decide whether consent has given voluntary, the following criteria could be taken into account: The contractual purpose in accordance to the party´s interests, as well as the means and circumstances of the data process that are used to fulfil the contractual performance. Voluntariness is an important part of self-determination, especially when the data controller service or product is seen as valuable

[101] Kühling/Buchner Art. 7 Rn. 21
[102] Kühling/Buchner Art. 7 Rn. 25
[103] Kühling/Buchner Art. 7 Rn. 25
[104] MMR 2014, 363
[105] Kühling/Buchner Art.7 Rn. 42
[106] Kühling/Buchner Art. 7 Rn. 44

society (e.g. services like amazon, WhatsApp or digital media). Consequently, unnecessary collection of personal data must be avoided and necessary personal data relevant for the purpose should be determined in the first place - in accordance with the principle of purpose limitation[107].

For the formal realization, the consent declaration should not only be perceptible with formal hints, but agreed with an "unambiguous indication of the data subject's agreement"[108]. An active participation in sense of an opt-in solution should substitute an already marked consent declaration in order to guarantee that the person concerned wilfully agreed in the process of his/her personal data. The European commission exemplary listed "ticking a box, choosing technical settings" as proper suggestion. Furthermore, combining a consent agreement for a contract with other agreements as "combined solution" is invalid, e.g. ticking the box "accepting the general terms and conditions" triggers ticking the box for giving consent.

Certainty

Combined with the principle of purpose limitation, the data processor has to formulate a legitimate, appropriate and unambiguous declaration that states for which purpose data are collected and proceeded. In this way, the person concerned can overview the process of his data and eventually interfere in the data processors action whenever he believes the data process of his data is not comprises by the agreed purpose anymore. Consequently, a precise and well declared purpose declaration as requirement for the data subject to be informed.

One exception applies for scientific research, as the commission allows that "the data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research". With this "broad consent", the GDPR allows a deviation from the basically required informed consent. Questionable is, how to differ between scientific research and not scientific research?

Knowledge and awareness

In order to reasonably agree the declaration of consent, the person concerned necessarily needs to be properly informed. This includes to be aware of the legal scope and the consequences of the data process[109]. With gaining consent, the data controller should prove that the data subject is aware of what data are collected and for which purpose[110]. The person concerned particularly needs to know the kind of data are used and the purpose of the data process in accordance to the general data protection principles as well as the identification and contact information of the data processor to begin of the data process with reference to article 13 GDPR[111]. The single obligations to inform are outlined in section 4.2.4 of this document. Being informed strongly

[107] Kühling/Buchner Art. 7 Rn. 46
[108] Recital (32 of the) GDPR: „statement or conduct which clearly indicates in this context the data subject's acceptance of the proposed processing of his or her personal data." - http://www.privacy-regulation.eu/en/recital-32-GDPR.htm
[109] Kühling/Buchner Art.7 Rn. 59
[110] Wybitul, Handbuch DS-GVO Introduction Rn. 287
[111] Kühling/Buchner Art. 7 Rn. 59

depends on the comprehensibility of the declaration, in what degree of abstraction the data process has been declared and the individual skills of the person concerned. Inherently, most Big Data architectures are highly complex and consists of several procedures all using (personal) data differently. Consequently, it could be necessary to inform the person concerned about the existence of all procedures respectively about all different stages where (personal) data is proceeded[112]. Especially in contractual relations, the data processor reps. contract party should take these criteria into account in order to effectively seek consent.

Checklist for consent

The guideline in annex 1 proposes particular criteria necessary for the consent requirements[113]. These criteria should be assessed in relation to the kind and the degree of the data process and its circumstances[114].

II. Other legal bases

The legal basis for processing data for performing contractual obligations is regulated in article 5 sec. 1 lit. c GDPR as "processing is necessary for compliance with a legal obligation to which the controller is subject". Gaining consent in such cases shall not be necessary, as the performance of a contract is in the interest of the data subject[115]. Ensuring the data protection principles shall ensure that the data process is necessary and limited as well as proportional to the performance of the contractual obligation preventing arbitrariness of the data controller[116].

Besides contractual obligations, a legal basis can derive from contractual relations, as one party has to guarantee particular protection obligations which include to prevent any harm for the legal interest of the other party. For example, an E-Mail provider can filter spam and fraud in E-Mails with appropriate fraud detection in Big Data Applications[117]. With regard of the increasing harm of cyber-criminality, e.g. credit card fraud, phishing or data theft, the interest of protection in a legal relation is more and more reasonably necessary[118].

In contrast to processing data within a legal relationship, the process of data can be lawful with presenting a legitimate interest by the data controller. The formulation is undetermined increasing legal uncertainty. The level of protection, especially in gaining consent, shall not be underpinned by this kind of legal basis[119]. Typical cases[120] are Business Intelligence Solution, relevant to create market analysis or to optimize strategic procedures whereby the basis is the process of

[112] ZD 2017, 213
[113] This overview was taken from: Wybitul, Handbuch DS-GVO Introduction Rn. 296
[114] Wybitul, Handbuch DS-GVO Introduction Rn. 296
[115] Wybitul, Handbuch DS-GVO Art. 6 Rn. 14
[116] Wybitul, Handbuch DS-GVO Art. 6 Rn. 15
[117] NJW 2014, 2985
[118] NJW 2014, 2985
[119] Wybitul, Handbuch DS-GVO Art. 6 Rn. 32
[120] Others are processing data for improving internal administration, Guaranteeing network,- and cybersecurity etc. - Wybitul, Handbuch DS-GVO Art. 6, Rn. 38/39

customer data[121]. The interest is based on economic interest in form of optimizing own products, accessing new customer groups or target-oriented advertisement[122].

With presence of a legitimate interest, the data process has to be objectively[123] necessary in the first place. Explicitly required is a careful weighing of both interest – the data subject and the data controller – as well as the basic rights of the data subject[124]. Only in case the data controller´s interest overweigh, the data process is to be considered as lawful. Highly relevant at this point is to determine *overweighing interest* of the data controller. The individual case has to be assessed on basis of the **specific situation of the data process[125]** as well as on the **level of intervention[126]** pursuant the interest of the data controller[127].

Recommendation: Due to the uncertainty of this legal basis and the high level of protection from the consent, the data controller shall pursuit processing data based on the data subject´s consent.

5.3.4.10. Data subject's rights

Privacy rights may be classified into two categories:

- **The rights of information**, consisting in the data subject's right to be informed by the Controller on the purposes and conditions of the processing activities to be carried out on his personal data.
- **The rights of intervention**, allowing the data subject to ask that certain actions are performed on his data and also to interfere in the data processing.

Chapter 3 of GDPR disciplines the rights of data subjects, including transparency and its modalities (Section 1), information and access to personal data (Section 2), rectification and erasure, including the right to data portability (Section 3), the right to object and automated individual decision-making (Section 4) and applicable restrictions (Section 5).

The following snapshot represent the rights of the data subject granted by the GDPR. The execution of granted rights represents an integral part of informational self-determination and is, as second component, dependent of being aware. Insofar, essential requirement for being able to execute certain granted rights according to this regulation is that the data controller has to provide as much information as necessary to ensure that the data subject has a chance to be appropriately informed – provided in previous section. Especially the right to information granted in Article

[121] NJW 2014, 2985
[122] NJW 2014, 2985
[123] The data process has to be necessary for both parties – data controller and data subject. - Wybitul, Handbuch DS-GVO Art. 6, Rn. 46
[124] Wybitul, Handbuch DS-GVO Art. 6 Rn. 49.
[125] Estimating whether the data subject has to anticipate the process of his personal data, e.g. using an app for food deliverance results in receiving advertisement.
[126] Meaning the way and degree of affecting the data subjects interest - Wybitul, Handbuch DS-GVO Art. 6, Rn. 50.
[127] Wybitul, Handbuch DS-GVO Art. 6, Rn. 50.

15 GDPR is derived from the transparency principle of Article 5 GDPR. Data protection would have no meaning, if the person concerned is not informed about his own rights and potential infringements of them resulting from incorrect data process[128].

The issue involved within this section concerns the De-anonymization. The rights according to this regulation only apply for the handling of personal data, whose presence is condition for the application of the following explanations, so the unwanted respectively unintended presence of personal has to be adequately verified or monitored.

With regard to Article 58 sec. 2 lit. (c), (g) GDPR, the data controller has to power to control the compliance of the data subject's requests to exercise his or her rights pursuant to this Regulation. Whereas lit. (c) of this article requires a request of the data subject, lit. (g) is independent of such request, so that the supervisory authority is able to order the rectification or erasure of personal data or restriction of processing pursuant to Articles 16, 17 and 18 GDPR. Guaranteeing compliance is not only recommended for the data subject´s interest, but for the data controller´s interest of avoiding fines pursuant article 81, 82, 83 GDPR.

I. Right to information

The right to information is designed as reactive information model. This means, the data subject can request the data controller to disclosure particular information, whereas in contrast Article 13 and Article 14 are represent an active information model obligating the data controller to appropriately inform the data subject at the time of obtaining personal data[129]. In contrast to the obligation to inform by the data controller according to Article 13, 14 GDPR, the right to information grants the data subject the right to request personal information concerning his person. In order to strengthen transparency, the data controller has to disclose information according to Article 15 sec. 1 and sec.2 without preconditions[130]. Subject of this right is whether personal data is processed and which personal data are affected.

Especially with regard to section 3, the data controller should take preventive steps to provide the information formulated in annex 4 on request according to Article 15 sec. 1 GDPR:

Article 15 section 3 - Providing a copy of personal data

The data subject is allowed to request a copy of all his personal data that is processed by the data controller. Insofar, subject of this entitlement are all personal data – implying completeness - that are present at the data controller´s storage, which requires that personal data is provided in unmodified condition[131]. The requested personal data are limited to the rights and interest of third persons pursuant to article 15 sec. 4 GDPR and other restrictions pursuant to article 23 GDPR. Regarding the form on which data is transmitted, as long as the data subject does not explicitly request a particular medium, the data controller has to choose a common medium readable

[128] Wybitul, Handbuch DS-GVO Art. 12-15 Rn. 2
[129] Ehmann/Selmayr, DS-GVO Art. 15 Rn. 4
[130] Kühling/Buchner Art. 15 Rn. 13
[131] Kühling/Buchner Art. 15 Rn. 40

without any constraints[132]. The first copy has to be provided for free, for additional copies, the controller can demand an appropriate payment.

Recommendation: See implementation strategy in section 4.3.2

II. Right to rectification

With the right to rectification, the data subject now has more influence on the data processing. Recital 59 requires Modalities provided by the controller that facilitate the exercise of the data subject´s rights. This seems necessary, as an accurate data process cannot be guaranteed. Insofar, accurate information could still result in incorrect datasets, e.g. missing information about a person´s profile could to lead to misinterpretation and wrong decisions[133]. This is relevant for cases where decisions are based on automatically created profiles for instance for granting a credit or to set insurance costs. In such cases the person concerned is dependent of the correctness of these information.

In order to affect these decisions and interfere with wrong or event not complete datasets, the person concerned has the right to rectification and completeness. Subject of this right are incorrect data where facts are objectively not compatible with the reality which means that their inaccuracy is verifiable[134]. In contrast, subjective judgements are not verifiable and basically not included within this claim[135]. Every wrong data/information regardless of its meaning falls under this right due to the fact they could be reused for instance in Big Data analysis and thereby be essential part of a wrong results or lead to a wrong profile of the person concerned[136]. Furthermore, the reason of an incorrect data or the responsiveness of a specific person is not relevant, but only the existence of a wrong data. However, if information are correct but incomplete, legally relevant and with regard of the respective purpose of the data process incomplete, the data processor is obligated to provide necessary data[137].

III. Right to restriction and right to erasure (right to be forgotten)

Right to restriction – article 18 GDPR

The right to restriction enables the data subject to limit the data process in case of unlawfulness or the contestation of incorrectness[138]. Article 18 names four reasons enabling the data subject to request restriction:

Criteria	Meaning

[132] Kühling/Buchner Art. 15 Rn. 41
[133] E.g. in automated scoring procedures of an insurance company.
[134] Kühling/Buchner Art. 16 Rn. 8
[135] Kühling/Buchner Art. 16 Rn. 9
[136] Kühling/Buchner Art. 16 Rn. 11
[137] Kühling/Buchner Art. 16 Rn. 26
[138] Wybitul, Handbuch DS-GVO Art. 12-15 Rn. 1

The accuracy of the personal data is contested	In case the accuracy of personal data has been contested by the data subject, the data controller has to exclude the data concerned for a period enabling the controller to verify the accuracy of the personal data. After this period exhausting all possibilities for verifying the correctness: 1. data are correct and the restriction is repealed 2. correctness cannot be determined. In this case, the burden of proof bears the data controller to proof the correctness and the data subject has the right to rectification or erasure. "Non liquid" - cases: The correctness of data cannot be proofed. The principle of data accuracy applies for the data process. As consequence, as long as the data controller cannot proof the correctness of (personal) data, the data process is unlawful.
The processing is unlawful	Consequence: Further processing of (personal) data is not allowed. Nevertheless, the data controller is obligated to clarify the desire of the data subject, to determine what he intends – reason: Erasure data and restrict the data process are mutually exclusive.
Purpose of data process is no longer valid, but personal data are required by the data subject for the establishment, exercise or defence of legal claims	In case that personal data is no longer needed by data controller, the data subject has the right to restriction, whenever (personal) data is actually needed for establishing, exercising or defending a legal claim that is present or actually expected in the near future. The pure potential possibility of a legal dispute is thereby not sufficient.
The data subject has objected to processing pursuant to article 21 GDPR	As long as it is not clarified whether the appeal against the data process is reasonable, the person concerned can demand the restriction of the data process.

Concerned data is to be highlighted in a way that it is recognizable to be processed for the limited purposes only. Specific means are mentioned in recitals 67 like "temporarily moving the selected

data to another processing system, making the selected personal data unavailable to users, or temporarily removing published data from a website"[139].

Recommendation: Implement control steps making the request to restriction possible in the data process.

Right to erasure – article 17 GDPR

Generally, the data subject does not have a general right to erasure after his personal data have been collected. The content of the right to erasure and the right to be forgotten are expressed in article 17 sec. 1 and 2 GDPR. Restrictions result from article 17 sec. 3 insofar, as interest of third parties[140] have to adequately be considered and weighed with the interest of informational self-determination of the individual.

Article 17 is divided into two components. The first section of article 17 allows that the data subject is able to request the data controller to erasure his data with presence of particular conditions.

Criteria	Meaning
The purpose of the data process no longer exists	If the purpose of the data process no longer exists, the data subject can request to erasure or request to restrict data those data. Erasure personal data without knowledge of the data subject can be unlawful and result in violating the data controller´s obligation[141] to enhance the data subject´s exercise of his own rights[142].
Consent has been revoked	In case of revoking consent, a formal request to erasure personal data is not necessary. The data controller should consider the expression of the revocation regarding the range and subject – assuming the revocation does not refer to the data process as whole[143].

[139] Recital 67 of the GDPR - https://gdpr-info.eu/recitals/no-67/
[140] E.g. the interest of the business, particularly the importance of the data process which is fundamental for his business activity
[141] Pursuant to article 12 sec. 2 GDPR
[142] Kühling/Buchner Art. 17 Rn. 10
[143] Kühling/Buchner Art. 17 Rn. 11

The person concerned has appealed against the data process – pursuant to article 21 GDPR	As long as not clarified whether the appeal against the data process is reasonable, the person concerned can demand the restriction of the data process. The obligation to erasure applies unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject[144].
The personal data have to be erased for compliance with a legal obligation in Union or Member State law	For compliance with a legal obligation in Union or Member State law to which the controller is subject, specific obligations to erasure can arise, dependent on the national legal system.

Section 2 of article 17 GDPR can be considered as an extent of the section 1. The subject derives from the ECJ judgement against Google ruling the application of data protection law for search engines, additionally applicable rights of the data subject.

The aforementioned reasons for claiming the right to erasure are restricted by the exceptions of section 3 of article 17, namely in cases where: The data process is:

- for exercising the right of freedom of expression and information
- for compliance with a legal obligation which requires processing by Union or Member State law
- for reasons of the public interest or the domain of public health
- for archiving interest, scientific or historical research-purposes or statistical purposes according to Article 89 sec. 1 GDPR
- for the establishment, exercise or defense of legal claims.

In order to appropriately comply with the aforementioned provisions, it is (again) recommended to document the whole data process and to create a data retention policy ruling the handling of data.

III. Right to object

Working with false or incorrect datasets can be very harmful for individuals, especially in cases of legal decisions. Article 21 GDPR enables the right to object which means the person concerned has the right to appeal against the data process on basis of particular situations. In case of lawful data process, the legislator provides the person concerned the chance to interfere with the data process under consideration of particular circumstances of the situation[145]. The meaning of this formulation is left open and has to be interpreted to which situations it refers. So the person concerned can take into account any legal or economic reason(s) that have effect on his interest

[144] According to Art. 21 sec. 1
[145] Sydow, DS-GVO Art. 21 Rn. 53

and rights[146]. Insofar, it is the burden of proof of the person concerned to demonstrate particular circumstances that justify an appeal of the data process[147]. With reference to article 6 lit. f - whereas a legitimate interest of the data controller to lawfully process (personal) data is accepted by generally weighing the interests of both parties - article 21 intends to adjust this weighing beyond by now considering specific presented reasons of the person concerned. Within this weighing, the data controller can proof reasons regarding his interest and rights as well that outweighs the interest of the person concerned[148].

Consequence of exercising this right, is that the legal basis no longer applies for the data process, and the data controller is obligated to terminate the data process. With reference to article 17 I, c) GDPR, the data controller has the obligation to erasure (personal) data in case he cannot claim legitimate interest[149].

5.3.4.11. Obligations of the Data Controller

5.3.4.12. Obligation to inform Article 12 – 14 GDPR

I. General information obligation according to Article 12 GDPR

Pursuant to Article 12 GDPR, the data controller shall provide and communicate information in such a comprehensive manner, that the data subject can exercise the rights according to this regulation. This is explicitly required according to article 12 sec. 2, as "*the controller shall facilitate the exercise of data subject rights under articles 15 to 22*". Comprehensively providing information about the processing, risks, guarantees and rights and the method of how these rights can be properly executed is inevitably necessary requirement to comply and realize the basic principle of data sovereignty of the individual according to Article 8 CFR.[150] Insofar, article 12 GDPR realizes the transparency principle of Article 5 constituting the obligation to inform general information relating to the data process and are specified in Article 13 and 14 as well as person-specific information according to Article 15ff GDPR. With regard to article 5 section 2 GDPR, it is highly recommended to comprehensively document the processing of personal data, as the principle of this regulation requires *accountability* and therewith to be able to demonstrate compliance with this regulation. This means that companies have the burden of proof in disputes to demonstrate to not have violated against this regulation[151].

The data controller shall take suitable measures to provide all information in concise, transparent, intelligible, easily accessible form and clear and plain language. Those measures shall be proportionally to the pursuit purpose[152]. Similar to the conditions of the consent requirements are

[146] CR 2016, 93
[147] Sydow, DS-GVO Art. 21 Rn. 58
[148] Sydow, DS-GVO Art. 21 Rn. 66
[149] Sydow, DS-GVO Art. 21 Rn. 72: Helfrich: „It is questionable, whether a request of the person concerned is necessary for the data controller´s obligation to erasure."
[150] Ehmann/Selmayr, DS-GVO Art. 12 Rn. 7
[151] Wybitul, Handbuch DS-GVO Introduction, Rn. 103
[152] Wybitul, Handbuch DS-GVO Art. 12-15 Rn. 10

the formal and language requirements. Different is the way of accessing these information, which have to be easy accessible. The common way is to clearly highlight the path or providing an URL to the respective electronic document or to set a reference in E-Mails or websites[153].

II. Information obligations pursuant to article 13 GDPR

At the time of collection, the data controller shall provide the following information:

Criteria	Meaning
▪ Identity and contact details	The data controller has to provide information concerning his name, identity and other contact details in such a comprehensive manner, that the data subject is able to unproblematically establish contact with the data controller.[154]
▪ Purposes and legal base	The data controller shall provide information concerning the purpose of the personal data to be collected. These information shall be detailed and complete enough in a way that the data subject recognizes what data is collected for what purpose.[155] Additionally, the data controller shall inform about the legal basis on which personal data has been collected. In case of collecting personal data according to article 6 sec. 1it. (f), the legitimate interest pursued justifying the handling of personal data shall be represented.

[153] Wybitul, Handbuch DS-GVO Art. 12-15 Rn. 10
[154] Kühling/Buchner Art. 13 Rn. 22
[155] Kühling/Buchner Art. 13 Rn. 26

▪ The recipients or categories of recipients	The data controller shall provide information about the recipients[156] which comprises contracting data processing and the flow of data within different business units. This information include expected recipients or specific categories of recipients, whereas the best detailed information has to be provided.[157]
▪ Transfer of personal data to third country or international organisation	In case of transfer data to recipients in a third country or international organisation, the data controller shall explicitly inform and hint about a(n) (expected) data transfer outside the EU including information about the risk estimation and, if possible, objections against the transfer.[158]
▪ Period for which the personal data will be stored	Providing information about the period for which personal data will be stored comprises either a specific date or the obligation to name according to which the period is measured.[159]
▪ Reference to the data subject´s rights	The data controller shall inform about the rights of the data subject including a brief description in a more general form: • Article 15: Right to access • Article 16: Right to rectification • Article 17: Right to erasure • Article 18: Right to restriction • Article 20: Right to data portability • Article 21: Right to object

[156] Pursuant to article 4 nb. 9 GDPR: Recipient "*means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not."*
[157] In this way excluding choosing between several options for the benefit of the data controller. – Kühling/Buchner Art. 13 Rn. 30
[158] Kühling/Buchner Art. 13 Rn. 34
[159] Kühling/Buchner Art. 13 Rn. 34

▪ Whether the provision of personal data is a statutory or contractual requirement	In case that the data controller is obligated to contribute to the collection of personal in order to not getting confronted with legal consequences, the data controller has to inform the data subject about this circumstance.[160]
▪ The use of automated decision-making	The data controller has to inform about the existence of automated decision making and profiling. This comprises information about the functionality of and the criteria relevant for these procedures as well as criteria about the scope and the consequences.[161]
▪ Whether processing data is intended on basis of a purpose other than that for which the personal data were collected	In case of reusing data for purposes other than the original purpose[162], the data controller shall inform about the new purpose in a detailed and comprehensive manner - if present, including the possibility to restrict the new data process, e.g. consent is necessary.[163]

III. Information obligations pursuant to Article 14 GDPR

The information to be provided correspond to the requirements of article 13 GDPR, with the constraint that the data controller has to inform the data subject within a reasonable period after obtaining the personal data, but at the latest within one month. Insofar, the following information have to be provided.

Criteria	Meaning
▪ Identity and contact details	The data controller has to provide information concerning his name, identity and other contact details in such a comprehensive manner, that the data subject is able to unproblematically establish contact with the data controller.[164]

[160] Kühling/Buchner Art. 13 Rn. 40
[161] Kühling/Buchner Art. 15 Rn. 25
[162] Pursuant to article 6 sec. 4 GDPR
[163] Kühling/Buchner Art. 13 Rn. 85
[164] Kühling/Buchner Art. 13 Rn. 22

▪ Purpose and legal base	The data controller shall provide information concerning the purpose of the personal data to be collected. These information shall be detailed and complete enough in a way that the data subject recognizes what data is collected for what purpose.[165] Additionally, the data controller shall inform about the legal basis on which personal data has been collected. In case of collecting personal data according to article 6 sec. 1it. (f), the legitimate interest pursued justifying the handling of personal data shall be represented.
▪ Categories of personal data	As the data subject does not participate in the collection of personal data, the data controller shall provide information about the categories of personal data in a precise and specific manner. Information beyond can be requested according to article 15 GDPR.
▪ The recipients or categories of recipients	The data controller shall provide information about the recipients[166] which comprises contracting data processing and the flow of data within different business units. This information include expected recipients or specific categories of recipients, whereas the best detailed information has to be provided.[167]
▪ Transfer of personal data to third country or international organisation	In case of transfer data to recipients in a third country or international organisation, the data controller shall explicitly inform and hint about a(n) (expected) data transfer outside the EU including information about the risk estimation and, if possible, objections against the transfer.[168]

[165] Kühling/Buchner Art. 13 Rn. 26
[166] Pursuant to article 4 nb. 9 GDPR: Recipient "*means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not."*
[167] In this way excluding choosing between several options for the benefit of the data controller. – Kühling/Buchner Art. 13 Rn. 30
[168] Kühling/Buchner Art. 13 Rn. 34

▪ Period for which the personal data will be stored	Providing information about the period for which personal data will be stored comprises either a specific date or the obligation to name according to which the period is measured.[169]
▪ Reference to the data subject´s rights	The data controller shall inform about the rights of the data subject including a brief description in a more general form: • Article 15: Right to access • Article 16: Right to rectification • Article 17: Right to erasure • Article 18: Right to restriction • Article 20: Right to data portability Article 21: Right to object
▪ Origin of personal data	The data controller has to inform about the origin of the data, which comprises where appropriate the subject as well as the mean of collecting data.[170]
▪ Whether the provision of personal data is a statutory or contractual requirement	In case that the data controller is obligated to contribute to the collection of personal in order to not getting confronted with legal consequences, the data controller has to inform the data subject about this circumstance.[171]
▪ The use of automated decision-making	The data controller has to inform about the existence of automated decision making and profiling. This comprises information about the functionality of and the criteria relevant for these procedures as well as criteria about the scope and the consequences.[172]

[169] Kühling/Buchner Art. 13 Rn. 34
[170] Kühling/Buchner Art. 15 Rn. 25
[171] Kühling/Buchner Art. 17 Rn. 40
[172] Kühling/Buchner Art. 15 Rn. 27

▪ Whether processing data is intended on basis of a purpose other than that for which the personal data were collected	In case of reusing data for purposes other than the original purpose[173], the data controller shall inform about the new purpose in a detailed and comprehensive manner - if present, including the possibility to restrict the new data process, e.g. consent is necessary.[174]

5.3.4.13. Confidentiality and security of data processing

The confidentiality and security of the processing are key issues for personal data protection, to be tackled with high precaution in AEGIS: the risks and threats to which personal data undergoing processing activities are exposed are becoming higher (both in number and danger), notably with regard to Internet and automated data processing activities.

Security and confidentiality precautions aim at protecting personal data both in the static and in the dynamic moment of the data processing, including their storage in databases and their transfer to third parties,

Security measure may be technical (e.g. anti-virus, firewalls, authentication and authorisation systems), organisational (e.g. internal privacy policies, instructions or guidelines, internal procedures) or physical (e.g. measures to control access and ensure security of the Controller's premises).

According to Article 17, the security measures have to be adopted: "to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access, in particular where the processing involves the transmission of data over a network, and against all other unlawful forms of processing.

Having regard to the state of the art and the cost of their implementation, such measures shall ensure a level of security appropriate to the risks represented by the processing and the nature of the data to be protected".

The Regulation addresses the issue in Articles 32, 33 and 34, stating that "taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate, pseudonymisation and encryption of personal data, the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident" and other measures.

[173] Pursuant to article 6 sec. 4 GDPR
[174] Kühling/Buchner Art. 13 Rn. 87

5.3.5. Directive 2002/58/EC "ePrivacy Directive"

The "ePrivacy Directive" (Directive 2002/58/EC on privacy and electronic communications) replaced the Directive 97/66/EC and was partially amended by Directive 2009/136/EC. The "ePrivacy Directive" pertains to the processing of personal data and the protection of privacy in the sector of electronic communications and transposes in the telecommunications sector, which is a "sensitive" area from a privacy perspective, the main principles and rules of the Data Protection Directive (therefore, now, GDPR).

Though the ePrivacy Directive is an important legal instrument for privacy in the digital age, notably as regards the confidentiality of communications and the tracking and monitoring, this text is expected to be updated, due to the entry into force of the GDPR. The European Commission acknowledged this need and published a proposal on 10 January 2017, in order to tackle the rapidly evolving technological landscape, with issues such as confidentiality of machine-to-machine communication (Internet of Things) or the confidentiality of individuals' communication on publicly accessible networks (such as public Wi-Fi). This proposal is currently under discussion in the European Parliament and the Council of the European Union.

The ePrivacy replacement texts are expected to be adopted in time to become applicable at the same time as the GDPR, in view of giving rise to a comprehensive modern framework for protecting privacy and for data protection.

However, Article 95 of GDPR states that there will not be additional obligations on natural or legal persons in relation to processing in connection with the provision of publicly available electronic communications services in public communication networks.

The main relevant provisions in relation to AEGIS are as outlined hereunder.

I. Security

Article 4. par. 1 sets forth to the obligation of adopting security measures: "the provider of a publicly available electronic communications service must take appropriate technical and organisational measures to safeguard security of its services, if necessary in conjunction with the provider of the public communications network with respect to network security. Having regard to the state of the art and the cost of their implementation, these measures shall ensure a level of security appropriate to the risk presented". The mandatory minimum precautions to be adopted were specified by the Directive 2009/136/EC, which amended Article 4.

The appropriateness of the security measures has to be assessed on a case by case basis, by making reference to the specific factual circumstances and conditions of the processing of personal data, to the state of the art technologies and to implementation costs.

In addition to these security obligations, the Controller, in case particular threats may occur for the network security, has to inform the users and also indicate possible remedies.

Article 4, as amended by the Directive 2009/136/EC, specifies the mandatory minimum precautions to be adopted. The security requirements should at least: i) "ensure that personal data

can be accessed only by authorised personnel for legally authorised purposes; ii) protect personal data stored or transmitted against accidental or unlawful destruction, accidental loss or alteration, and unauthorised or unlawful storage, processing, access or disclosure, and, iii) ensure the implementation of a security policy with respect to the processing of personal data".

The Directive 2009/136/EC introduced also the definition of data breach, as follows: "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed in connection with the provision of a publicly available electronic communications service in the Community." In case of data breach, there is the 'duty to warn', consisting in the Controller's obligation to notify security breaches occurred in the course of the processing, detailing the procedures and rules for such a notification, towards both the competent national data protection authority and the interested data subject. The latter does not apply if the Controller adopted appropriate technological security measures that make data unintelligible to anyone who has no access authorisation.

II. Protection to confidentiality of the communications among individuals

According to Article 5, adequate protection has to be devoted to confidentiality in the communications. It may be limited only in case of specific circumstances.

An exemption to the prohibition of interception of communications (e.g. storing or other kinds of surveillance of communications and the related traffic data) occurs when such an interception is performed with specific precautions or by specifically authorised subjects (e.g. when users provided their consent, or applicable law provisions authorise it, or storage is functional to conveying the communications and without prejudice to the confidentiality principle).

The use of a deployment of electronic communications networks with the aim to store or have access to information kept in the user's terminal equipment is legitimate, provided that such user receives the mandatory information established by the Data Protection Directive and that he/she can oppose this data processing. The exemption is when this kind of activity is necessary from a technical point of view.

The protection of confidentiality of communications covers the communication itself, the users' terminal equipment (or other tool used by user to communicate electronically) and the information and data stored in such equipment and tools.

In case of deployment of invasive and tracking technologies (e.g. tags, spy wares, hidden identifiers and cookies), stringent provisions are set forth, considering the serious threat for users' privacy and confidentiality (e.g. it is possible to map and track users' online activities, to collect data from the technical equipment deployed).

As regards cookies, the user has to be provided with the mandatory information required under the Data Protection Directive and to be allowed to intervene on cookies, turning them down.

Similar provisions also apply to the other tracking technologies. Furthermore, it is necessary to have the data subject's consent for storage and gathering of information that is in turn stored on

his/her terminal equipment (e.g. cookies and other tracking technologies), and user-friendly information has to be given to the data subject in order to enable him to willingly express his preferences, including his right to refusal.

III. Traffic data and location data

Traffic data is "any data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service" (Article 2 letter C): therefore, it is personal information linked to communications and use of the Internet.

Given that traffic data poses serious concerns from a data protection standpoint and that possible potential threats regard concern surveillance, misuse and the pervasive encroaching into an individual's private sphere, its legitimate processing is subject to strict requirements.

Though some exceptions are indicated, "Traffic data relating to subscribers and users processed and stored by the provider of a public communications network or publicly available electronic communications service must be erased or made anonymous when it is no longer needed for the purpose of the transmission of a communication" (Article 6).

User' consent is considered as a tool for the protection of data subject's freedom of expression and rights to data protection and confidentiality in the communications. However, the set of mandatory information to be provided to the same, is larger than that identified under the Data Protection Directive (e.g. additional details on the types of traffic data collected and processed and on the specific time length of the processing activities). In addition, "processing of traffic data… must be restricted to persons acting under the authority of providers of the public communications networks and publicly available electronic communications services handling billing or traffic management, customer enquiries, fraud detection, marketing electronic communications services or providing a value added service, and must be restricted to what is necessary for the purposes of such activities" (Article 6).

The legitimate traffic data processing activities are only those strictly necessary and functional to achieve the specific legitimate purposes, and the traffic data may be kept and processed only for the time strictly necessary and functional to such purposes (this derives from the necessity, proportionality and time storage principles).

Location data are a type of traffic data. They "may refer to the latitude, longitude and altitude of the user's terminal equipment, to the direction of travel, to the level of accuracy of the location information, to the identification of the network cell in which the terminal equipment is located at a certain point in time and to the time the location information was recorded" (Recital 14). The concept was extended by the Directive 2009/136/EC, thus including also personal data processed by an electronic communications service.

Location data may be lawfully processed only "when they are made anonymous, or with the consent of the users or subscribers to the extent and for the duration necessary for the provision of a value added service" (Article 9). Informative requirement has to be followed as well (e.g. type of data, time length, extent, etc.).

A data subject can withdraw his/her consent at any time, and this kind of data may be accessed and processed only by persons under the authority of the Controller (or the third party providing value added services), whilst data collection and processing activities have to be limited to what is strictly necessary.

IV data retention

Article 15 refers to data retention. It assumed a key role since 2014, when the "Data Retention Directive" (Directive 2006/24/EC) was declared invalid by the Court of Justice because it did not meet the principle of proportionality and entailed a wide-ranging and particularly serious interference with fundamental rights. In fact, the retained data could provide a clear insight of data subject's private lives (e.g. his/her habits of everyday life, daily movements, frequent activities, social relationships, etc).

The annulment of Directive 2006/24/EC implied the need to refer to both ePrivacy Directive and to the guarantees of the European Convection on Human Rights and its interpretation.

The latter set forth the following principles:

- need to strict necessity and proportionality of collection, retention and transfer of data;
- rejection of the blanket data retention of unsuspicious persons and indefinite or even lengthy retention period of data retained;
- need of link between a threat to public security and the data retained for such purposes;
- need for effective procedural rules, like independent oversight and access control;
- need to address the risk of stigmatisation stemming from the inclusion of data in law enforcement databases.

The ePrivacy Directive, in Article 15, par. 1, though gives Member States the possibility to exceptionally introduce data retention schemes deviating from the general prohibition to collect and store data, underlines the need to have a very strict and detailed measure of compatibility with fundamental rights standards, taking into account the formulation of Article 8 of the European Convention on Human Rights.

5.3.6. Regulatory Framework in the selected jurisdictions

The following paragraphs provide a concise overview of the regulatory framework implementing European privacy and data protection legislations respectively in Italy, Austria and Greece, the countries where the use cases and demonstrators will be located. Some of the information reported here corresponds with that inserted in D9.2, where a first snapshot of demonstrators' ethical, privacy and data protection concepts was provided.

As pointed out hereabove, GDPR "shall be binding in its entirety and directly applicable in all Member States". The national provisions analysed in this chapter, which were adopted pursuant to the Directive repealed by GDPR, can still be considered in the regulatory framework relevant to AEGIS. Nevertheless, in case of conflict, as a general rule (with exceptions for instance as regards constitutional laws), GDPR prevails.

5.3.6.1. Demonstrator 1: Road Safety Indicator

The automotive and road safety demonstrator will be developed in three versions, Broken Road Indicator, Safe Driving Indicator, and Regional Driving Style Risk Estimator. The three versions of the automotive demonstrator are aimed to provide the following benefits:

- Provide insights into road conditions based on exploiting individual vehicle sensor data, traffic data, and map data.
- Infer the driver's safety style and then calculate a safety index, through utilising vehicle sensor data along with environmental information and other content.
- Calculate a regional driving safety risk metric for certain regions including intersections, streets, cities or countries.

During the project runtime the automotive and road safety demonstrator will involve human participants as volunteers for:

(a) generating driving data in the field (vehicle usage data) as well as in laboratory settings using a driving simulator (vehicle simulator data), and

(b) evaluating usefulness and usability of the developed services & applications running in a browser and/or on a mobile phone.

Before the experiments begin, an informed consent procedure will be applied. All participants who want to volunteer in the experiments of the automotive demonstrator have to sign a declaration of consent. Study participants will be made aware on of the project goals as well as of their role in the experiments. Each volunteer will be clearly informed on of the possibility to refuse to enter or to retract at any times with no consequences. All experiments will be designed and implemented according to the Data protection and privacy ethical guidelines from the European Commission and to the main sources of national legislation relevant to AEGIS in Austria, in particular the "Datenschutzgesetz 2000 - DSG 2000" (Federal Act concerning the Protection of Personal Data), which is the current data protection act and the foundation of data protection law, the Telecommunications Act 2003 (TKG 2003) and Austrian Federal Constitutional Law.

5.3.6.2. Demonstrator 2: Smart Home and Assisted Living

Considering the specificities of the 2nd project demonstrator, the first step was to investigate and study the laws which are associated with the activities of the project. Beside the directives of the EU, the legislation of the countries where the demonstrator will be established (Greece) has been taken into consideration. Concisely, the legislation with which the AEGIS framework has to conform includes:

Greece – Law 2472/97 (amendments: 3471/06 & 3917/11)

The AEGIS project has to abide by the national laws of the countries that are involved in the pilots or in other activities of the project. In this section, **some key articles** will be mentioned

underlying the legal and ethical scope of the AEGIS framework in the Smart Home and Assisted Living demonstrator.

1. An Authority (NDPA) has been created, as described in the following article, in order to enforce it.

Chapter D – Article 15

1. A Personal Data Protection Authority (hereinafter: the Authority) is hereby created with the task to supervise the implementation of this law and all other regulations pertaining to the protection of individuals from the processing of personal data as well as to the exercise of the duties assigned to it each time.

2. The Authority constitutes an independent public authority and will be assisted by its own Secretariat. The Authority shall not be subject to any administrative control. In the course of their duties the members of the Authority shall enjoy personal and functional independence. The Authority reports to the Minister of Justice and its seat is in Athens.

3. All necessary appropriations for the operation of the Authority shall be entered in a special code which shall be integrated in the annual Budget of the Ministry of Justice. The authorising officer for the expenditure is the President or his substitute.

2. Data Controllers must respect the provisions of Law 2472/1997 (and 3471/2006 regarding electronic communications) and more specifically:

They must collect personal data fairly and lawfully.

They must process only the data which are necessary for one or more specified purposes.

They must make sure that they keep data accurate and up to date.

They must retain data only for as long as is deemed necessary for the purpose of the collection and process thereof.

In order to carry out the data processing, the Controller must choose employees with relevant professional qualifications providing sufficient guarantees in terms of technical expertise and personal integrity to ensure such confidentiality.

The Controller must implement appropriate organisational and technical measures to secure data and protect them against accidental or unlawful destruction, accidental loss, alteration, unauthorised disclosure or access as well as any other form of unlawful processing.

If the data processing is carried out on behalf of the controller, by a person not dependent upon him, the relevant assignment must necessarily be in writing.

The controller must respect the data subject's rights to information, access and objection.

They must meet their obligations vis-a-vis the DPA (notification, granting of permit).

They must be kept informed on any Decisions, Directives or Recommendations issued by the DPA that may be important to them.

3. More specifically and based on Article 4 - Law 2472/97 (Characteristics of personal data):

1. Personal data, in order to be lawfully processed, must be: a) collected fairly and lawfully for specific, explicit and legitimate purposes and fairly and lawfully processed in view of such purposes. b) **adequate, relevant and not excessive** in relation to the purposes for which they are processed at any given time. c) **accurate and, where necessary, kept up to date**. d) kept in a **form which permits identification** of data subjects for no longer than the period required, according to the Authority, for the purposes for which such data were collected or processed.

Once this period of time is lapsed, the Authority may, by means of a reasoned decision, allow the maintenance of personal data for historical, scientific or statistical purposes, provided that it considers that the rights of the data subjects or even third parties are not violated in any given case.

2. It shall be for the Controller to ensure compliance with the provisions of the previous paragraph. Personal data, which have been collected or are being processed in breach of the previous paragraph, shall be destroyed, such destruction being the Controller's responsibility. The Authority, once such a breach is established, either ex officio or upon submission of a relevant complaint, shall order any such collection or processing ceased and the destruction of the personal data already collected or processed.

4. Article 6 defines the notification process towards contacting the NDPA for getting full consent about exploiting datasets.

The Controller must notify the Authority in writing about the establishment and operation of a file or the commencement of data processing.

In the course of the aforementioned notification, the Controller must necessarily declare the following:

a) his/her name, trade name or distinctive title, as well as his/her address. (The second item is deleted, as it is no longer valid)

b) the address where the file or the main hardware supporting the data processing are established.

c) the description of the purpose of the processing of personal data included or about to be included in the file.

d) the category of personal data that are being processed or about to be processed or included or about to be included in the file.

e) the time period during which s/he intends to carry out data processing or preserve the file.

f) the recipients or the categories of recipients to whom such personal data are or may be communicated.

g) any transfer and the purpose of such transfer of personal data to third countries.

h) the basic characteristics of the system and the safety measures taken for the protection of the file or data processing.

i) (The item was deleted pursuant to paragraph 2 of article 8 of Law 2819/2000, Official Gazette A/84)

3. The data referred to in the preceding paragraph will be registered with the Files and Data Processing Register kept by the Authority.

4. Any modification of the data referred to in paragraph 2 must be communicated in writing and without any undue delay by the Controller to the Authority'.

5. Article 7a- Exemption from the obligation to notify and receive a permit

1. The Controller is exempted from the obligation of notification, according to Article 6, and the obligation to receive a permit, according to Article 7 of the present Law in the following cases:

a. When the processing is carried out **exclusively for purposes relating directly to an employment** or project relationship or to the provision of services to the public sector and is necessary for the fulfilment of an obligation imposed by law or for the accomplishment of obligations arising from the aforementioned relationships, and upon prior announcement to the data subject.

b. When the processing involves clients' or suppliers' personal data, provided that such data are **neither transferred nor disclosed to third parties**. In order that this provision may be applied courts of justice and public authorities are not considered to be third parties, provided that such a transfer or disclosure is imposed by law or a judicial decision. Insurance companies, for all types of insurance, pharmaceutical companies, companies whose main activities involve trading of data, credit and financial institutions, such as banks and institutions issuing credit cards are not exempted from the obligation of notification.

c. When the processing is carried out by societies, enterprises, associations and political parties and relates to personal data of their members or companies, provided that the latter have given their consent and that such data are neither transferred nor disclosed to third parties. Members and partners are not considered to be third parties, provided that said transfer is carried out among said members and partners for the purposes of the aforementioned legal entities or associations. Courts of justice and public authorities are not considered to be third parties, provided that such a transfer is imposed by law or a judicial decision.

d. When the processing involves medical data and is carried out by doctors or other persons rendering medical services a, provided that the Controller is bound by medical confidentiality or other obligation of professional secrecy, provided for in Law or code of practice, and data are neither transferred nor disclosed to third parties. In order for this provision to be applied, courts of justice and public authorities are not considered to be third parties, provided that such a transfer or disclosure is imposed by law or judicial decision.

e. When the processing is carried out by lawyers, notaries, unpaid land registrars and court officers or companies formed by the aforementioned and involves the provision of legal services to their clients, provided that the Controller and the members of the companies are bound by an obligation of confidentiality imposed by Law and that data are neither transferred nor disclosed to third parties, except for those cases where this is necessary and is directly related to the fulfilment of a client's mandate.

f. When the processing is carried out by judicial authorities or services, with the exception of the authorities referred to under item b of paragraph 2 of Article 3, in the framework of attributing justice or for their proper operation needs.

For further information please visit the Hellenic Data Protection Authority (www.dpa.gr).

While the laws establish some core principles both at European and National level, they do not establish clear lines for the field of research. The AEGIS consortium will abide by the above-mentioned legislation and will act with respect to the rights of any human being that is involved in the project either as a participant or not, according to the "*Data Protection and Privacy Ethical Guidelines*" of the Ethical Review in HORIZON 2020.

5.3.6.3. Demonstrator 3: Insurance Sector. Personalised Early Warning System for Asset Protection

The main source of regulation relevant for the Personalised Early Warning System for Asset Protection Demonstrator is the Italian Data Protection Code or Privacy Code (Legislative Decree n. 196/2003). It came into force on 1 January 2004 and superseded previous laws, in particular Data Protection Act 1996 n. 675/1996. In respect of this, the Privacy Code adopted a more practical approach, especially by removing all the previous requirements that resulted in mere formalities. The Data Protection Code, which is still in effect, was amended by a series of subsequent instruments.

The code, which is mainly applicable to all processing within the State and its territories, consists of three parts, respectively setting forth:

- the general data protection principles, applying to all organisations;
- additional measures that will need to be undertaken by organisations in certain areas (e.g. healthcare, telecommunications);
- sanctions and remedies.

The first Article of the Code expressly acknowledges that: "Everyone has the right to protection of personal data concerning himself".

The key guiding principles behind such Code are simplification, harmonisation, and

effectiveness. Other important points are as follows:

- The codes encompasses the element of data minimisation and boosts organisations in making use of non-personal data whenever possible;
- Data subjects are allowed to exercise their rights and instigate proceedings in an easier manner, so that to better safeguard and promote their data protection rights. In relation to compliance and enforcement, in case data subject have been prevented from exercising his/her rights, he/she can settle disputes either through the courts or by lodging a complaint with the Garante;
- International data transfers (outside the EU), according to Article 42-45, on the one hand, businesses have to provide notification only when such a transfer is able to prejudice data subjects' rights, and, on the other hand, notifications have need not to be yearly resubmitted yearly. The transfer of processed personal data to a non-EU Member State shall also be permitted if it is authorised by the Garante on the basis of adequate safeguards for data subjects' rights;
- In case of processing of personal data, Article 26 of the Codes provides the need of the Garante's authorisation. "General Authorisations", targeted to industry sectors and/or specific categories of data, were issued by the Garante, in compliance of Article 40, to prevent private-sector data controllers from having to apply for ad-hoc authorisations;
- The processing operation related to electronic communication data is addressed in Title X "Electronic Communication". Here we can mention only some of its provisions:
    - Article 121 clearly defines the extent of application of the title: "processing of personal data in connection with the provision of publicly accessible electronic communication services on public communications networks".
    - Section 122 states that:
        - "1. Subject to paragraph 2, it shall be prohibited to use an electronic communication network to gain access to information stored in the terminal equipment of a subscriber or user, to store information or monitor operations performed by an user.
        - 2. The Code of conduct referred to in Article 133 shall lay down prerequisites and limitations for a provider of an electronic communication service to use the network in the manner described in paragraph 1 for specific, legitimate purposes related to technical storage for no longer than is strictly necessary to transmit a communication or provide a specific service as requested by a subscriber or user that has given his/her consent based on prior information as per Artcile Article 13, whereby purposes and duration of the processing shall have to be referred to in detail, clearly and accurately.
    - Section 123, in relation to traffic data, states that:
        - "1. Traffic data relating to subscribers and users that are processed by the provider of a public communications network or publicly available electronic communications service shall be erased or made anonymous when they are no longer necessary for the purpose of transmitting the electronic communication, subject to paragraphs 2, 3 and 5.
        - 2. Providers shall be allowed to process traffic data that are strictly necessary for subscriber billing and interconnection payments for a period

not in excess of six months in order to provide evidence in case the bill is challenged or payment is to be pursued, subject to such additional retention as may be specifically necessary on account of a claim also lodged with judicial authorities.

- 3. For the purpose of marketing electronic communications services or for the provision of value added services, the provider of a publicly available electronic communications service may process the data referred to in paragraph 2 to the extent and for the duration necessary for such services or marketing, on condition that the subscriber or user to whom the data relate has given his/her consent. Such consent may be withdrawn at any time.
- 4. In providing the information referred to in Article 13, the service provider shall inform a subscriber or user on the nature of the traffic data processed as well as on duration of the processing for the purposes referred to in paragraphs 2 and 3.
- 5. Processing of traffic data shall be restricted to persons in charge of the processing who act — pursuant to Article 30 — directly under the authority of the provider of a publicly available electronic communications service or, where applicable, the provider of a public communications network and deal with billing or traffic management, customer enquiries, fraud detection, marketing of electronic communications or the provision of value-added services. Processing shall be restricted to what is absolutely necessary for the purposes of such activities and must allow identification of the person in charge of the processing who accesses the data, also by means of automated interrogation procedures…".

- Section 126, in relation to location data states that:
  - "1. Location data other than traffic data, relating to users or subscribers of public communications networks or publicly available electronic communications services, may only be processed when they are made anonymous, or with the prior consent of the users or subscribers, which may be withdrawn at any time, to the extent and for the duration necessary for the provision of a value added service.
  - 2. The service provider must inform the users or subscribers, prior to obtaining their consent, of the type of location data other than traffic data which will be processed, of the purposes and duration of the processing and whether the data will be transmitted to a third party for the purpose of providing the value added service.
  - 3. Where consent of the users or subscribers has been obtained for the processing of location data other than traffic data, the user or subscriber shall continue to have the possibility, using a simple means and free of charge, of requesting to temporarily refuse the processing of such data for each connection to the network or for each transmission of a communication.
  - 4. Processing of location data other than traffic data in accordance with paragraphs 1, 2 and 3 shall be restricted to persons in charge of the

processing acting pursuant to Section 30 under the authority of the provider of the publicly available communications service or, as the case may be, the public communications network or of the third party providing the value added service. Processing shall be restricted to what is necessary for the purposes of providing the value added service and must ensure identification of the persons in charge of the processing that access the data also by means of automated interrogation operations".

  - As regards traffic data retention other than for purposes of dealing with disputes over billing and subscriber services, according to Article 132 it is possible for communications service providers (CSPs) to retain traffic data for thirty months;
  - Article 133 and 134 deal with the codes of conduct and professional practice and enhance their importance in respect of the protection of personal data: their adoption is encouraged in highly significant sectors such as processing of data via the Internet.
- Title IV provides the definitions of the actors that perform the processing: data processor, controller and persons in charge of processing: Article 28. 29, 30;
- The security measures are set forth in Annex B;
- Article 13 refers to the set of information to be given to the data subject, orally or in writing. The usual practice is to provide him with a written information statement. Besides this, for traffic data (Article 123) and location data additional (Article 126), further information must be given. Only in restricted exemptions the Controller is exempted from the obligation of giving the information to the data subject (Article 13, par. 4).
- Article 23 and Article 24 respectively linger over the data subject's consent and excemptions. A data subject's consent has to be: express, free, specific, informed, given in advance, documented in writing in case of processing of personal data (the consent for sensitive data must be given through written instrument). In case of network monitoring, it is relevant the specific purpose for which it is performed, to determine if there is or not the necessity to obtain the data subject' s consent. According to Articles 123 and 126, for the processing of traffic data and of location data, usually consent usually is necessary, also for performance of value added services. As to sensitive data processing, it is necessary an authorisation issued by the Garante and data subject's written consent (save for limited exemptions).
- Title VII, in Article 42 – 45, deepens the transborder data flow and, in general, the transfer of data.

5.4. Project implementation phase

The AEGIS EP Strategy, based on the aforementioned regulatory framework, is structured into two main parts. The first moves around the project's implementation phase and refers to all the issues relevant during project's development, including ethics processes, Ethics Advisory Board's set-up and operations, AEGIS demonstrators, as well as an overview of Ethics procedures and Roadmap and hints for data protection impact assessment methodology. The second part refers mainly refers to AEGIS solutions and requirements to be complied with.

5.4.1. Ethics Advisory Board

The Ethics Advisory Board (EAB) was set up and is working closely with AEGIS Consortium during the course of the project on tackling ethical and data privacy issues that will have to do with the retrieval, the processing, and the retaining of these data. The EAB's role is directed to evaluate the AEGIS's progress and the results generated and supervise the operation of the project, in order to ensure that European and national regulations regarding data protection are fully observed and that the framework and its implementation adhere to a minimum set of ethical and legal requirements. At the same time EAB advises the Project Partners how to proceed with the research activities in an ethically correct way and in compliance with the applicable legislations.

The EAB is coordinated by Dr Maurizio Ferraris, as EAB Coordinator, who is responsible for interfacing with it.

Upon demand of GFT, the EAB will perform the following activities:

a. provide expertise in specific ethics and privacy areas (as instructed by the Consortium and the EC) during the whole duration of the project and contribute to provide independent opinions and thoughts and to advise both the technical and the research partners on issues regarding the AEGIS methodology, the development of the platform and its components and the piloting operation.
b. contribute to propose the Assessment Methodology to be described in D9.1 and followed in WP1 and WP5, including, if opportune, the provision of templates at an early stage and the coherence with the Ethical Risk Table already named in the AEGIS Annex I;
c. participate and/or contribute to AEGIS workshops or meetings, which will be conducted during the project;
d. co-create and/or review selected parts of the ethics and privacy related deliverables;
e. periodically report to the commission on the implementation of the ethical issues in project and compliance with applicable national and EU regulations. The Ethics Advisory Board's Report will summarise the evaluation activities of the Ethics Advisory Board and will contain the Ethics Advisory Board's recommendations. The reports will be based on a common assessment methodology as introduced in D9.1 and will be submitted as AEGIS Deliverable 9.3, as attachment to the AEGIS Periodical Reporting in Project Month 18 and, at the end of the Project, in an updated version as attachment to the AEGIS Periodical Reporting to be submitted in Project Month 30.

5.4.2. Demonstrators/use cases: final ethics and data protection remarks

5.4.2.1. Demonstrator 1: Automotive and Road Safety Demonstrator

The AEGIS Automotive and Road Safety Demonstrator explores how vehicle driving data and other road safety related data including e.g. weather data to name one concrete source can be meshed and modelled, aggregated, and semantically annotated in order to extract meaningful,

safety-relevant information. For this, various combinations of vehicle driving datasets and datasets from other domains will be investigated to determine which of them provides the most valuable insights into driving styles and driving behaviour. Beneficiaries including drivers and other stakeholders will enhance their (business) value by using the AEGIS platform to create services for safer driving and safer roads.

The automotive and road safety demonstrator will be developed according to three different scenarios, Broken Road Indicator, Safe Driving Indicator, and Regional Driving Style Risk Estimator. The three different corresponding versions of the automotive and road safety demonstrator are then aimed to provide the following benefits to the users of the services:

- Provide insights into road conditions based on exploiting individual vehicle sensor data, traffic data, and map data (Broken Road Indicator).
- Infer the driver's safety style and then calculate a safety index, through utilising vehicle sensor data along with environmental information and other content (Safe Driving Indicator).
- Calculate a regional driving safety risk metric for certain regions including intersections, streets, cities or countries (Regional Driving Style Risk Estimator).

The final automotive and road safety demonstrator will include all three versions, Broken Road Indicator, Safe Driving Indicator, and Regional Driving Style Risk Estimator.

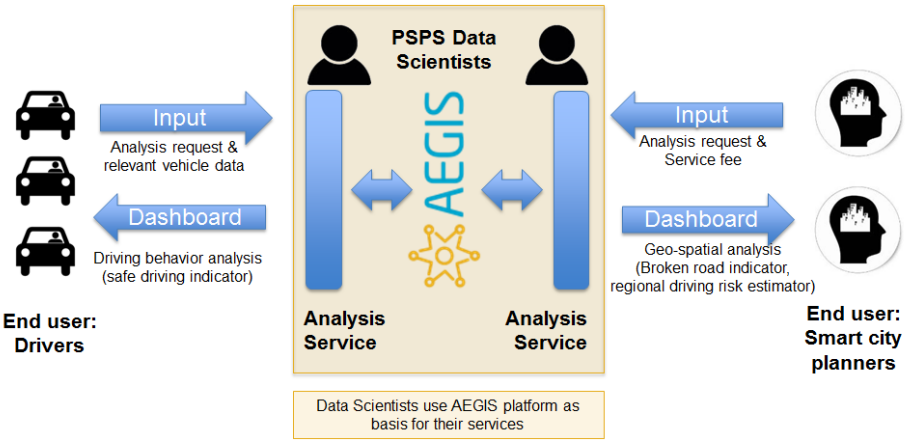

Figure 5-2: Actors of the automotive and road safety demonstrator

The automotive demonstrator is 'located' in Greater Graz area in Austria as the majority of vehicle trips have been recorded in this area. PSPS data scientists from VIF will use the AEGIS platform to implement the automotive demonstrator on the platform. Furthermore, VIF will provide vehicle data to the platform to enable service creation as well as develop algorithms to detect safety-relevant events. PSPS data scientists from VIF are responsible to develop the automotive demonstrator, which is mainly an analysis service for vehicle data and other sources of relevant data to detect road damage as well as safety-related events and visualise them on geographic maps allowing also comparisons between different regions. Trip data is generated by various drivers employed at VIF differing in age, sex, and driving experience (an informed

consent procedure has been implemented). Additional relevant data for driving analytics (e.g. weather data) is supposed to be accessed via the platform.

The responsible national data protection authority in Austria is Austrian Data Protection Authority (in German: "Datenschutzbehörde"), a governmental authority charged with data protection. The data protection authority is the Austrian supervisory authority for data protection, the equivalent of a national data protection commissioner in other countries.

Despite the automotive and road safety demonstrator in the AEGIS project will not involve processing any personal data, according to the corresponding business scenarios and business models developed in the project and aiming to scale these applications to the market, a future collection of personal data might be taken into account. A collection of personal data for establishing novel data-driven services in the automotive domain applies e.g. if a future user of one of these applications might link the data he or she generates during the operation of a vehicle with his or her social media / web accounts, e.g. to inform his social network about how he attained a safe driving style. A user might for instance use his or her Facebook or Twitter account to log in or to share information with peers, which requires a professional data protection concept to safeguard ethics and privacy for future exploitation. However, this only affects the post-project exploitation phase.

Nevertheless, in parallel to the activities conducted during the project runtime, Virtual Vehicle will therefore approach the Austrian National Data Protection Authority to discuss the requirements for data protection, if Virtual Vehicles foresees any linkage of personal data in the post-exploitation phase of the AEGIS project for services related to automotive and road safety building on the results of the AEGIS project. This will ensure that services developed in the post-project exploitation phase will be developed according to 'privacy by design'.

Data to be collected during the experiments is **sensor data** (**technical data**) and/or **simulation data**. Sensor data is generated through connecting a device developed at VIF 'termed vehicle data logger' to the on-board diagnostic (OBD2) interface of a car. Sensor data will include for instance vehicle speed, vehicle rpm, or vehicle acceleration to name a few types. Simulation data is generated by study participants using a driving simulator at VIF and may include many additional values. Both sensor data and simulation data has to be stored on a research server at VIF to allow the development of algorithms for inferring events including broken roads, patterns of safe and unsafe driving, or driving risks. Sensor and simulation data will be kept on this server till the end of the project.

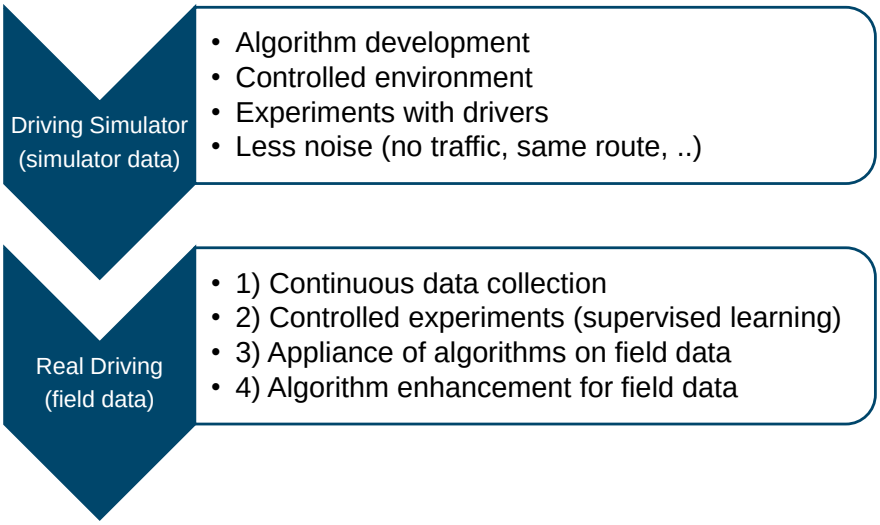

Figure 5-3: Simulator data and field data

During the AEGIS project, the automotive and road demonstrator involves the development and evaluation of applications running in a browser together with volunteers. During these automotive and road safety data related experiments, no identification data will be electronically stored on a server. Furthermore, no sensible personal data on health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction, etc. will be collected at all. The figure below shows data sources related for the automotive and road safety demonstrator.

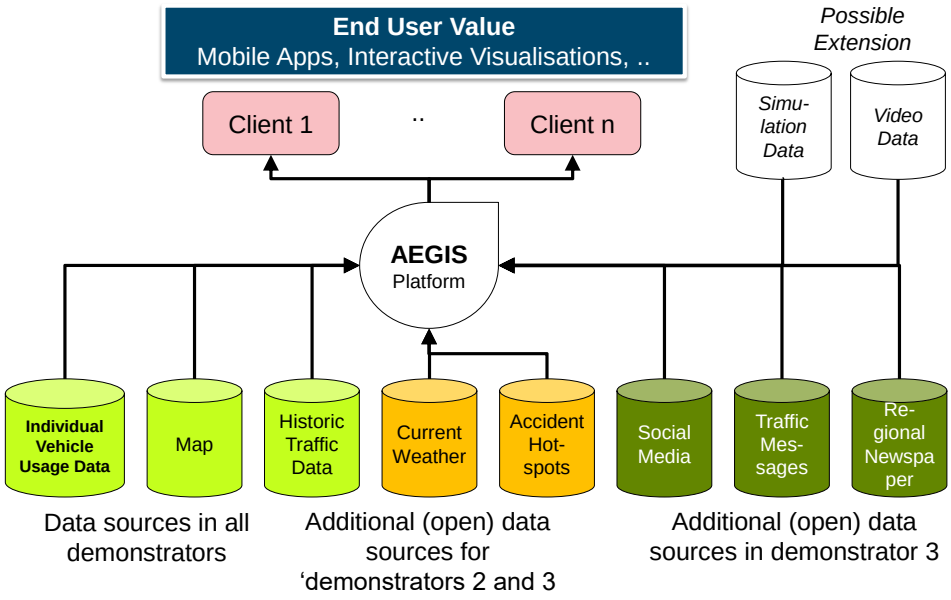

Figure 5-4: Data sources relevant to the automotive demonstrator

5.4.2.2. Demonstrator 2: Smart Home and Assisted Living

The smart home and assisted living demonstrator will implement two main services, with respective scenarios, that can be offered by a care service provider to at-risk individuals and/or their (in)formal carers. In particular, the services are the following:

i. Monitoring and analysis of an individual's well-being conditions, physical activity, positioning and wearable information and external environment data (e.g. weather, crime, news, social media), towards provision of a service for personalised notification and recommendation system for at-risk individuals, including notifications for carers.
ii. Additional service pertaining monitoring and analysis of weather, indoor environmental conditions, energy and operational device data towards the provision of a smart home application, which can be offered by care providers to at-risk people for increased indoor comfort and welfare.

Towards the demonstration of Smart Home and Assisted Living services in AEGIS project, a detailed overview of the ethics and data protection remarks have been detailed in deliverable D1.2. Here we concentrate on specific updates on this material, as resulted from the project and demonstrator developments during the intervening period. In particular, detailed stories and test cases have been established and documented in deliverable 5.2. The list of datasets has not been extensively modified. Minor updates are reflected in the table below. Towards the demonstration of Smart Home and Assisted Living services, the following data types as retrieved from sensors and metering devices will be considered.

Motion data
Luminance
Indoor Air Quality
Indoor temperature and humidity
Control actions over lighting and HVAC
HVAC Energy Consumption
Wearable Sensor Data (Fitbit and/or Apple watch)
Smartphone Sensors (Accelerometer/GPS)
Personal Health Data (Dummy data)
Expert Rules Data

Figure 5-5: List of Datasets - Smart Home and Assisted Living Demonstrator

It must be highlighted that the smart home demonstration will be implemented and tested within and utilizing the demonstrator participants' premises and personnel and will not have a direct interaction with human individuals (end users are the demonstrators - the demonstrator phase will deliver applications but will not be pushed to a group of selected users rather the tests will be performed as part of the research activities of the demonstrator). Nevertheless, all required safety and security procedures (anonymization, local storage, dissociation) will be adopted and implemented, so as to guarantee that the services are market ready.

It is important to reiterate that the smart home and assisted living demonstrator evaluation does requires the installation of equipment and usage of wearable devices. By taking into account the national legislation about the installation of sensors, we are presenting indicative guidelines in the field:

- All sensors utilised during the demonstrator should be privacy-preserving and should neither acquire sensitive personal data nor violate personnel's privacy.

- The controller of the study or his representative, if any, must notify the supervisory authority (Ethical Advisory Board) before carrying out any data collection process. The information to be given in the notification shall include at least:

    - the name and address of the controller and of his representative, if any;

    - the purpose or purposes of the processing;

    - a description of the category or categories of data subject and of the data or categories of data relating to them;

    - the recipients or categories of recipient to whom the data might be disclosed;

    - proposed transfers of data to third countries;

    - a general description allowing a preliminary assessment to be made of the appropriateness of the measures taken to ensure security of processing.

- All offices/areas that will be monitored and controlled with any type of sensors and equipment should be appropriately marked with **Notification Posters**, describing in detail equipment used and monitoring procedures taking place towards project's objectives.

- All occupants, whose working offices/areas will be monitored during the pilot, should be thoroughly informed and their informed consent should be requested as specified above.

All experiments are designed according to the Data protection and privacy ethical guidelines from the European Commission as defined in "H2020 Guidance —How to complete your ethics self-assessment". In addition, considering the need to have a clearance about any possible ethical concerns in the project, HYPERTECH (leader of Smart Home and AAL demonstrator) has contacted the national data protection authority in Greece to get a full commitment from HDPA

about the AEGIS project activities. The sign from HDPA about the full clearance for AEGIS project activities will be available once received from the National Data Protection Authority.

5.4.2.3. Demonstrator 3: Insurance Sector. Support, Warning and Personal Offering

As outlined in D5.1 and D5.2, 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. customer location, insured/uninsured asset types, …). 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 (to be 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.

In this demonstrator, volunteers will be involved through the installation of the Mobile App and accepting the secondary use of their data; no sensitive data will be stored on the platform: the HDI in-house datasets will be anonymised through the Anonymisation tool provided by AEGIS before the upload on the platform.

In coherence with the project-level ethical, privacy and data protection overall strategy, a fine-tuning policy was elaborated for the Insurance Sector Demonstrator's Application by taking into account Italian regulatory system. It is fully described in D9.2.

Here it is important to remark the key requirements that have to be complied with, thus setting the frontiers of legally acceptable or affordable AEGIS measures and tools in the insurance sector demonstrator, with a particular focus on data processing.

- The Italian Informed Consent Procedure for gathering the volunteers' consent will meet the specific requirements set forth by the Italian Privacy Code. In particular:
    - Article 13 refers to the set of information to be given to the data subject, orally or in writing. The usual practice is to provide him/her with a written information statement. Besides this, for location data additional (Article 126), further information must be given. Only in restricted exemptions the Controller is exempted from the obligation of giving the information to the data subject (Article 13, par. 4);
    - Article 23 and Article 24 respectively linger over the data subject' consent and excemptions. The data subject's consent has to be: express, free, specific, informed, given in advance, documented in writing in case of processing of personal data (the consent for sensitive data must be given in writing). In case of network monitoring, it is relevant the specific purpose for which it is performed, to determine if there is or not the necessity to obtain the data subject' s consent. According to Articles 123 and 126, for the processing of location data usually consent is necessary, also for performance establishing of value added services. As to sensitive data processing, it is necessary to have an authorisation issued by the Garante and data subject's written consent (save for limited exemptions).

- The security measures, as "Technical specifications on minimum data security measures", indicated by Annex B of the Privacy Code can be split in minimum and adequate measures. The first former represent the minimum standard to be adopted to have a lawful processing, while the others latter, though not specifically defined by the Code, are those considered suitable by the same Controller in relation to the specific processing having regard to the goal of minimising any possible risk that may jeopardise the personal data or that may harm the data subject. The general criteria to be followed by the Controller, according to the Code, is that, taking into consideration technological innovations, their nature and the specific features of the processing, personal data shall be kept and controlled in such a way as to minimise, by means of suitable preventative security measures, the risk of their destruction or loss (whether by accident or not), of unauthorised access to the data or of processing operations that are either unlawful or inconsistent with the processing purposes. For the processing of location data, as written hereabove, stricter measures are compulsory (Article 123 and Article 126)[175]. These technical and organisational measures are also functional to ensure anonymity.
- the Data Controller and Data Processors (and, in case, sub-processors, if any) will be appointed and the set of responsibilities set for by the legislation will be assigned to them.
- The notification procedure to the National Data Protection Body (NDPB) will be completed. The Italian NDPB is the so-named "Garante per la protezione dei dati personali". It is an independent Authority set up in 1997, with the function to ensure respect for individuals' dignity and to safeguard fundamental rights and freedoms in connection with the processing of personal data. The Garante is very active in this role and promotes a set of initiatives aimed at fostering the correct enforcement of the Privacy Code. Article 37 of the Code requires the notification to the Garante only in case of processing of higher-risk categories of data, by stating as follows: "1. A data controller shall notify the processing of personal data he/she intends to perform exclusively if said processing concerns:

  a) genetic data, biometric data, or other data disclosing geographic location of individuals or objects by means of an electronic communications network;

  d) data processed with the help of electronic means aimed at profiling the data subject and/or his/her personality, analysing consumption patterns and/or choices, or monitoring use of electronic communications services except for such processing operations as are technically indispensable to deliver said services to users;

[175] In 2008 the Garante issued a General Regulation on Security In Telephone And Internet Traffic Data, containing details on the physical, organizational and technical data security measures that have to be implemented with regard to the processing and storage of personal data

f) data stored in ad-hoc data banks managed by electronic means in connection with creditworthiness, assets and liabilities, appropriate performance of obligations, and unlawful and/or fraudulent conduct".

5.4.3. Ethics Procedures, Roadmap and Data Protection Impact Assessment Methodology

I. Composition, selection and appointment process status

The EAB includes relevant external, independent experts and practitioners with knowledge and experience regarding ethical and privacy issues. In particular, the following experts were appointed:

- Prof. Gert G. Wagner
- Avv. Marina Da Bormida, PhD
- Ing. George D. Karagiannopoylos

The EAB was formed through unanimous approval of all invited members by the Consortium partners and in agreement with the EC.

The Ethics Advisory Board is coordinated by the EABC, who is responsible for interfacing with the Ethics Advisory Board. This role has been assigned to Mr. Maurizio Ferraris (GFT).

At first, the Ethical Advisory Board Experts signed a Non-Disclosure Agreement (NDA), which was prepared by Fraunhofer, and immediately after the Expert Agreement (EA), which was prepared by GFT.

The template of both of them is included in this deliverable, respectively as Appendix B and C.

II. Reporting activities

The EAB will periodically report to the Commission on the implementation of the ethical issues in the project and on the compliance with applicable national and EU regulations. The EAB will submit the reports along with the periodic activity reports of WP8 (Coordination and Project Management), namely D8.2 at M18 and D8.3 at M30: the EAB reporting at M18 will timely ensure that the project is on the right tracks just before the completion of WP1 (AEGIS Data Value Chain Definition and Project Methodology), while at M30 just before the project end.

III. Ethics peer-review activities

If opportune, key project deliverables will be evaluated by the EAB's experts in the framework of their oversight activities. In this case, in order to gather diversified and balanced viewpoints, GFT will circulate the document to each of them separately and will collect their feedback. After this initial phase, the experts will be encouraged to discuss the concerns eventually identified and to collectively propose recommendations.

Deliverables in which ethical issues are involved and/or relevant from a privacy, data protection and ethics perspective can be identified as follows:

- Deliverable D1.2 “Aegis Methodology and High Level Usage Scenarios” (M6) and its updated release in D1.3 “Final AEGIS Methodology” (M15)
- D2.1 “Semantic Representations and Data Policy and Business Mediator Conventions” (M8)
- D2.3 Update on Semantic Representation and Data handling and Analytics Methods (M18)
- D5.2 Demonstrators Readiness Documentation and Execution Scenarios (M14)
- D5.6: Final Evaluation, Impact Assessment and Adoption Guidelines (M30)
- Deliverable D6.3 - Data Management Handling Plan (M6)
- WP9 deliverables

In case of need, also some parts of WP3-4-5 deliverables can be subjected to EAB’s ethics peer-review.

IV. Extraordinary procedures (in case of ethical issues)

According to the DoA, in case of ethical issues partners consult:

1) at first, their own ethics departments

2) in a second time, the Ethics Advisory Board

The AEGIS partners will adhere to the recommendations of ethics departments and/or of the EAB and will implement the adequate mitigating actions, countermeasures necessary in order to reinforce ethical safeguards and fully comply with both ethical standards/best practices and regulatory obligations or constraints.

V. Ethics workshops

AEGIS consortium is considering to organise some public discussion of the privacy issues arising from the project research as part of the dissemination and public outreach activities

VI. Data Protection Impact Assessment

A comprehensive Data Protection Impact Assessment (DPIA) methodology will be elaborated in the framework of WP9, in particular in D9.1. This methodology will be strongly based on the **ethical, privacy and data protection requirements** as set forth in this deliverable and will contribute to reinforce the ethical safeguards, as well as to provide an in-depth exploration of the **societal consequences** (positive or negative) of the introduction of AEGIS system, as well as to approach data protection and ethical issues in a more comprehensive manner (going beyond the use of high-level data security solutions, as appropriately proposed by the project). As regards the exploration of the societal consequences (positive or negative) of the introduction of AEGIS system, it is important to bear in mind that trust reflects the sense of a general acceptance, in the meaning that the societal affirmation that in AEGIS a good equilibrium has been found between, on the one hand, individual privacy and ethical values and, on the other hand, interests as security, safety, economic growth. Otherwise, mistrust reflects exactly the opposite: the sense of a general unease and potential renunciation implying societal objection.

The DPIA should assess the particular **likelihood and severity of each risk** to data protection,

taking into account "the nature, scope, context and purposes of the processing and the sources of the risk". The starting point will be the ethical risk table inserted into the DoA and referred to in D9.2. The impact assessment will also include "the measures, safeguards and mechanisms envisaged for mitigating each risk, ensuring the protection of personal data". The key questions driving the DPIA will include the following: what is gained, what is lost, by whom, how is this framed and measured and shared, by whom, and how is this articulated to decision-making processes related to AEGIS technologies?

The DPIA Framework is going to comprise the **assessment of the pros as well as the cons** (lock-ins, limits and constraints, SWOT analysis) of AEGIS technologies in general and of demonstrators' applications. The impact assessment will conduct **balancing assessment**, between, on the one hand, privacy/data protection tensions and, on the other hand, societal expectations and public interests related to PSPS solutions.

In relation to DPIA, it is useful to mention Article 35 of the new Regulation. It indicates that "where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data".

Lingering over such a methodology, the **mid-term and final assessment of AEGIS operations, framework and architecture** will be elaborated, in particular within the ethics report, assessing to what extent the legal requirements have been taken into account and offering recommendations where appropriate.

5.5. Overall AEGIS platform and components

5.5.1. Methodology

Privacy-awareness and ethical compliance is one of the main objectives in designing, developing, and using AEGIS system: the system design takes privacy issues into appropriate account, bearing in mind that, from a wider perspective, a balancing operation has to be conducted between this kind of requirements and other kind of requirements (e.g. usability requirements, economic requirements). Consequently, the selection of ethical, privacy and data protection requirements and the assessment of their implementation play a pivotal role.

The approach taken for the identification and analysis of such requirements in AEGIS was not tackled from a purely legal perspective, but also rotates on the underlying ethical values, like individual's self determination, which implies both the possibility for individuals to be in control and data minimisation.

The AEGIS approach to privacy-awareness and ethical compliance is based on the combination of Privacy by Design and by Default method and Privacy Protection Goals method, as follows.

5.5.1.1. Privacy by Design and by Default

With reference to the responsibility principle, this part of the document will outline the technical and organizational requirements necessary to comply with this regulation. This will comprise privacy-friendly basic settings, *privacy by default*, as well as privacy enhancing technologies to consider privacy aspects within the design phase and executing phase, granting *privacy by design*

I. Privacy by Default

Privacy by default guarantees the user to have "effective" security settings enabled during the first use reps. after registration. This requirement arises from the fact that the user does not have sufficient knowledge and experience about the process of the concerning technology and thereby about the choice of "optimal" data protection settings. The responsibility of privacy by default is question of how data is collected and in which way (personal) data is collected.

II. Privacy by Design

This method addresses the design of the technical system as well as the business processes and relies on the idea that there is the need of putting privacy principles into the design process of data processing systems since the very beginning. The seven principles to be considered in the design process, as conceived by Cavoukian are: "1. Proactive not reactive – preventative not remedial 2. Privacy as the default setting 3. Privacy embedded into design 4. Full functionality – positive-sum, not zero-sum 5. End-to-end security – full lifecycle protection 6. Visibility and transparency – keep it open 7. Respect for user privacy – keep it individual and user-centric".

The implementation of privacy strategies during the design stage in Big Data is fundamental. Especially Big Data conceals a bunch of privacy risks whose damage occur in a late stage. According to Article 25 sec. I GDPR "*the controller shall [...] implement appropriate technical and organisational measures [...] which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects."*

The article includes two main requirements:

- Implement security safeguards to protect the infrastructure against harmful attacks - security against attacks from "outside"
- Implement appropriate means to comply with the data protection principles, and the data protection issues of the data subject.

Differences between privacy and security

What privacy comprises is illustrated in the section *informational self-determination.* The subject of guaranteeing privacy represents reliable measures against wanted and unintended attacks to protect certain parts of privacy from the individual. Whereas, security respectively cybersecurity seeks to enforce policies relating to several different aspects of the internal and extern handling

from data. [176] Subject of such policies can include aspects of the identity and authentication of certain persons having access to certain kinds of internal data under particular authorization policies. [177] Beyond, the defence against external attacks aiming to gain access to protected data sources, either by exploiting defective software or access controls as well as taking advantage of authorized persons to get authorized access.[178] On the contrary, privacy concerns mostly arise from the data process itself with applications running over not known sources of data with progressive algorithms with expecting unknown result able to represent harsh violations. The implementation of appropriate safeguards against new kinds of violation is also not determinable at this point. Insofar, a differentiation is noticeable with the classification of internal and external actions meaning that appropriate safeguards have to direct against the root of these attacks. As result, the following issues and suggestions for appropriate safeguards focus on the protection of internal processes.

Appropriate technological and organisational measures

First step here is to outline the demanded scale of protection for compliance. Concerning the question of what scale has to be applied, the general data protection regulation requires to take into account all appropriate and proportional means that are necessary to ensure data protection in regard of the societal function of the used technology[179].

1. Nature, scope, circumstances, purpose of the data process as well as the probability and weight of the risks

According to Article 32 GDPR for the security level, the following criteria shall evaluate occurrence probability and weight of specific risks in relation to the nature of the technology. As the general data protection regulation intends to be *technological-neutral* that have to be taken into account are not determined and are dependent from the single technology in question.

For the questions of what possible risks should be taken into account, the recital 75 of the general data protection regulation identified the following categories:

- physical, material or non-material damage
- discrimination
- identity theft or fraud
- financial loss
- damage to the reputation
- loss of confidentiality of personal data protected by professional secrecy
- unauthorised reversal of pseudonymisation and anonymization
- or any other significant economic or social disadvantage; where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data

[176] Big Data and Privacy p. 34
[177] Big Data and Privacy p. 34
[178] Big Data and Privacy p. 34
[179] ZD 2017, 59

- reveal racial or ethnic origin, political opinions, religion or philosophical beliefs, trade union membership, and the processing of genetic data, data concerning health or data concerning sex life or criminal convictions and offences or related security measures
- analysing or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles
- process of personal data of vulnerable natural persons, in particular of children
- processing involves a large amount of personal data and affects a large number of data subjects.

If one of these risks are likely to occur "should be determined by reference to the nature, scope, context and purposes of the processing"[180].

In relation to Big Data application, the *harm-based approach*[181] should be interpreted in way, that the data controller shall not only find solutions adjusted to the possible harm, but to always comply with the data protection principles and create a condition in which the data subject is able of exercise his rights[182].

2. State of the technology

The effort required comprises those technologies that are available, acquirable, implementable, proven and usually used in the concerning domain[183].

3. Implementations costs

The data controller shall be engaged to realize means with unproportional effort addressing those means where the relation between the economic effort (material costs, time costs and human performance/workforce) and the benefit for the data protection are reasonable.

Insofar, the following recommendation offers a possible approach to evaluate the data processes on basis of particular criteria:

1. Outline possible harm
2. Evaluate the probability of occurrence, the possible harm and the degree of harm
3. Implement appropriate safeguards according to the aforementioned criteria

5.5.1.2. Privacy Protection Goal

This approach, in which the private individual's point of view play a key role, considers the protection goals as central element for deriving requirements to be complied with in system design, as well as for identifying risks, countermeasures and in an evaluation perspective. Besides the well-known security protection goals named "Classic CIA Triad" (consisting of

[180] Recital 76 of the GDPR - https://gdpr-info.eu/recitals/no-76/
[181] *„Harm-based approach"* means to implement protective measures corresponding to the level of risk
[182] ZD 2015, 351
[183] Kühling/Buchner Art. 25 Rn. 21

confidentiality, integrity, and availability), three further specific privacy protection goals are encompassed: unlinkability, transparency and intervenability. Protection goals promote the balance of the following privacy and security requirements against other protection goals:

- **Confidentiality,** which refers to the protection of the information from disclosure to unauthorised parties. It can be ensured by measures/tools like encryption, enforcing file permissions and access control list to restrict access to sensitive information;
- **Integrity**, which lingers over the protection of the information from being modified by unauthorised parties. Commonly used methods to protect data integrity includes cryptography, hashing the data received and comparing it with the hash of the original message, use existing schemes such as GPG (GNU Privacy Guard) to digitally sign the data;
- **Availability** of information, which dwells upon the need to ensure the access to information by authorised parties when needed, at the right times. Data availability may be ensured, for instance, by backup, redundancy, off-site location ready to restore services relate to data centre;
- **Unlinkability**, aiming at separating data and processes, in order that processes are operated in such a way that the privacy-relevant data may not be linked across privacy domains or used for a different purpose than originally intended. The minimisation of possible infringements to the individual's privacy is connected to the minimisation of processing of personal data or, in case that data processing takes place, to the minimisation of the possible linkability and actual linkages. It may be obtained for instance by applying effective anonymisation, by separating data that are processed for different purposes, avoiding central points where personal data are or could be collected. This is coherent with the protection of the available data against misuse, where the focus is on the security protection goals.
- **Transparency**, directed to grant an adequate level of clarity of the personal data processes, including all privacy-relevant properties and actions, so that at any time it is possible to understand and reconstruct the collection, processing, and use of the information, both actual and planned. A sufficient level of transparency is a prerequisite for all kinds of control and intervention. Information has to be provided in form and extent adequate to the recipient of the information: in relation to different user groups, different ways of information concerning channels, granularity, language, etc., can be opportune.
- **Intervenability**, functional to assure that parties involved (in particular, data subjects, operators, and supervisory authorities) are able to interfere with the ongoing or planned data processing, including, if necessary, putting in place corrective measures and counterbalances, like data erasure, blocking or destruction, shutting off the system. Data subject's intervenability implies also the right to: i) withdraw consent, ii) obtain rectification and erasure of data; iii) lodge a claim or to raise a dispute to achieve remedy.

5.5.2. Key principles, legal evaluation and assessment of technologies in AEGIS

The legal evaluation of AEGIS technologies has to start with reflecting on the aim of the project, being the presence of a legitimate aim the first requirement for the lawful data processing.

As indicated by the DoA and reminded in D9.2, data processing in AEGIS is functional to create a curated, semantically enhanced, interlinked & multilingual repository for public & personal safety-related Big Data, delivering a data-driven innovation that expands over multiple business sectors and considers structured, unstructured & multilingual datasets, rejuvenates existing models and facilitates organisations in the PSPS linked sectors to provide better and personalised services to their users.

By delivering services addressing the main challenges of cross-domain & multilingual applications through data identification, collection, harmonisation, storage & utilisation, the project aims to generate value and renovate PSPS sector. AEGIS technologies positively influence the welfare and protection of the general public and of individuals through prevention and protection from dangers affecting safety such as accidents or disasters. In this perspective, the AEGIS solutions is aligned with the general interest and common good.

In fact, the project contributes to face some of the main PSPS' challenges, consisting in: i) the lack of data discoverability and on the lack of a common structure and semantic model even for data that bear the same information type and come from similar sources; ii) the lack of data and knowledge sharing mechanisms that in the case of safety issues are important to be properly exploited in order to timely disseminate key findings and promote the adoption of validated solutions. Project solutions, by introducing new business models through the breed of an open ecosystem of innovation & data sharing principles, will enable the creation of value chains towards more accurate risk models and proactive thinking and will revolutionise semantic technologies in Big Data, Big Data analytics & visualisations as well as security & privacy frameworks.

This aim is not only lawful, but also implies a set of positive impacts both for the society (both in terms of economic growth and of enhanced public security) and for the individual (mainly in terms of improved safety and well-being). The set of benefits derived from AEGIS data collection and processing will strengthen value generation for PSPS sector and includes, as reported in D9.2:

- Unified representation of knowledge;
- Accelerated, more effective & value packed cycles of intelligence extraction & of services & applications development;
- Introduction of novel business models for the data sharing economy & establishment of AEGIS as a prominent Big Data hub, utilising cryptocurrency algorithms to validate transactions & handle effectively IPRs, data quality & data privacy issues though a Business Brokerage Framework. Besides capturing a portion of the total addressable market, AEGIS is also expected to enlarge it by creating additional uncaptured value based on small data integration in typical Big Data repositories & algorithms.

In addition to such range of benefits, other advantages arise, in particular by facilitating and promoting the collaboration in PSPS related domains, including public sector, insurance,

environment, health, automotive, smart home, etc. AEGIS, in fact, will facilitate all companies and organisations in the PSPS linked sectors to provide better and personalised innovative services to their users, eventually of cross-domain nature and leveraging the plethora of data sources (from other domains) which, by adequate processing and combination, could further enhance and add value in the baseline services, and thus will allow smart collaborations for maximising the value offered to the end users.

In defining and, at a later stage, assessing and certifying the privacy-friendliness of AEGIS, the findings expressed by the European Group on Ethics in Science and New Technologies play an important role. Such findings suggest to go beyond the traditional drastic trade-off between two goals, security/safety and freedom (including the right to privacy).

Analysing the trade-off narratives, we can see that some doctrine considers security as requirements of the state to protect the lives, welfare and basic freedoms of all citizens, and requires some trade-off between such rights to be protected and the freedom rights (including the right to privacy). Another doctrine argues that, being new technologies connected to competitiveness, jobs and economic growth, this requires to 'trade' away freedom rights, both at the policy level, for removing hindrances to the success of particular enterprises (premised on certain uses of Big Data like in AEGIS) and at the individual level, for exploiting the opportunities provided by such companies, especially online services.

The EGE Group believes that these framings underestimate the difficulty associated with the sensitive equilibrium between freedom and security/safety and constrain the reasoning, corralling it towards limited options and avenues, whereas it is necessary to open up new possibilities for thought as well as for individual and collective actions.

First of all, the EGE Group remarks that human dignity, which is intimately associated with freedom and responsibility "is the core principle of the European moral framework, and as such it cannot be traded off".

Given this, **the right to privacy and the right to data protection**, or the right to information and transparency, are not absolute rights. Therefore, such rights **must be balanced against other rights** and balanced against the rights of other persons or groups. Some kind of balancing, weighing, or choice between priorities is always necessary, in the meaning of need to find an equilibrium between rights of persons, on the one hand, and rights among persons, on the other hand.

In this regard, a rich jurisprudence of the European Court on Human Rights and the Court of Justice of the European Union[184] (ECJ/CJEU) has repeatedly stated that a balancing exercise with other rights is required when applying and interpreting Article 8 of the Charter of Fundamental Rights, setting forth the right to the protection of personal data.

[184] E.g. CJEU, Joined cases C-92/09 and C-93/09, Volker and Markus Schecke GbR and Hartmut Eifert v. Land Hessen, 9 November 2010

This need for equilibrium and balancing is important for the legal evaluation and assessment of the AEGIS technologies. As regards the opposite interests relevant for AEGIS technologies, they are, on the one hand, economic growth, in conjunction with public safety, well-being and personal security, and, on the other hand, the right to privacy and the right to data protection. Personal data collected within the three AEGIS demonstrators (e.g. by the use of tracking technologies) will be immediately anonymised through local dedicated services for anonymisation and filtering of data. These services will allow to process, anonymise the data and strip them of any private or sensitive information, on a local environment before uploading to private containers in order to avoid communication of any personal data outside of the data provider infrastructure. Therefore, AEGIS platform will not collect any personal data.

The AEGIS system overview and description of technologies was inserted in D9.2, Chapter 2, to which reference is recommended. As regards AEGIS demonstrators, the reference has to be made to D5.1, which provides a snapshot of the three AEGIS Data Value Chain Early Community Demonstrators from three different Public Safety & Personal Security (PSPS) domains, namely (1) Automotive, (2) Smart Home & Assisted Living, and (3) Insurance within WP5.

Here it is important to start a preliminary assessment of AEGIS solutions from a legal, privacy, data protection and ethics viewpoint.

Besides the local dedicated services for anonymisation and filtering of data, it is important to remark that in the framework of Data Aggregation and Harmonisation Layer, and in particular of AEGIS Data Value Chain Bus and of its annotations will serve not only for delivering robust, flexible and tailor-made data handling operations and semantic tagging, but also for tagging data with different policies: the data policy library will be used to specify the visibility in terms also of security and privacy/trust levels, as well as of IPR clearance, of each dataset. This data tagging will be based on the Data Policy Framework.

Furthermore, the produced output of the Data Aggregation and Harmonisation Layer is going to be stored in a public or private repository, depending on the type of the data and the policies of the corresponding SME/enterprise/organisation (in particular, during project implementation, this regards the demonstrators).

This is very important both from a privacy perspective and from an ethical perspective, taking into account also IPR issues: decisions on where to store the output/ harmonised data (within the public repository, or private repository, in the meaning of internal repositories Aggregated Local Linked Data Space - ALLDS) will rely also upon security and privacy/trust level and on IPR issues of each dataset (or even dataset element). In other terms, the selection of the repository for the storage depends on the applied disclosure and IPR policy.

Considering that in AEGIS, information exchange among the SLOD space and the private repositories of each SME/enterprise/organisation is going to be supported and that publication/consumption of the produced linked data is going to be realised in real time, AEGIS Consortium’s efforts in designing AEGIS platform has to be directed to make these operations compliant with the applied disclosure and IPR policy and to take safeguard measures to avoid or minimise any privacy risk or IPR infringement regarding the data stored in the internal

repositories. This safeguard measures will be facilitated by the fact that this layer is interconnected with the microservices repository and the Data Policy Repository.

Finally, as regards the Business Intelligence and Analytics Layer, AEGIS Consortium will ensure that privacy-friendly and IPR-preserving modalities and tools are adopted when applying to private harmonised data stored into each private repository ALLDS and, notably, to produced linked data resulting from the information exchange among ALLDS and the SLOD space. The Partners will put great attention in defining how and where algorithms (such as Classification and Text Analysis Algorithms), data analysis techniques and Big Data solutions (like Hadoop, Spark, Storm, Flink) can be applied for the extraction of linked data analytics from such data stored in private repositories or produced linked data with them.

Another consideration concerns the propagation to the SLOD space of the achievements generated by the analysis, which makes this knowledge re-usable in the future and leading to the design of advanced customised solutions. Such a propagation has to be aligned with the Data Policy Framework's disclosure and IPR policies, considering the type of the data and the policies of each SME/enterprise/organisation whose internal dataset was used for the extraction of data analytics. This notably applies in case of linked data resulting from the information exchange among the SLOD space and the private repositories of an SME/enterprise/organisation. On this, special attention should be paid when the core offerings, which will be made available by the top level of the layer, will be offered to the various stakeholders.

In this perspective, the following elements are relevant. First of all, a key role will be especially played by the mechanism to be implemented by Business Broker, for resolving Data Policies of each dataset under request and determining how these can be exchanged between different organisations. AEGIS will follow the notion of a "virtual currency" (in terms of "points") that will be used to safeguard the proper data sharing principles of the platform and will make use of blockchain technology. Secondly, also the Open API Communication sub-layer will be useful for the aforementioned purpose, being responsible for monitoring the usage and verifying that each transaction with the Business Access Layer is verified and thus accepted or rejected.

As underlined here above, it is evident that the set of services comprised by Big Data research in AEGIS opens promising avenues in terms of competitiveness, jobs and growth. AEGIS technologies (technologies of traceability, on-line applications, machine to machine communication, cross-correlation data analytics, predictive analytics and algorithms, etc.) touch not only on new ways to produce growth but also on new ways to produce knowledge, notably "intelligence" and scientific knowledge, as well as opportunities for the individuals. AEGIS intelligence-driven solution fuelled by Big Data analytics represents a powerful tool for identifying trends, patterns, or relationships among data, for improving the individuals' quality of life safety and well-being, as well as for strengthening public security.

Nevertheless, as mentioned, they may give rise to ethical and privacy dilemmas. What is more important? Competitiveness, growth and jobs, public safety and personal security or privacy, data protection, informational self-determination, and individual freedoms?

However, rather than reasoning through a drastic trade-off paradigm, AEGIS partners prefer to concretely operate in line with the prioritisation approach fostered by the EGE's Group, based

on the prioritisation of rights and interests, not giving up on any of the rights and interests and, finally, acknowledging that priorities may differ in different contexts (in particular the different sectors addressed by AEGIS). Following this prioritisation paradigm, AEGIS Ethical, Privacy and Data Protection Strategy, including requirements, has been conceived and will be implemented during project life and in the post-project phase in a way able to guarantee the proper handling of any ethical and privacy issues and the adherence to national, EU wide and international law and directives.

The pillars of it consist of:

1. focusing on notice and on choice (consent) of the data subject prior to data collection;
2. regulatory compliance and continuous legitimate ground of data processing;
3. setting ethical, privacy and data protection requirements to be complied with, elicitated through the Privacy Protection Goal approach, combined with the Privacy-by-Design approach;
4. elaboration of the Data Protection Impact Assessment Methodology, to be delivered in D9.1, including risk analysis and assessment scheme for evaluating the different proposed uses of AEGIS technologies, as well as a set of measures to minimise the privacy and ethics risks;
5. technological fixes, including deidentification/anonymisation/pseudonymisation of personal data: the AEGIS project is going to resort to privacy enhancing technologies, like this CloudTeams Anonymiser (developed by NTUA), and to design and develop its solutions relying upon the "Privacy by Design and by Default" approach.

5.5.3. Ethical, Privacy, Data Protection and IPR Requirements list

The handling and use of personal data is mainly regulated by GDPR, setting out data subjects' rights and providing general rules on the lawfulness and fairness of the processing of personal data. Therefore, in the elicitation of AEGIS ethical, privacy and data protection requirements, references to it will be made. Nevertheless, considering the chosen holistic approach in setting these requirements, we considered other legal instruments applicable, such as the European fundamental rights framework and the national legislations applicable on a case-by-case basis, as well as ethics.

This requirements list clearly lays out a first guideline on how to conceive, develop and use AEGIS architecture and tools, without forgetting checkpoints. Anyhow, the list reflects the insights on the basis of current project progress and, in case of need, may be further refined or revised in a later stage of the project, as the AEGIS architectural design develops.

The main AEGIS ethical, privacy and data protection requirements are as follows.

Number	Short name	Description	Assessment method	Phase	Notes
EPR.1	**Legitimate aim & purpose limitation**	This requirement implies that: i) AEGIS system and technologies have to serve a specific, explicit and legitimate aim; ii) the data have to be collected for such a purpose and not further processed in a way incompatible with that purpose; iii) adequate safeguards against misuse have to be taken. This requirement is also quoted by the DoA (Section 5.1.1): "No data collected will be sold or used for any purposes other than the current project".	DoA, D1.2 itself and D1.3, where AEGIS Methodology is respectively defined and updated, D2.1 and D2.3, where Data Policy, Data handling and Analytics Methods are respectively elaborated and refined, WP4 deliverables (D4.1-D4.4), which refer to AEGIS Platform in each of its improved releases; D9.2, Mid-term and Final Ethics Reports	All	GDPR provisions refers to the legitimate purpose with substantially unchanged formulation in respect to the previous Data Protection Directive. An extended analysis of the purpose of data processing in AEGIS is reported in D9.2
EPR.2	**Propotionality and data minimisation, including anonymisation**	The data minimisation principle is set forth by GSOR and is also quoted by the DoA (Section 5.1.1): "A data minimisation policy will be adopted at all levels of the project and will be supervised by the Ethics Panel. This will ensure that no data which is not strictly necessary to the completion of the current study will be collected". The benefit potentially resulting from the use of that kind of data has to be clear. This requirement also implies adopting anonymisation as much as possible. The de-identification of datasets has to occur since the beginning of the processing: AEGIS datasets have to be stripped of any direct identifiers and, in addition, adequate technical and organisational safeguards have to be taken for mitigating the risks of re-identifying the individuals. In the same perspective, this requirement implies minimising linkability and linkage: efforts have to be done to minimise possible linkability and actual linkages. Fostering unlikability in this way will reduce the risk of data breach and allow to safeguard the securing of the anonymity of the datasets.	D2.2, D2.3, D3.1-D3.5, D4.1-D4.4, D5.6, D6.3, D6.5, mid-term and final Ethics Reports, D9.1	All	The principle of proportionality is expressly recognised by the Recital 4 of the new Regulation: "The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality". AEGIS Data Policy framework addresses also the issue of privacy and data anonymisation through specific micro-services to be developed.

		Regarding anonymisation, it is necessary to comply with what the DoA states: "The data to be stored in the platform will anonymised and held securely using state o the art encryption methods". Tools like the CloudTeams Anonymiser (developed by NTUA), allowing real-time efficient data anonymisation with cross domain scalability, have to be widely and timely adopted and used.			
EPR.3	**Data storage/retention minimisation**	The key rule is that "Personal data must be… kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed". After the Court of Justice' annulment of the Data Retention Directive, reference has to be made to each legal system concerned, which has its own rules on data retention. Therefore, data retention period relevant for AEGIS' demonstrators are those respectively stated by the legal system coming into relevance (e.g., for the insurance demonstrator, Italian regulatory framework). Access to the database has to be allowed only to authorised personnel, whose access is controlled through secure authentication techniques.	D2.2, D2.3, D5.6, D6.3, D6.5, Mid-term and Final Ethics Report	All	It is necessary to comply with what the DoA states: "After the end of the project, all collected data that can related to individuals will be deleted from the platform". Moreover, as regards demonstrators, the DoA specifies that "personal data will be used solely for the specific case, and will be completely destructed and removed from the AEGIS system after the case's finalisation". Regarding ancillary /shadow) data, though the plan is to minimise its gathering as much as possible, in case ancillary is obtained during the course of the research, it must be immediately cancelled.
EPR.4	**Avoidance of discrimination , harm and social sorting**	The Consortium has to avoid that AEGIS demonstrators or AEGIS overall system facilitate discrimination (race, gender, age, religion, disabled) or social sorting. Any possible different treatment has to rely on a rationale and project's solutions have to avoid to cause undue or unjustified harm to anyone, including wrongfully stigmatisation.	Mid-term and Final Ethics Report	D, Ex	The European Charter of Fundamental Rights prohibits any kind of discrimination (Article 21).
EPR.5	**Assignment of responsibilities**	The data controller has to be appointed, as well as the data processors and, in case, the data sub-processors. Also the data protection officer has to be designated by the controller and the processor in the	D5.6, Final Ethics Report	D, Ex	GDPR maintains the set of provisions regulating the entities involved in data handling and adds the figure

		circumstances set forth by Article 37 of GDPR. In relation to the role covered, each entity involved in the processing (data controller and data processor or sub-processor) is bound by obligations to be met and principles to be followed. Such obligations ensure that AEGIS data processing conforms to privacy laws and that the data subjects maintain the right to control what information is collected about them, how it is used, who has used it, who maintains it, and what purpose it is used for. Given that the main responsibility for data processing is in charge of the data controller, most duties and obligations are assigned to this figure, whilst the data processor has fewer and limited legal responsibility.			of the Data Protection Officer in the cases outlined in Article 37. Such cases include "b) the core activities of the controller or the processor consist of processing operations which, by virtue of their nature, their scope and/or their purposes, require regular and systematic monitoring of data subjects on a large scale" and the sensible data (e.g. health data)
EPR.6	**Informed Consent**	The data subject's informed, explicit and free given consent to the transmission and processing of their data is one of the criteria for rendering the data processing legitimate. Consent is principally explicit under the legal framework in force and is an important legal basis of lawful processing in AEGIS (particularly as regards sensitive data). Also when not required as legal ground, seeking consent in AEGIS has to be regarded as best practice. The following specific conditions make the consent valid: • Unambiguity: unambiguous expression of data subject's wishes (no doubt should exist); • Specificity: expression must be intelligible and distinctive, referring "clearly, precisely to the scope and consequences of the data processing". This condition is closely related with the next requirement ("informed"). As an example of invalid consents we can refer to blanket consent; • Information: consent has to be based on accurate, full and understandable information of all relevant issues (the nature of the data processed, purposes of the processing, the recipients of possible transfers, and data subject's rights);	AEGIS Consent Form in D9.2, D5.6, Final Ethics Report	D, Ex	This requirement is based on the transparency principle. Article 4, 11) of GDPR defines the "consent of the data subject" as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her". Recitals 32 specifies that it can consist of a written statement, including by electronic means, or of an oral statement, provided that the data subject's behaviour clearly indicates his/her acceptance of the data processing. It is relevant to AEGIS also Recital 33 which states that, being often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection, data subjects should be allowed to give their consent to certain areas of scientific research (or parts of research projects) when in keeping with recognised ethical standards for scientific research. Recital 42 specifies that "…For consent to be informed, the data subject should be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended. Consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment". Recital 54 clarifies that "The

		• Free exercise of choice: the consent has to be freely given, in absence of any sort of intimidation, coercion or risk of negative consequences. This requirement is interlinked with the next requirement; • Possibility of withdrawal: data subject may be able to change is mind and make a different choice at a later time, thus withdrawing the previously given consent and preventing any further processing. Withdrawal may not be retroactive; • Timing: the consent has to be given before the starting of the processing; Consent form has also to be aligned with the applicable national legislation. Each voluntary participant to AEGIS demonstrators has to be provided with the clear information on AEGIS project and on the specific research activity related to the demonstration activity, as well as the information to be collected, how that information will be used and how to exercise his/her rights (e.g. of withdrawal).			processing of special categories of personal data may be necessary for reasons of public interest in the areas of public health without consent of the data subject". However, suitable and specific measures in order to protect the rights and freedoms of natural persons have to be taken. Public health refers to "all elements related to health, namely health status, including morbidity and disability, the determinants having an effect on that health status, health care needs, resources allocated to health care, the provision of, and universal access to, health care as well as health care expenditure and financing, and the causes of mortality".
EPR.7	**Use of private environment/cloud as much as possible**	Being privacy and control more easily retained in a private environment, they should be used when possible for the storage or processing of personal data, in order to retain bigger control of the data being processed.	D3.2-D3.5, D4.1-D4.4	All	-
EPR.8	**Respect for data subject's rights**	The main categories of data subject's rights relevant to AEGIS can be split into two categories: - rights of information - rights of intervention (including rectification and erasure as well as, according to the new Regulation, data portability). This categories relies upon the intervenability protection goal and guiding principles, that encompasses the control exercised by the data subject and the other parties involved in AEGIS processing system. This includes the possibility for them to	D5.6, Final Ethics report.	D, Ex	The entire chapter III of GDPR is dedicate to data subject's right and has to be taken into considerations: it describes transparency and its modalities (Section 1), information and access to personal data (Section 2), rectification and erasure (Section 3), the right to object and automated individual decision making (Section 4) and restrictions to the data controller and processor (Section 5). Transparency is considered fundamental by the new Regulation, that includes the same in the key principles.

		intervene if necessary. The chance to withdraw the consent can be attributed to this category. The first category comprises transparency or feedback of information, which refers to a set of data subject' rights, first of all his right to access the data stored and processed about him.			
EPR.9	**Data Quality, including Data Accuracy and Data Security**	GDPR at Article 5 letter d) expressly refers to data accuracy, stating that "personal data shall be…accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay" (Article 5, letter d). In AEGIS data accuracy has to be connected to the concept of data quality in data sharing and handling: predefined data handling policies have to be able to ensure data quality and trust. Data Quality, in a privacy-driven perspective, also requires Data Security and Integrity. Personal data shall be "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures" (Article 5, letter f of GDPR). According to level of security has to be appropriate to the risk taking into account "the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons" (Article 32 GDPR). AEGIS has to use state-of-the-art technologies for secure storage, delivery, access and handling of personal information, for encryption and anonymisation, as well as for managing the rights of the users. It is necessary to have the complete guarantee that the accessed, delivered, stored and	D2.1, D2.2, D2.4, D9.1, D5.6, Mid- term and Final Ethics report.	D, Ex	As regards Data Accuracy, this requirement relies upon ethical principles. On the other hand, Data Security and Integrity are two aspects encompassed by the CIA Triad protection goals, which, in addition to privacy protection goals, have been considered as essential for AEGIS methodology for the identification of privacy, data protection and ethical requirements. AEGIS data handling policies have to be able to ensure data quality and trust, besides privacy compliance. The quality level will be described by performing the necessary annotations both at dataset and on dataset element level: this has to be ensured by AEGIS Data Policy framework, which will be used upon insertion of any kind of data into the platform.

		transmitted content will be managed by the right persons, with well-defined rights, at the right time. Where possible (depending on the facilities of each organisation) the data should be stored in a locked server, and all identification data should be stored separately. Tools for monitoring anomalies and activate restraint policy if needed should be used. The Data Policy Framework has to detail the security measures and other tools to be used for ensuring data protection and data quality.			
EPR. 10	**Privacy by design and by default**	Security-by-Design, Privacy-by-Design, as well as Security-by-Default and Privacy-by-Default design methodology, has to be adopted in order to minimise the risks of compromising privacy. Efforts should be directed towards compliance with the voluntary standars developed by CEN-CENELEC/JWG 8 'Privacy management in products and services' for implementing data protection by design and by default rules and good practices, and more generally for privacy protection. According to Article 25 of GDPR, the controller, considering a set of circumstances, shall implement appropriate technical and organisational measures: "such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects". Data protection by default ensures "that, by default, only personal data which are necessary for each specific purpose of the processing are processed".	D3.2-D3.5, D4.1-D4.4	R	Privacy by Design is at the core of AEGIS approach for the elicitation of privacy and data protection requirements, whilst data protection by default is coherent with data minimisation requirement.
EPR. 11	**Record of processing activities**	"Each controller and, where applicable, the controller's representative, shall maintain a record of processing activities under its responsibility" (Article 30 GDPR)	D5.6, Final Ethics Report	D, Ex	This is a provision of the reform, that specifies also the information that has to be contained in the recording.

EPR. 12	**Data protection impact assessment**	The need for a data protection impact assessment in AEGIS derives by the DoA (WP9), but is also coherent with Article 35 of GDPR. Article 35 states that "where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data".	D9.1, Mid-term and Final Ethics Report	R, D	_
EPR. 13	**Application scrutiny to local/national boards if required by national legislation concerned**	As regards the demonstrators, "authorisation or notification by the National Data Protection Authority must be submitted, where applicable" (WP9). National legislations provide that data controllers and processors have to register at the competent authorities, in order to be allowed to process personal data, and impose differing national requirements for such a registration/authorisation, ranging from none to extensive authorisation processes. In most Member States registration for transfer to another EU Member State is not required, unlike for cross-border data transfer, where additional or separate requirements may exist (e.g. registration or authorisation or mandatory additions to the standard contractual clauses).	D9.1, Mid-term and Final Ethics Report	D, Ex	Unlike the Data Protection Directive, the new Regulation doesn't provide "for a general obligation to notify the processing of personal data to the supervisory authorities". Such an obligation has to rely on "effective procedures and mechanisms which focus instead on those types of processing operations which are likely to result in a high risk to the rights and freedoms of natural persons by virtue of their nature, scope, context and purposes. Such types of processing operations may be those which in, particular, involve using new technologies, or are of a new kind and where no data protection impact assessment has been carried out before by the controller…" (Recital 89)
EPR. 14	**Confidentiality and access restriction**	People in charge of collecting, using or accessing personal data in AEGIS must be subject to an enforceable duty to keep them confidential and secure. Therefore, a confidentiality clause or agreement should be concluded by all research staff that will be having access to personal data in AEGIS. A closed user group has to be established, composed of only authorised persons, contractually obliged to keep confidentiality and meet data security rules. It is recommended an authentication and authorisation infrastructure in AEGIS.	D9.1, Mid-term and Final Ethics Report	D, Ex	This requirement is ascribable also to ethical principles and to the chosen privacy protection goal as load-bearing method for elicitating and analysing data protection, privacy and ethical requirements in AEGIS. In particular, it is one of the three well-known security protection goals, named "Classic CIA Triad"

		In addition to the technical measures that will be taken in view of ensuring confidentiality, publication of AEGIS result will not reveal the data subjects.			
EPR.15	**Involvement of AEGIS Ethics Advisory Board**	This ethical requirement concerns the need to involve this committee to i) monitor ethical and legal issues in the project and report to the Commission; ii) work closely with the consortium in order to address the ethical and legal issues and data privacy concerns, that may arise from accessing user related information	D9.3, Mid-term and Final Ethics Report	D	AEGIS Ethics Advisory Board is expected to act as a sort of Data Protection Officer internal to the project and it has to periodically report to the Commission on the implementation of the ethical concerns (issues) in project and compliance with applicable national and EU regulations.
EPR. 16	**Set of requirements referring to the voluntary participation to AEGIS demonstrators**	The following requirements apply: i) AEGIS Recruitment Procedures for the selection of the voluntary participants for the AEGIS trials have to avoid any sort of discrimination/social sorting and be assessed by the Ethics Advisory Board of the project; ii) informed consent has to be obtained: partners must inform voluntaries and distribute the consent form, to be signed by each voluntary before trials' operations start; iii) Volunteers' dignity has to be safeguard and direct/indirect incentives for participation must not affect it.	D5.2, D5.6, D9.1, Mid-term and Final Ethics Report	All	_
EPR. 17	**Adequate mechanism and tools for safeguarding IPRs on data artefacts and data usage**	This requirement calls for carefully addressing the data ownership aspect and for effectively handling IPRs of each dataset and dataset element.	D2.1, D2.2, D2.4, D9.1, D5.6, Mid- term and Final Ethics report	D	These requirements refers also to the emergent of the Human Data Interaction (HDI) topic, aiming at putting the human beings at the centre of the data driven industry and thus calling attention to address the data ownership aspect more carefully (e.g. who owns this data captured by the sensors? And who should have access to it?)

EPR: Ethical, Privacy and Data Protection Requirement
R: Research phase
D: Demonstration phase
Ex: Exploitation phase of the AEGIS system
All: all the phases, both during the project and after its end.

5.5.4. Guiding principles and recommendations for AEGIS Data Policy Framework

AEGIS cloud based Big Data solution is going to become a distributed database of secure transactions, removing the need for centralised ledgers, trusted parties copies and manual interventions.

AEGIS data sharing, homogenisation and reusability value chain will allow the exchange and documentation of data in an unanimously understandable manner, facilitating knowledge exchange.

It will therefore rely on collaboration activities of diverse sectors and stakeholders. Therefore, AEGIS will offer services supporting data IPR handling, security and privacy. These services will be able to overcome the limits characterising existing solutions, where data confidentiality, privacy protection and IPRs hinder the ability to exchange information in a trustful and transparent manner.

AEGIS Blockchain powered Security, Privacy, Quality and IPR Data Policy Framework (DPF) will power AEGIS platform with a methodology encompassing aspects related to the exchange of data from business value point of view, focusing on the quality, the IPRs and the privacy of data. Indeed, AEGIS DPF, in its on-going development under T2.2, is exactly devoted to set the appropriate security, data privacy, data quality probing and IPR policies to resolve on-the fly how data can be handled by each stakeholder group, based on its content, its value and peer-to-peer agreements that will be reached between the collaborating entities. Hence, AEGIS DPF under development has to allow the creation of a trustful and rigorous data sharing community, by focusing on data anonymisation and privacy preservation, secure data channels, IPRs on data artefacts and data usage, as well as data quality, going beyond current practises in data sharing and handling.

In this way, organisations, including those previously reluctant, are expected to contribute to AEGIS new data sharing ecosystem around Pubic Safety PSPS related information: stakeholders will be allowed to securely exchange data, on the basis on national, international and business ethics and regulations, and will not incur in the risk of limiting their competitive business advantages.

The cloud based nature of AEGIS infrastructure will made possible the inclusion of new data and knowledge, as well as new sector data and services. Also in this case, the alignment to AEGIS DPF will be requested for enabling producers and consumers to collaborate for the shared value generation and expansion of the overall solution.

AEGIS Data Policy Framework has been driven and will be further driven by the AEGIS Ethical, Privacy, Data Protection and IPR Strategy and resulting requirement list, as provided in this deliverable on the basis of the EU and national legislations, as well as ethical standards. As written hereunder, the DPF is going to be elaborated in T2.2 “Data Policy and Business Brokerage Frameworks”. The DPF and the core methods to be used have been initially addressed in D2.1 “Semantic Representations and Data Policy and Business Mediator Conventions” (M8) and will be further refined in D2.3 “Update on Semantic Representation and Data handling and Analytics Methods” (M18).

Being the DPF strictly interrelated with the practical implementation of the requirements and recommendations set forth in the AEGIS Ethical, Privacy, Data Protection and IPR Strategy, in this chapter we intend to briefly underline its key role and providing some guidelines and insights for its development.

The DPF has been conceived for exploiting the new opportunities arisen in the areas of security and privacy through the use of Blockchain technology. In fact, it is intended to represent a novel method of using this technology and microservices for checking data quality, security, trust and IPRs. As stated in the DoA, "a blockchain can be defined as a digital, chronologically updated, distributed and cryptographically sealed record, of all data transfer activity", which "enables the transfer of digital assets, representing various manifestations of value or possessing inherent value within themselves. It is a secure way to enable such transactions and various other digital activities as everyone can participate, there is no identity disclosure and as already mentioned, manipulation is difficult due to the distributed nature of blockchain".

In relation to AEGIS DPF, two of Blockchain's aspects are particularly relevant: the employment of self-governance of transfer of ownership and the use of cryptography for preserving purposes. The DPF will be used upon insertion of any kind of data into the platform, performing the necessary annotations both at dataset and on dataset element level, ensuring that the accumulated data are fully described with respect to their IPRs, quality and privacy levels.

The semi-automatic negotiation of micro-contract regarding data exchange based on existing IPR schemes will be made possible by the Business Brokerage framework, by exploiting the DPF core methods, notably the IPR annotations, as well as by utilising Blockchain technology. In particular, following the annotations, all data exchanges are going to be supervised by the Business Brokerage service, which will generate on-the fly micro contracts for data sharing between the different data collaborating parties.

The AEGIS DPF is expected to introduce a prototype usage case of blockchain for ensuring the creation of a trusted and secure channel supporting the communication and data exchange between different stakeholders and users of the platform in a distributed environment without the need of intermediaries. The choice of the Hyperledger Fabric blockchain infrastructure and the relevant modules that offer secure communication within the blockchain network to be deployed, offers the ability to overcome security incidents and risks that may occur to AEGIS applications and databases.

The AEGIS platform will manage and process closed data (proprietary data) as well as open data and will allow the exchange of data with different IPR. The former will be stored encrypted in the Security Linked Open Data (SLOD) space, whilst the latter will be stored unencrypted and published under an open data license and will be publicly accessible through the platform. IPR and data sharing agreements through semi-automatic negotiation of micro-contracts utilising Blockchain technology will be based on the predefined data handling policies, schemes and annotations defined in the DPF. They will ensure IPRs on data artefacts and data usage in relation to the data to be contributed to the platform.

In this perspective, AEGIS is expected to use a Blockchain-based IP Model, which makes possible unprecedented forms of transparency in copyright information and management. In fact,

Blockchain technology is expected to allow all users to have clear insights and access to all copyright information on the dataset and on any dataset element and, at the same time, to be able to bring an easier pay out system to IP owners and licensors of data.

A good example to consider is Ascribe "Ownership Layer", which provides a powerful tool allowing proof-of existence on Blockchain for IP and innovation (in AEGIS, for dataset or data element) and makes easier the whole process of licensing and copyright transfer. Ascribe tackles the compelling need for a workable solution to the ownership and attribution issues by ensuring "ownership processing", that makes ownership actions of digital property universally accessible. Ascribe's approach is twofold, being based both on a registry with easy and secure legals and on visibility of data on usage/provenance of the content. It has two components, respectively ensuring IPR transparency and management, based on an ownership registry for easy secure disposition of rights. There is an ownership registry with easy and secure legals, which formalise (via a creator and consumer-friendly Terms Of Service) existing copyright rights on digital objects traditionally difficult to be leveraged, whilst the bitcoin-inspired blockchain serves for securely recording ownership transactions. In the registry it is possible to register a work, transfer ownership, grant licenses, loans and rentals. The registry also provides the time-stamping evidence of ownership actions through bitcoin-inspired blockchain. Ascribe enables to record intellectual properties on the Bitcoin-inspired blockchain, which is used as a distributed database to store the registry records (that track the history of ownership, the so-called "provenance"). It, thanks to the combination with cryptography, is able to make the registry global, robust, and impairment-resistant, whilst shielding the parties' personal identity (thanks to cryptography again). Ascribe has been proven in several domains and is being used both by individual creators and by institutions (e.g. marketplaces, libraries, archives, museums, galleries) and organisations, including new startups.

The following part of this paragraph will outline the main findings and insights relevant for the definition of AEGIS Data Policy Framework in relation to each of the demonstrators.

Demonstrator 1: Road Safety Indicator

In the Automotive Demonstrator VIF engages a number of volunteers to generate **vehicle driving data** and/or **vehicle simulation data** during experiments after having signed an informed consent. Both sources of data are anonymised and do not include any personal information. VIF will provide the collected & anonymised field data as well as the collected and anonymised simulation data to be published at the AEGIS platform for further data analysis and service generation. Both are sources of non-confidential data (technical data) and can be anonymously shared with the AEGIS consortium to be processed in the AEGIS platform to establish data-driven services. However, other sources of data relevant for the automotive demonstrator and the three scenarios (e.g. weather data, traffic data, or map data) will not be generated by the people volunteering in experiments. They will have to be accessed through other data repositories (e.g. OpenWeatherMap for weather data, or OpenStreetMap for map data), which will have their own licenses and disclaimers for using data and services, which have to be studied in detail.

Both vehicle simulation and vehicle usage data can be considered as technical data and are **time series data**.

- Vehicle simulation data is generated through the software running on the simulator, e.g. through CarMaker - a professional software which has been developed for testing passenger cars and light-duty vehicles in real world test scenarios.
- Vehicle usage data is generated by the vehicle data logger, a device developed at VIF and used for research purposes.

While the **data quality** for simulation data is expected to be very high, the data quality for vehicle usage data measured by sensors from the car as well as by sensors installed on the vehicle data logger may vary. There will certainly by missing or wrong sensor values included in the data which have to be eliminated before (in a data cleaning process) exploiting it to establish data-driven services and reports. For instance, GPS data may include wrong values, if a vehicle is driving under a bridge or through a tunnel as a result of reflections. OBD2 data such as vehicle speed or vehicle rpm may include false values from time to time, too, which have to be corrected, e.g. by applying interpolation mechanisms. Furthermore, the data will be interpolated before it can be used for analysis.

The collection of vehicle data from the field – and especially the smart combination of this data with data from other sources – will facilitate the generation of many innovative digital products, third party services and business models. However, such digital services based on vehicle data can only be successful if a critical mass of vehicles shares driving data (in a later exploitation phase of the project). Raising awareness in the society on what kind of data a vehicle generates, processes, stores, and potentially transmits to a third party is a challenging yet crucial task. The '***My Car My Data***' campaign launched by Federation Internationale de l'Automobile (FIA) educates car drivers about the potentials and pitfalls of connectivity. The My Car My Data campaign believes that the driver should be the one deciding if vehicle data should be shared and with whom. Europeans should be entirely free to choose with what party they share their vehicle data in the future (eventually on a market that allows services providers to compete in offering the drivers the most added-value for shared data), unless mandated by law. VIF supports this strong statement of the FIA initiative mycarmydata.eu in its demonstrator, and emphasizes data protection, free choice (of service providers), as well as fair competition (through a variety of service providers) as the three main principles.

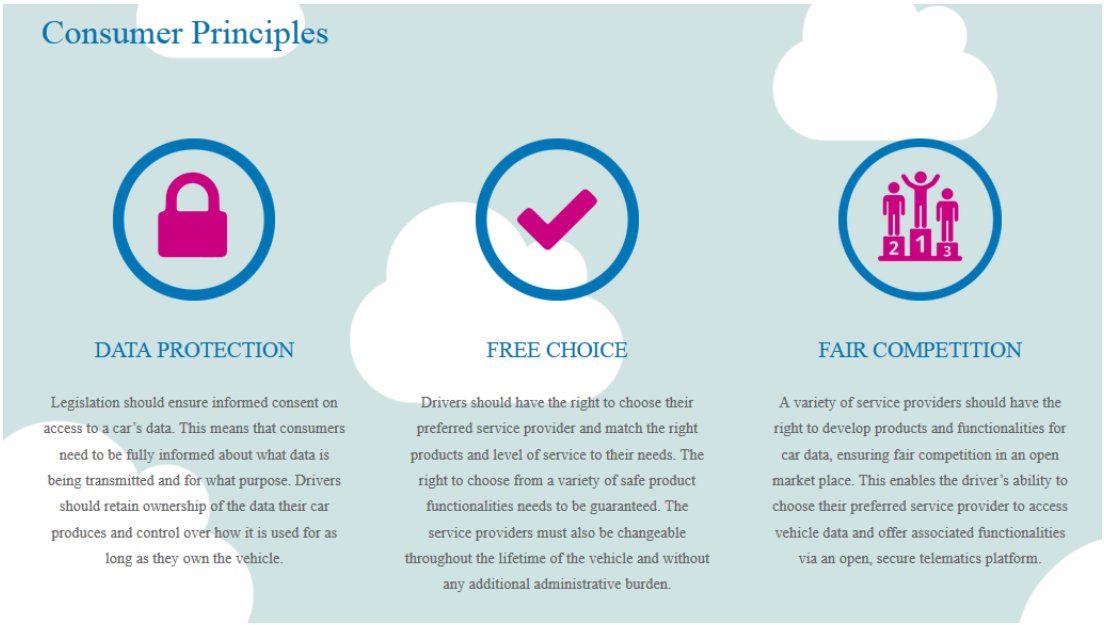

Figure 5-6: Consumer principles for vehicle data sharing (Source: FIA MyCarMyData)

In the project, people participate on a voluntary base in the automotive demonstrator and also share their data voluntarily. Nevertheless, VIF has installed an informed consent procedure to ensure that voluntaries are well-informed on the project's goals, the purpose of data collection, the nature of the collected data, and the services to be developed on top of collected data. These three services are Broken Road Indicator, Safe driving indicator, and Regional Driving Safety Risk Estimator. All three services are the results of a similar data process, (1) manually upload data, (2) transform data using the platform, (3) save transformed data on the platform, (4) identify events in the saved data, (5) save identified events to a new dataset, and finally (6) visualize this new dataset as an overlay on a geographic map.

As state before, access to vehicle data is crucial for novel digital services as well as for new business models based on vehicle data. Regarding (direct) access to vehicle generated data, **car manufacturers** are in a comparably lucky position. However, they were so far not very successful in exploiting this market yet to establish a digital ecosystem. The potential to exploit car lifecycle data for purposes other than driving currently remains almost untapped by automotive OEMs. According to the EU research project AutoMat (AutoMat, 2016), the automotive industry has not yet been able to successfully establish an ecosystem for apps and services equivalent to that of smartphone manufacturers. The project mentions three reasons why OEMs are currently struggling: Brand-specific business approaches dominate, and as a consequence there is a lack of brand-independent car lifecycle data. Current proprietary car services focus on the individual customer, what leads to privacy concerns, and few ideas exist how anonymised car data can be used to establish other services. The implied or required collaboration between OEMs on car data and services is considered risky in terms of competition.

Two recent position papers from **VDA - the German association of the automotive industry** discuss the role of German-speaking car manufacturers towards establishing digital ecosystems based on vehicle data. The first position paper '***Data protection principles for connected vehicles***' (VDA 2014) refers to the continuous transformation of vehicles towards 'connected

vehicles' with a permanent uplink to the internet and the feasibility to connect various data sources for establishing new services. The position paper suggests three principles for VDA members to handle the advancements in connectivity and the new services associated with respect to responsible data handling as well as with data protection:

- Transparency: The members of the VDA strive for adequate information about the data in connected vehicles and the use of these data.
- Self-determination: The members of the VDA are striving to enable customers to determine themselves the processing and use of personal data through various options.
- Data-security: The members of the VDA strive to implement the strong safety culture in the automotive industry also in the connected vehicle.

The short paper closes with a chart of data categories in connected vehicles and their relevance for protection.

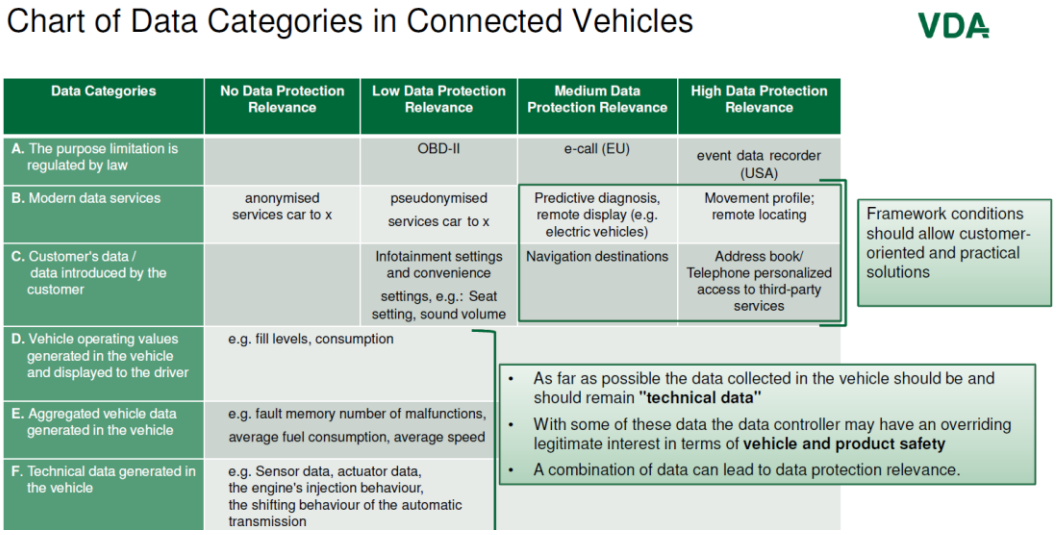

Figure 5-7: Data categories in connected vehicles (Source: VDA)

The second position paper titled '***Access to the vehicle and vehicle generated data***' (VDA 2016) which has been developed in accordance with the '*EU Commission* ***C-ITS platform project final report****'* discusses data-centric requirements for security, privacy, and discrimination free innovation. The C-ITS report cited in the position paper lists five guiding principles to apply when granting access to in-vehicle data and resources:

1. 'Data provision conditions: consent': The data subject (vehicle owner) decides if data can be provided and to whom, including the concrete purpose of the data including an opt-out option.
2. 'Fair and undistorted competition': All service providers should be in an equal position to offer services to the data subject.
3. 'Data privacy & data protection': There is a need for the data subject to have vehicle and movement data protected for privacy reasons.
4. 'Tamper-proof access and liability': Services making use of in vehicle data and resources should not endanger the proper safe and secure functioning of the vehicle.

5. ‘Data economy’: Standardised access favours interoperability between different applications and facilitates the common use of same vehicle data.

According to the VDA report, each OEM holds the role of a system administrator and is hence responsible for the safe and secure transfer of car data to a business to business (B2B) OEM interface. Third parties can access this car data directly over the OEM B2B interface or via neutral servers, which gather the data from the cars. If the access to the OBD2 interface will be limited by OEMS, many business models based on vehicle data might be endangered.

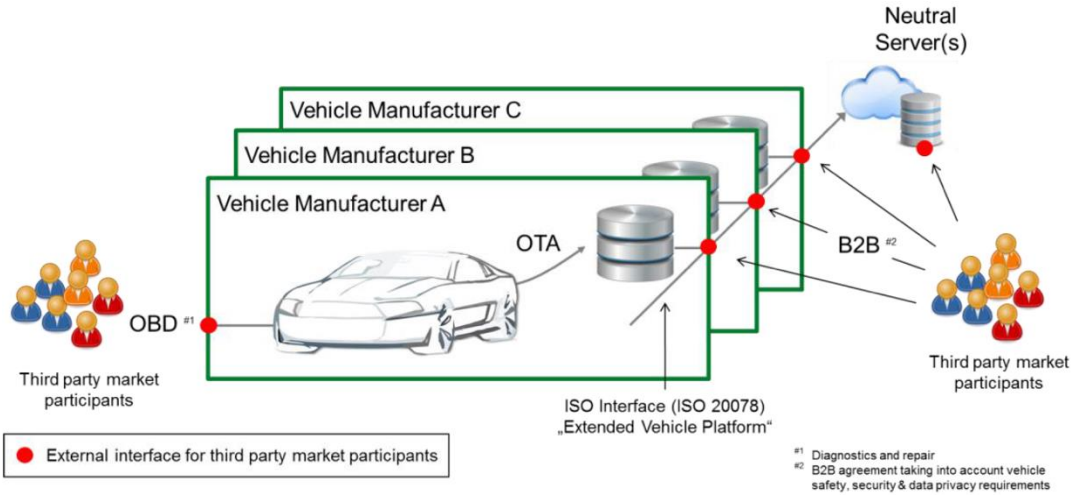

Figure 5-8: Access to the vehicle (Source: VDA)

The VDA report summarises that the vast majority of vehicle generated data is raw technical data, which is used locally within the vehicle and never stored. The VDA has defined four usage categories for vehicle generated data: Data for the improvement of road traffic safety, data for cross brand services, data for brand specific services, data for component analysis and product improvement, and personal data.

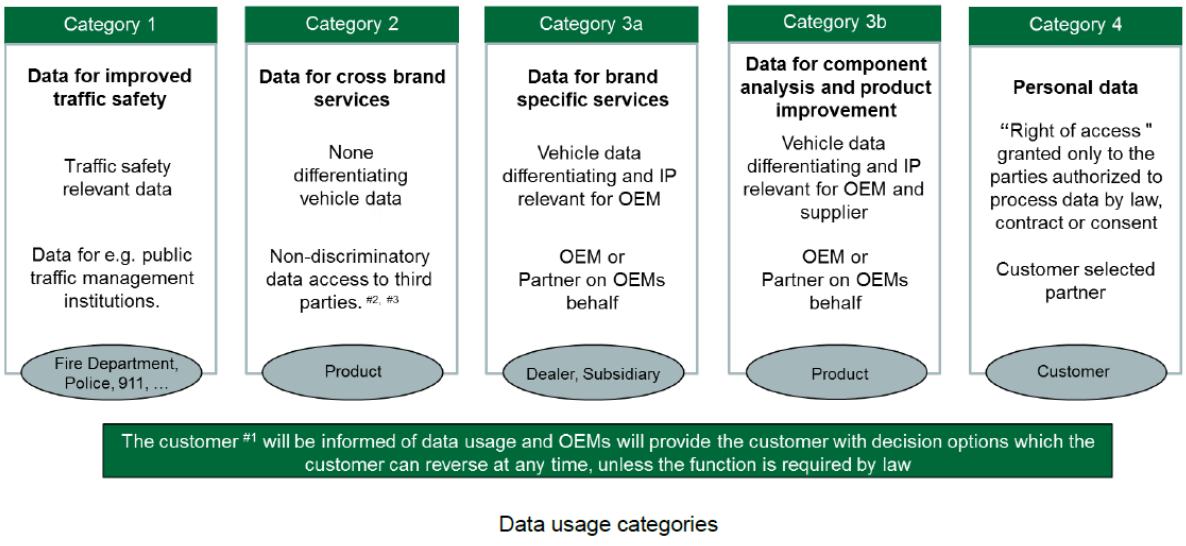

Figure 5-9: Data usage categories (Source: VDA)

Demonstrator 2: Smart Home and Assisted Living

The details regarding the data policy framework for the AAL and Smart Home demonstrator have already been defined in the previous deliverable of WP1. Here we reiterate the information, with some minor updates regarding the services and data sources associated with/to the demonstrator.

The main objective of Smart Home and AAL demonstrator is to correlate information coming from ubiquitous IoT devices (building sensors, smartphone data and wearable devices) with open datasets towards the extraction of meaningful services. As Internet of Things (IoT) becomes a growing reality, more ubiquitous devices are embedded in our daily lives, serving us in a broad range of purposes in everyday life from: personal healthcare to home automation to location based services.

IPR considerations

These devices primarily collect data that is about or produced by people, be it the energy footprint of an individual's home or her location and other situational context. As this unprecedented amount of data is collected, we are challenged with one fundamental research question: **who owns this data and who should have access to it?**

Specifically, the emergent of the Human Data Interaction (HDI) topic which aims to put the human at the centre of the data driven industry, calls attention to the IoT community to address the data ownership aspect more carefully. In this note, it is fundamental within the project to clearly clarify IPR issues on **data artefacts** and **data usage** for Smart Home and AAL demonstrator. The analysis should not only address the demonstrator specific requirements, rather to exploit the potential of commercial exploitation of the AEGIS platform towards providing home automation and AAL services.

Nowadays, the main IP rights in relation to data are copyright, database right and confidentiality. Patents and rights to inventions can apply to software and business processes that manipulate and process data, but generally not in relation to data itself. Trademarks can apply to data products, but again, generally not in relation to the actual data. IPR in relation to data is of uncertain scope at the moment, and the law in this area is likely to continue to develop in the coming years: historically, IPR development has followed the commercialising of innovation and as the value of Big Data rises, so likely will the IP rights underpinning it.

In essence, the owner of machine-generated data (MGD), which covers virtually all of the IoT, is the entity who holds title to the device that recorded the data. In other words, the entity that owns the IoT device also owns the data produced by that device. However, it's not always clear that whomever has possession of the device and/or its output data actually "owns" it. Data may be owned by one party and controlled by another. Possession of data does not necessarily equate to title. Possession is control. Title is ownership. Referred to as usage rights, each time data sets are copied, recopied and transmitted, control of the data follows it. Conversely, transfer of ownership requires a legal mechanism to convey title.

Contract rights in relation to data are technically entirely separate from IPR and their value was confirmed in a UK High Court case in 2006 where the judge said that an owner of data: *is entitled in principle to impose a charge for use of its data by users whether or not it has IP rights in respect of that data.*

By taking into account the high level principles towards managing IPR issues for data coming from IoT devices, we are defining as data controllers: the data scientists of the company providing the equipment to the end users. This is the common case in industry as the IoT solution provider is also the controller of the data streams (IPR on data usage).

Data streamed off devices in home has so far been handled by companies which treat the users as clients **only with no say on how their data should be used**. However, we believe that given a transparent framework and regulations, many users would be willing to share their data. As such, we propose the notion of Data Market as an instrument to enable users to share their personal data locally and globally with monetary benefits, i.e., an individual can trade data produced at her personal space with interested business entities. Such model is currently being considered by a number of data exchange companies, where monetary incentives are offered to end users for correcting erroneous sensor data. The challenge in this case is how to design future infrastructure so to make users aware of the commodity of their data along with the risks of sharing it.

Data Quality Considerations

Extracting high-quality and real data from the massive, variable, and complicated data sets becomes an urgent issue. Data quality is not necessarily data that is devoid of errors. Incorrect data is only one part of the data quality equation. Amongst others, there are several conditions that contributed to the data quality problem such as lack of validation routines, data valid but not correct, mismatched syntax, formats, and structures, unexpected changes in source system, lack of referential integrity checks, poor system design and data conversion errors.

High-quality data are the precondition for analysing and using Big Data and for guaranteeing the value of the data. Figure 2 shown the first five attributes (i.e. Accuracy, Integrity, Consistency, Completeness and Validity) generally pertain to the content and structure of data, and cover a multitude of sins that we most commonly associate with poor quality data: data entry errors, misapplied business rules, duplicate records, and missing or incorrect data values. But defect-free data is worthless if knowledge workers cannot understand or access the data in a timely manner. The last two attributes (Timeliness and Accessibility) above address usability and usefulness, and they are best evaluated by interviewing and surveying business users of the data.

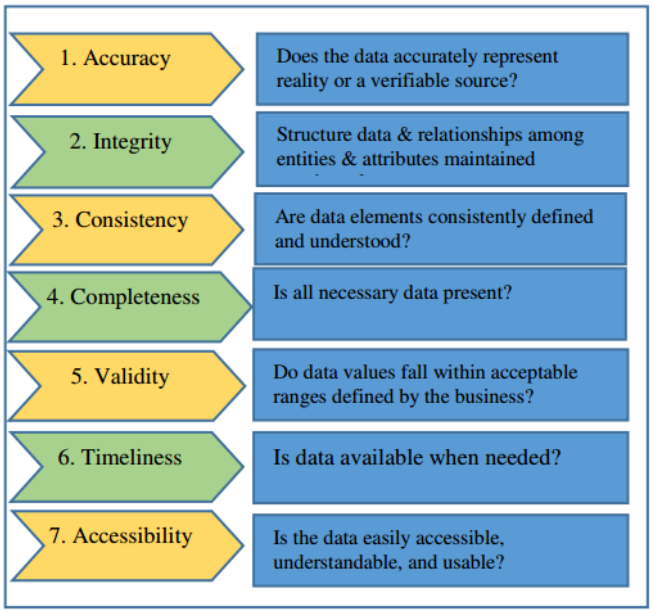

Figure 5-10: Data Quality Attributes

Data quality is an essential characteristic that determines the reliability of data for making decisions in Smart Home and AAL demonstrator. High data quality is:

- Complete: All relevant data such as in home environment data, wearable devices datasets and open datasets are available.
- Accurate: Common data problems like misspellings, typos, and random abbreviations have been cleaned up.
- Available: Required data are accessible on demand; users do not need to search manually for the information.
- Timely: Up-to-date information is readily available to support decisions.

It is very important to define the associated microservices in AEGIS platform that will meet the demonstrator requirements towards accessing high quality data. This is actually a process highlighted in the definition of the high-level usage scenarios for the associated demonstrator.

Considerations on Data Privacy, security and trust

Considering the nature and the type of the datasets available in Smart Home and AAL demonstrator it is mandatory to adopt Data Privacy, Security and trust policies towards handling the data streams required for PSPS services. Data privacy is suitably defined as the **appropriate use of data**. Privacy assures that personal information are collected, processed (used), protected and destroyed legally and fairly. On the other hand, data security provides protection for all types' information, in any form, so that the information's confidentiality, integrity, and availability are maintained.

We have highlighted the importance of data privacy and security in the description of the high-level usage scenarios for the demonstrator. Visiting the AEGIS platform, the data scientist (of the IoT equipment provider) creates an organisation profile for the company and then creates a

project marking it as a "private" project. In the project, the data scientist may invite other users providing classified access to the same project, giving them partial access to the data. At the same time, invitations may be delivered to external collaborators with specific rights to upload data to the project's repository, indicating also the data structures expected and the schemas that need to be uploaded. In line with classified access to the datasets, we are highlighting the importance of anonymisation over the streams of personal data, to meet privacy and security objectives. A list of tools should be provided by the AEGIS platform, to ensure that a prompt anonymisation will be performed over the streams of data by the data scientist.

Finally, some further considerations regarding the aspects that the "Data Policy Framework" will have to take into consideration as regards the Smart Home and Assisted Living Demonstrator.

Purpose of the processing of personal data

The purpose of processing the streams of personal data is defined through the specification of the high-level usage scenario reported in this deliverable. More specifically, the scope of Smart Home and AAL demonstrator is threefold.

- Monitoring and analysis of an individual's well-being conditions, physical activity, positioning and wearable information and external environment data (e.g. weather, crime, news, social media), towards provision of a service for personalised notification and recommendation system for at-risk individuals, including notifications for carers.
- Additional service pertaining monitoring and analysis of weather, indoor environmental conditions, energy and operational device data towards the provision of a smart home application, which can be offered by care providers to at-risk people for increased indoor comfort and welfare.

A more detailed list of processing mechanisms is defined by UBITECH as part of user requirements

Origin of personal data and its collection method

Smart Home and assisted living demonstrator evaluation requires the installation of equipment and usage of wearable devices. The list of datasets to be examined in this demonstrator were presented in Figure 5-5.

Towards gathering the required information, a limited number of installations will be performed as part of the demonstrator. Namely we are considering the installation of:

- PIR sensors for tracking occupancy information;
- Luminance, temperature humidity, VOC and CO2 sensors for acquiring information about illuminance, indoor temperature and humidity and IAQ levels;
- HVAC controllers and actuators (smart thermostats, actuator interfaces) towards acquiring information about HVAC operation;

- Lighting devices controllers and actuators (dimmers, 0-10V actuators, smart lamps) towards acquiring information about lighting devices operation;
- Smart metering equipment, clamps and plugs towards gathering information about energy consumption.

In addition to in-building equipment installation, datasets will made available from:

- Smartphone devices with build-in sensors as a daemon is running to track accelerometer, gyro and GPS data;
- Wearable devices (activity trackers) with build in sensors to track activity and health related parameters.

Finally, self-reporting data about personal health conditions is an option considered in the project. As these are sensitive data, a data specific ethics handling methodology has been presented above.

Technological component of the overall AEGIS system processing personal data relevant in this demonstrator

Having defined above 1) the list of datasets (personal data) available in Smart Home and AAL demonstrator and 2) the purpose of processing these datasets, we are further highlighting the list of technological components of the AEGIS system to support the analytics process. The overall analysis takes into account the data value chain schema towards the definition of the associated technological component.

- The end users should be able to **upload data to the project's repository**, indicating also the data structures and the schemas that needs to be uploaded. A service should support end users to easily upload datasets examined in Smart Home and Assisted Living demonstrator. Different ways of uploading datasets should be considered in the project as some data are initially uploaded only for experimentation purposes (e.g. some samples of datasets), while also there is the option to connect data to the platform through a project specific API endpoint.
- The platform should support **tools** to **anonymise, clean and transform data** in order to meet the expectations of the data scientist. **Anonymisation** and data privacy preservation methods are required so that no sensitive data is transferred to the platform. Furthermore, **cleansing and normalisation services** should be supported to ensure high quality datasets availability. Finally, **semantic curation** of the datasets should be supported considering the nature of the applications developed in the project (PSPS services).
- Having all the data in one place, the data scientist is now able to invoke several analyses, choosing which data to combine as well as the algorithms to utilise. Those come out of a

predefined **algorithms library,** while it is also possible for the end users to **upload their own algorithm** and conduct an analysis. The overall results (datasets and analytics results) are then presented in a dashboard that visualised the outputs of the analyses, where access can be provided to any member of the project, while the results can be also exported in various formats.

- What is especially interesting is the option to **export the data through an API** that also allows the analyses to be executed remotely by providing an external signal. This can come either from an external stimulus, such as a weekly call from an external system, or from triggers specified and enabled in the AEGIS platform, such as the updating of a dataset or the occurrence of an event.

The definition of the AEGIS technological components towards handling personal datasets is in line with the overall data value chain definition in previous section.

Risks identified when dealing with personal data

Having presented above the overall framework for Smart Home and AAL demonstrator and the usage of personal datasets, an indicative list of risks when dealing with personal data is presented in the following table:

No.	Description
1	Loss of Privacy Control. Participants will be monitored and their personal data will be collected.
2	Difficulty in ensuring the security of shared Personal Data
3	Storage and Process of personal Data - Confidentiality
4	Lack of transparency
5	Delegation of Control Privacy - Incidental Findings
6	Improper use of IT equipment

Demonstrator 3: Insurance Sector. Support, Warning and Personal Offering

The Insurance Demonstrator will exploit the AEGIS technologies to achieve a more personalised mode of calculation of the risk associated with each customer, to provide alert systems and to adopt new insurance models.

In this demonstrator, as in the others, volunteers will be involved and their personal data will be collected and handled.

A unique interface will enable to access and analyse information coming from diverse and heterogeneous data sources (geospatial information, social media, broadcasted news, etc.) and will be combined with the in-house datasets of the Company.

In this way, HDI will deliver personalised intelligence for preventing specific catastrophic incidents related to people lives/assets through early warning notifications. These notifications take the cue from the incidents and threat situations (e.g. severe weather conditions) and will be diffused by the insurance company, together with different recommendations, based on the anticipated impact that the given incidents and threat situations might have, on the basis of an expert's model that store and evaluate their criticality and severity.

The proposed Scenarios (see D5.1 and D5.2) will apply semantic analysis to the gathered data and will realise the intelligence extraction from multiple data sources. There will be investigated multiple types of threats and refer to the location/asset type, besides capitalising on the already available open data knowledge.

The warning notification will target a restricted group of customers by filtering out the ones belonging to groups that are really in danger (due to location, etc.).

This personalised proactive alert system will allow to risk reduction (of insurance companies' customers) by assisting people to proactively take the necessary precautions and facilitating claims management, thus resulting in savings for both the insurance companies and the individuals.

Privacy considerations

In addition to the specific laws on protection and treatment of personal data, HDI's Ethic Code establishes that recipients who, in the exercise of their activities, acquire documents, studies, work plans (including business plans), technological processes, data and Information of any kind related, directly or indirectly, to the activities of HDI, have the obligation to guard and protect them in an appropriate and continuous manner in compliance with the security measures adopted by the Company pursuant to Italian D. Lgs. 196/2003 ("Codice Privacy"). In particular, personal information collected must be processed in accordance with the principles established in the "Codice Privacy" in a consistent and appropriate way to the purpose of their collection.

It is in any case compulsory to refrain from seeking confidential information, which is not functional in the exercise of their functions.

Data Protection considerations

Appropriate security measures are taken against unauthorised access to, or alteration, disclosure or destruction of the data and against their accidental loss or destruction:

- Particular focus is placed on the security of personal data held on portable devices, with appropriate security measures such as encryption applied.
- Robust procedures for limiting access to personal data are in place and that staff are aware of these limits.
- An appropriate external access policy is in place to ensure that only the data subject or their clearly chosen representative has access to their personal data during the course of a policy or claim.
- A confidentiality policy is in place pertaining to the collection, processing, keeping and use of medical and sensitive data.

Access to sensitive data is restricted to authorised staff. In particular, it is expected that access to sensitive medical information should be restricted to relevant underwriters, claims assessors and persons needing to access a particular file as part of their role.

Origin of personal data and its collection method

The Insurance demonstrator gathers data both from internal enterprise and both from external sources.

The list of dataset to be examined in the demonstrator are presented in the following table:

In-house dataset from Company CRM system
In-house dataset from Company Portfolio system
In-house dataset from Company Claims system
Institutional Italian databases
Private databases from third parties
National and local press releases
Reports and statistics on Insurance facts
GPS customer data
Weather statistics
Social media (e.g. Twitter, Facebook)

Natural Disaster datasets
Trending topics
Floods
Earthquakes

Concerning internal enterprise data sources, data will be adequately anonymised before being processed.

The Insurance demonstrator, may require the installation of a Mobile App (Scenario 2), in order to collect GPS data. To deal with this data HDI will involve a number of volunteers to generate data during the experimentation; Informed Consent Procedure for gathering the volunteers' consent to the transmission and processing of their data will be followed. Moreover, the simulation of this kind of data is an option considered in the project.

Concerning external data sources, they have their own licences with will be taken into consideration during the experimentation phase.

Purpose of the processing of personal data

The purpose of processing the personal data is defined through the specification of the high-level usage scenario reported in this deliverable. More specifically, the scope of Insurance demonstrator is **to provide customers with efficient added value services**.

- By correlating risk information with internal enterprise datasets, the Demonstrator is able to identify assets potentially involved in the risk. This information can be used in order to directly contact customers and to assist them in the event of damages to customer's assets insured with the Company.
- By correlating risk information with real time GPS data, the Demonstrator is able to identify if a customer may be involved in the risk. This information can be used to provide the customer with a fast and efficient assistance. For example, in the case the customer is found to be potentially involved in a hailstorm, our operators can contact him and suggest him the nearest garage in the area that can assist him.

An advanced customer segmentation enables the Company to provide its customers with personalised offerings that better fit his behaviour and needs.

6. CONCLUSION

The current deliverable constitutes a report on activities performed during the second iteration of all WP1 tasks (T1.1-T1.5).

Its first goal was to update the previously provided stakeholder analysis and needs identification. In order to achieve this, the consortium members prepared and conducted an updated, more targeted survey to elicit stakeholder requirements. The survey served a dual purpose: on one hand to understand the needs of the potential AEGIS stakeholders so as to develop a Big Data analytics platform of real added value, on the other hand to examine how the AEGIS platform impacts the market considering all the steps of the AEGIS Big Data Value Chain, spanning from data collection to data and service sharing in real scenarios. The survey contained alternative paths depending on the participant's role inside their organisation to ensure that each respondent was directed to questions best suited to their knowledge, making the questionnaire more engaging whilst obtaining more valuable answers. It should be noted that participation in the survey was significantly lower compared to the previous one conducted in the beginning of the project. This could be an outcome of the decision to design questions that go into greater detail regarding the way big data are leveraged inside PSPS organisations, which may have been difficult for some respondents to address. However, precisely due to the level of detail in the acquired responses, results were very insightful at this stage of the project.

The next objective of the current deliverable was to update the AEGIS Big Data Value Chain definition. As a first step, the data sources relevant to the 11 AEGIS stakeholder categories were grouped in order to discuss per stakeholder the expected challenges in handling them. These challenges, along with the challenges imposed by the 4 Vs of big data (volume, velocity, variety, veracity) were then discussed and, where possible, addressed in the updated definition of the data value chain. The big data value chain comprises five steps, namely data acquisition, data analysis, data curation, data storage and data usage, for each of which an overview of the way it is performed in AEGIS was provided.

The core contribution of the deliverable was the provision of the final integrated project methodology towards data driven innovation in the PSPS domains. As AEGIS aims to facilitate big data analysis through iterative exploration and experimentation of data and data-enabled services, the scope of its functionalities spans across an impossible to depict number of possible workflows. Therefore, the updated methodology definition focused on identifying and highlighting the common steps needed to accomplish possibly very diverse big data analysis tasks by very diverse users. As a means of validation of its correctness and completeness, two core workflows for big data analysis were selected to showcase how the methodology is instantiated and were described in detail.

Finally, this deliverable defined the final AEGIS Ethical, Privacy, Data Protection and IPR Strategy, where, besides an in-depth analysis of the provisions of the current European and national regulatory instruments relevant to AEGIS implementation and overall architecture, key aspects are described for both the project implementation phase and the AEGIS solutions, including ethics and data protection insights for each of the demonstrators, key principles of legal evaluation and assessment of technologies in AEGIS, methodology for the elicitation and

analysis of Ethical, Privacy, Data Protection and IPR Requirements and list of them, guiding principles and recommendations for AEGIS Data Policy Framework, both at project-level and at demonstrator-level.

APPENDIX A: AEGIS QUESTIONNAIRE

Data Analysis in your Organisation: do you really use it at its best?

This survey has been developed jointly by the partners of the AEGIS project, dedicated to "Advanced Big Data Value Chains for Public Safety and Personal Security" and is co-funded by the European Commission under the Horizon 2020 Programme (H2020-ICT-2016) under Grant Agreement No. 732189.

AEGIS aims to drive a data-driven innovation that expands over multiple business sectors and takes into consideration structured, unstructured and multilingual data sets, rejuvenate the existing models and facilitate all companies and organisations in the Public Safety and Personal Security(PSPS) linked sectors to provided better and personalised services to their users. Moreover, the project will introduce new business models through the breed of an open ecosystem of innovation and data sharing principles.

This is the second survey of the AEGIS project, the first one has been sent at the beginning of the project in order to define the preliminary user requirements and information sources of the stakeholders that are potentially interested in AEGIS data value chain. The analysis of the first survey is available at https://www.aegis-bigdata.eu/what-is-the-current-and-expected-use-of-big-data-technologies-a-glimpse-to-our-aegis-questionnaire-results/ (https://www.aegis-bigdata.eu/what-is-the-current-and-expected-use-of-big-data-technologies-a-glimpse-to-our-aegis-questionnaire-results/).

At the moment we are moving to the second part of the project, and we have developed a second version of the survey to collect further suggestions from the potential stakeholders, evaluating their concers and expectations related to Big Data. In particular, thanks to your replies to the questionnaire, we would like to understand how the AEGIS platform impacts to the market considering all the steps of the AEGIS Big Data value chain, from the collection to the sharing in real scenarios.

The questionnaire takes about ten minutes to complete.
The outcome will contribute to the AEGIS project towards the creation of a Big Data value chain for public safety and personal security.
Responses will be analysed so that no individual person or organisation can be identified.
Any information or answers to the questionnaire you provide will not be used for other purposes except the development of the AEGIS activities and will not be sold, rented, leased or forwarded to any third party.

You are more than welcome to submit additional input! Please send an email to: info@aegis-bigdata.eu (mailto:info@aegis-bigdata.eu).

Thank you for your time and input!

The AEGIS Team

General Information:

Your Name

Name of your organisation

Country

Email address

* Type of organisation:

- ○ Private, Company
- ○ Private, Research Centre
- ○ Public, Academia
- ○ Public, Institution
- ○ ONG

* Sector:

* Number of employees:

- ○ Micro organisation (<10)
- ○ Small organisation (10-49)
- ○ Medium-sized organisation (50-249)
- ○ Large organisation (>250)

* Which is your role in the organisation?

- ○ Manager
- ○ IT Technical Operator, Computer Systems Analyst
- ○ Data Handling Operator, Data Scientist

Manager

* To what extent does your organisation have experience in Big Data?

- ○ No experience
- ○ Planning to use Big Data
- ○ Beginning in the use of Big Data

* Who perform Data Analysis in your organisation?

- ☐ Internal team (main activity)
- ☐ Internal team (occasional activity)
- ☐ External consultant (inside the organisation)
- ☐ External providers

* Which are from your point of view the added values of Big Data Analysis?

- ☐ Cross-domain analysis
- ☐ Predictive analysis
- ☐ Real-time analysis
- ☐ Cost reduction
- ☐ Fast decision making
- ☐ Improvement of the offered services
- ☐ More effective marketing

- [ ] Better customer service
- [ ] Competitive advantages over rivals
- [ ] Other

* Which are the main issues related to Big Data handling in your organisation?

- [ ] Lack of confidence in the real benefit
- [ ] High management cost
- [ ] Difficulty of finding trained staff in Data Analysis
- [ ] Difficulty of handling Big Data
- [ ] Lack of performance of the available tools
- [ ] Legislation about privacy and security
- [ ] Other

* How many people work on Data Analysis in your organisation?

- ○ 0
- ○ 1
- ○ 2-7
- ○ >7

* Which of these steps have already been implemented in your organisation?

- [ ] Data Acquisition
- [ ] Data Analysis
- [ ] Data Curation
- [ ] Data Storage
- [ ] Data Usage
- [ ] Other

* Do you have any dedicated budget for Big Data and Analytics?

- ○ Yes, the investment is adequate
- ○ Yes, but the investment is not enough
- ○ No, but it is in our plan
- ○ No

* Do you have the proper hardware to manage Big Data?

- ○ Yes
- ○ No
- ○ I don't know

If no, why?

* Which are the data involved in the analysis of your organisation?

- [ ] Internal of the organisation
- [ ] Internal, related to customers
- [ ] External, customers (e.g. data from social media or sensors)
- [ ] External, open data
- [ ] Purchased data
- [ ] External, real-time data
- [ ] Other

* Does the agreement with your data providers include any reference to further processing of previously collected personal data?

- ( ) Yes
- ( ) No
- ( ) I don't know

* Do you use/would like to use interlinking datasets from different domains/ data sources for your analysis?

- ( ) Yes, we do
- ( ) No, we don't
- ( ) No, we don't, but it would be useful
- ( ) I don't know

* Are there in your organisation alert, warning or monitoring systems based on Big Data analytics?

- [ ] Yes, to prevent events
- [ ] Yes, to support after an event
- [ ] No
- [ ] I don't know

If yes, which kinds of events are you taking into account?

- [ ] Weather disaster
- [ ] Riots
- [ ] Geological disaster
- [ ] Terrorist attacks
- [ ] Other

* The analysis are shared:

- [ ] With external, entities
- [ ] With colleagues of the same office/team/department
- [ ] With colleagues of other offices of the same organisation
- [ ] As open data
- [ ] With customers
- [ ] I don't share analysis

* Would you been interested on a tool:

	Not at all	Slightly	Moderately	Very	N/A
Online and free	○	○	○	○	○
Where you can buy and sell assets	○	○	○	○	○
With a set of open assets	○	○	○	○	○
Where you can connect in-house streaming datasets	○	○	○	○	○
Where you can store your analysis and assets	○	○	○	○	○
Where you can manage the metadata related to your data	○	○	○	○	○
Where you can query your datasets and access a set of related visualisations	○	○	○	○	○
Where you can set and save the steps of your analysis	○	○	○	○	○
Where you can share the information with a selected group of users	○	○	○	○	○
Where you can set different restrictions about data visibility	○	○	○	○	○

IT Technical Operator

* To what extent does your organisation have experience in Big Data?

- ○ No experience
- ○ Planning to use Big Data
- ○ Beginning in the use of Big Data
- ○ Effectively using Big Data

* Who perform Data Analysis in your organisation?

- [ ] Internal team (main activity)
- [ ] Internal team (occasional activity)
- [ ] External consultant (inside the organisation)
- [ ] External providers

* Which are the main issues related to Big Data handling in your organisation?

- [ ] Lack of confidence in the real benefit
- [ ] High management cost
- [ ] Difficulty of finding trained staff in Data Analysi
- [ ] Difficulty of handling Big Data
- [ ] Lack of performance of the available tools
- [ ] Legislation about privacy and security
- [ ] Other

* How many people work on Data Analysis in your organisation?

- 0
- 1
- 2-7
- >7

* Which of these steps have already been implemented in your organisation?

- Data Acquisition
- Data Analysis
- Data Curation
- Data Storage
- Data Usage
- Other

* Do you have the proper hardware to manage Big Data?

- Yes
- No
- I don't know

If no, why?

* Which are the data source collected/ would you like to collect?

	Used	I would like to use
Log		
Transactions		
Events		
Emails		
Social media		
Sensors		
Open data/Public sector information		
Phone usage		
Reports to authorities		
External feeds		
RFID scans or POS data		
Earth observation and space		
Other geospatial		
Free-form text		
Audio		

Still images/video

* Which are the data involved in the analysis of your organisation?

- [ ] Internal of the organisation
- [ ] Internal, related to customers
- [ ] External, customers (e.g. data from social media or sensors)
- [ ] External, open data
- [ ] Purchased data
- [ ] External, real-time data
- [ ] Other

* How often do you acquire data?

- [ ] Continuous data streaming from specific data sources (e.g. sensor data, social media)
- [ ] Scheduled data streaming - Daily
- [ ] Scheduled data streaming - Weekly
- [ ] Scheduled data streaming - Monthly
- [ ] Data acquisition when needed

* Do the employees of your organisation have different restrictions about data visibility?

- ( ) Yes
- ( ) No
- ( ) I don't know

* Does your organisation have scheduled automated analysis?

- [ ] Yes
- [ ] No
- [ ] I don't know

* Are there in your organisation alert, warning or monitoring systems based on Big Data analytics?

- [ ] Yes, to prevent events
- [ ] Yes, to support after an event
- [ ] No
- [ ] I don't know

If yes, which kinds of events are you taking into account?

- [ ] Weather disaster
- [ ] Riots
- [ ] Geological disaster
- [ ] Terrorist attacks

- [ ] Other

* Does your organisation provide data to third parties?

- [ ] Yes, we publish data
- [ ] Yes, we sell data
- [ ] Yes, we rent data
- [ ] No
- [ ] I don't know

* The analysis are shared:

- [ ] With external, entities
- [ ] With colleagues of the same office/team/department
- [ ] With colleagues of other offices of the same organisation
- [ ] As open data
- [ ] With customers
- [ ] I don't share analysis

* Would you been interested on a tool:

	Not at all	Slightly	Moderately	Very
Online and free	○	○	○	○
Where you can buy and sell assets	○	○	○	○
With a set of open assets	○	○	○	○
Where you can upload your in-house datasets (eventually anonymizing them)	○	○	○	○
Where you can store your analysis and assets	○	○	○	○
Where you can manage the metadata related to your data	○	○	○	○
Where you can query your datasets and access a set of related visualisations	○	○	○	○
Where you can set and save the steps of your analysis	○	○	○	○
Where you can share the information with a selected group of users	○	○	○	○
Where you can set different restrictions about data visibility	○	○	○	○

Data Scientist

* Which are from your point of view the added values of Big Data Analysis?

- [ ] Cross-domain analysis
- [ ] Predictive analysis
- [ ] Real-time analysis
- [ ] Cost reduction
- [ ] Fast decision making
- [ ] Improvement of the offered services
- [ ] More effective marketing
- [ ] Better customer service
- [ ] Competitive advantages over rivals
- [ ] Other

* How many people work on Data Analysis in your organisation?

- ( ) 0
- ( ) 1
- ( ) 2-7
- ( ) >7

* Which of these steps have already been implemented in your organisation?

- [ ] Data Acquisition
- [ ] Data Analysis
- [ ] Data Curation
- [ ] Data Storage
- [ ] Data Usage
- [ ] Other

* Which are the data involved in the analysis of your organisation?

- [ ] Internal of the organisation
- [ ] Internal, related to customers
- [ ] External, customers (e.g. data from social media or sensors)
- [ ] External, open data
- [ ] Purchased data
- [ ] External, real-time data
- [ ] Other

* How often do you acquire data?

- [ ] Continuous data streaming from specific data sources (e.g. sensor data, social media)
- [ ] Scheduled data streaming - Daily
- [ ] Scheduled data streaming - Weekly
- [ ] Scheduled data streaming - Monthly
- [ ] Data acquisition when needed

* Which are the data sources you use/ would you like to use to enhance the quality of your analysis?

	I use	I would like to use
Log	○	○
Transactions	○	○
Events	○	○
Emails	○	○
Social media	○	○
Sensors	○	○
Open data/Public sector information	○	○
Phone usage	○	○
Reports to authorities	○	○
External feeds	○	○
RFID scans or POS data	○	○
Earth observation and space	○	○
Other geospatial	○	○
Free-form text	○	○
Audio	○	○
Still images/video	○	○

* Do you have the proper analytic tools related to your needs?

○ Yes

○ No

If yes, which tools do you use?

If no, which tools would you like to use and why?

* Which are the algorithms involved/would you like to involve in your analysis?

	Used	I would like to use
Estimation of correlations between variables	☐	☐
Linear regression	☐	☐
Predictive analysis	☐	☐
Clustering algorithms	☐	☐
Simulations	☐	☐

Other

* Do you use/would like to use interlinking datasets from different domains/ data sources for your analysis?

- Yes, we do
- No, we don't
- No, we don't, but it would be useful
- I don't know

* Which are the formats of the output of your analysis?

- Tabular
- Textual
- Graph
- Other

* Do the employees of your organisation have different restrictions about data visibility?

- Yes
- No
- I don't know

* Does your organisation have scheduled automated analysis?

- Yes
- No
- I don't know

* Are there in your organisation alert, warning or monitoring systems based on Big Data analytics?

- Yes, to prevent events
- Yes, to support after an event
- No
- I don't know

If yes, which kinds of events are you taking into account?

- Weather disaster
- Riots
- Geological disaster
- Terrorist attacks
- Other

* The analysis are shared:

- [ ] With external, entities
- [ ] With colleagues of the same office/team/department
- [ ] With colleagues of other offices of the same organisation
- [ ] As open data
- [ ] With customers
- [ ] I don't share analysis

* Would you been interested on a tool:

	Not at all	Slightly	Moderately	Very
Online and free	○	○	○	○
Where you can buy and sell assets	○	○	○	○
With a set of open assets	○	○	○	○
Where you can upload your in-house datasets (eventually anonymizing them)	○	○	○	○
Where you can store your analysis and assets	○	○	○	○
Where you can manage the metadata related to your data	○	○	○	○
Where you can query your datasets and access a set of related visualisations	○	○	○	○
Where you can set and save the steps of your analysis	○	○	○	○

APPENDIX B: NON-DISCLOSURE AGREEMENT TEMPLATE

Non-Disclosure Agreement

between

Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e. V., Hansastrasse 27c, 80686 Munich, Federal Republic of Germany

- hereinafter referred to as »Fraunhofer«-

as legal entity for its

Fraunhofer Institute for Open Communication Systems FOKUS,

Kaiserin-Augusta-Allee 31, 10589 Berlin, Federal Republic of Germany

- hereinafter referred to as »Fraunhofer FOKUS« -

and

GFT Italia SRL, Via Campanini Alfredo, 20124 Milano, Italy

Kungliga Tekniska Hoegskolan, Brinellvagen 8, 100 44 Stockholm, Sweden

UBITECH Ltd., 26 Nikou & Despinas Pattchi, 3071 Limassol, Cyprus

Kompetenzzentrum – Das virtuelle Fahrzeug, Forschungsgesellschaft mbH, Inffeldgasse 21 A, 8010 Graz, Austria

National Technical University of Athens – NTUA, Heroon Polytechniou 9, Zographou Campus, 15780 Athina, Greece

Ecole Polytechnique Federale de Lausanne, Batiment CE 3316 Station 1, 1015 Lausanne, Switzerland, c/o College of Management of Technology (hereinafter referred as to "CDM"), represented by Prof. Andreas Mortensen, Vice-Provost for Research, and Prof. Christopher Tucci, Head of CDM

Suite5 Ltd., Wenlock Road 20-22, N1 7GU London, United Kingdom

Hypertech (Chaipertek) Anonymos Viomichaniki Emporiki Etaireia Pliroforkis Kai Neon Technologion, 32 Perikleous Street, 15232, Chalandri Athina, Greece

HDI Assicurazioni S.p.A., Via Abruzzi 10, 00187 Rome, Italy

- hereinafter together referred to as »Project Partners« -

and

…….

……..
[name, address]

- hereinafter referred to as »Board Member « -

Preamble

The Project Partners co-operate in the project "Advanced Big Data Value Chain for Public Safety and Personal Security" (hereinafter: Project) which is partly funded by the European Commission in the H2020 Research Framework Programme under Grant No. 1290/2013, and which is coordinated by Fraunhofer. The Project Partners have agreed that certain aspects of their work under the Project shall be assessed by experts of an Ethics Advisory Board so to benefit from the Board Members' expertise in order to optimize the Project results. To protect the results and information exchanged between the Project Partners and the Board Member, the following agreement is concluded:

§ 1 Purpose of the Board

The Project Partners will establish the Ethics Advisory Board that will include relevant external, independent expertise to evaluate the Project's progress and the results generated thereunder and to advise the Project Partners how to proceed with the Project ethically correct. Additionally, the Ethics Advisory Board shall supervise the conduct of the Project to ensure that European regulations regarding data protection are fully observed.

The Board Member will be invited to the Project meetings in order to learn about the Project's objectives and approach but shall not have any voting rights.

The Board Member shall contribute to the Ethics Advisory Board's Report that summarizes the evaluation activities of the Ethics Advisory Board and contains the Ethics Advisory Board's recommendations. The report shall be submitted as AEGIS Deliverable 9.3 as attachment to the AEGIS Periodical Reporting in Project Month 18 At the end of the Project, the Ethics Advisory Board shall update its report. The updated report as attachment to the AEGIS Periodical Reporting shall be submitted in Project Month 30.

The Board Member commits itself to actively contribute to the Ethics Advisory Board activities.

The Board Member will get access to deliverables and Results generated by the Project Partners as well as to intended publications from the Project whose drafts need to be treated in confidence.

§ 2 Confidentiality

For the purposes of this Agreement »Confidential lnformation« shall mean

- any technical and/or commercial Information, including – but not limited to – any documents. drawings, sketches or designs, materials or samples disclosed by any of the Project Partners or their subcontractors to the Board Member;
- information obtained from another member of the Ethics Advisory Board;
- deliverables/results generated by the Project Partners or their subcontractors.

The Board Member agrees to treat as confidential all and any Confidential Information – whether obtained directly or indirectly – and to use the same only for the purpose of the execution of its duties as a Board member and not to use or exploit such Confidential Information for its own or any third party purpose, disclose it to any third party or allow any third party access to such Confidential Information, except with the prior written consent of the disclosing Project Partner.

The Board Member will take all necessary precautions to ensure the confidentiality of the Confidential Information.

The restrictions on the use and disclosure of Confidential Information shall not apply to information which is:

(a) proven to have been known to the Board Member prior to the time of its disclosure pursuant to this Agreement; or

(b) in the public domain at the time of disclosure to the of Board Member or thereafter enters the public domain without breach of the terms of this Agreement; or

(c) lawfully acquired by the Board Member from an independent source having a bona fide right to disclose the same; or

(d) independently developed by the Board Member provided that it has not had access to any of the Confidential Information of the disclosing Project Partner.

§ 3 Liability

The Board Member shall be held liable for any damage caused to the Project Partners by breach of its duties under this Agreement.

The Parties agree that any Confidential Information is made available »as is« and that no warranties

are given or liabilities of any kind are assumed with respect to the quality of such Confidential Information, including, but not limited, to its fitness for the purpose, non-infringement of third party rights, accuracy, completeness or its correctness.

§ 4 Non-assignment

This Agreement is personally binding the Board Member and shall not be assigned by the latter without the Project Partners' prior written consent.

§ 5 Intellectual Property

The Board Member agrees not to exploit Confidential Information, in particular not to apply for the registration of intellectual property rights.

All Confidential Information supplied pursuant to this Agreement shall remain the property of the Project Partner disclosing or supplying the same and nothing contained in this Agreement shall be construed as granting to or conferring upon the Board Member any rights by license or otherwise, express or implied, to the Confidential Information, accompanying know how or any underlying intellectual property rights of the Project Partners.

Should any results generated by the Board Member – in particular from evaluating the Project results or deliverables – be eligible for protection under intellectual or industrial property laws or be protected under copyright law, the rights to use such results shall be assigned by the Board Member to the Project Partners. This shall especially apply to the Board Member's contributions to the documents and reports mentioned under § 2. Additionally, the Project Partners shall be entitled to use the contributions and recommendations generated by the Board Member unrestrictedly in time, place and content. For the avoidance of doubt, this right of use contains also the right to implement and develop such contributions and recommendations.

§ 6 Entry into force and Term

This Agreement shall come into force on the date of the last signature and shall thereafter be valid until September 2020, the current planned end date of the AEGIS reporting.The obligation of confidentiality hereunder shall continue to be valid for a period of 10 years after the end of the term of this Agreement.

Upon request of the disclosing Project Partner, any document, sample or material shall be returned by the Board Member to the disclosing Project Partner without delay and at the end of this Agreement at the latest.

§ 7 Miscellaneous

Amendments and additions to this Agreement must be made in writing to have legal effect.

This Agreement is subject to and governed by the laws of the Federal Republic of Germany.

If any provision of this Agreement is determined to be illegal or in conflict with the applicable law, the validity of the remaining provisions shall not be affected. The ineffective provision shall be replaced by an effective provision which is economically equivalent. The same shall apply in case of a gap.

Signed on behalf of the Project Partners, acting through the Coordinator

Place, date

Place, date

_____

Signature of the Board Member

APPENDIX C: EXPERT AGREEMENT – TEMPLATE

Agreement ("Ethics Advisory Board" Expert)

between

NAME of the expert,
Address

hereinafter referred to as "Party or Expert or Board Member"

and

GFT ITALIA SRL (GFT) SRL, 1549351, established in VIA SILE 18, MILANO 20139, Italy, VAT number IT00819200478

hereinafter referred to as "GFT"

WHEREAS,

I. In cooperation with other Beneficiaries, the Coordinator (Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e. V.) has been awarded a Grant Agreement by the European Commission (EC) number 732189 entitled "Advanced Big Data Value Chain for Public Safety and Personal Security, - AEGIS", hereinafter referred to as the Grant Agreement or GA. From this Grant Agreement including its Annexes certain rights and obligations result between the EC and the Coordinator. The Grant Agreement provides the participation of the Expert for certain part of the work;
II. As AEGIS may rise some concerns regarding ethical and privacy issues due to the use of users' personal data after their written consent, the AEGIS Consortium (Annex 1) has decided to put together an advisory board named Ethics Advisory Board (EAB), comprising of known domain experts and practitioners who will work closely with the overall Consortium during the course of the project on tackling ethical and data privacy issues that will have to do with the retrieval, the processing, and the retaining of these data. The EAB will provide independent opinions and thoughts and will advise both the technical and the research partners on issues regarding the AEGIS methodology, the development of the platform and its components and the piloting

operation. The Ethics Advisory Board will be coordinated by the EAB Coordinator, who will be responsible for interfacing with it;

III. GFT, as EAB Coordinator, is in charge for setting up and coordinating the EAB and of subcontracting for engaging the Experts;

IV. In performing the work as member of the Ethics Advisory Board it is anticipated that GFT, the Coordinator and the other partners of the "AEGIS" Consortium disclose to the Expert technical and/or commercial information of a confidential nature presently in their possession and wish to ensure that the same remain confidential. The Expert, the Coordinator and the other AEGIS Project Partners have previously signed a Non-Disclosure Agreement on 00.00.2017, which is still valid and binding.

Now, therefore, it is hereby agreed as follows:

2. The Expert will collaborate with the Coordinator, GFT and the other partners of the "AEGIS" Consortium in order to tackle any ethical issue raised by the project and monitor the legal issues to continuously assess and ensure that the framework being proposed adheres to a minimum set of ethical and legal requirements. The Board Member commits itself to actively contribute to the Ethics Advisory Board activities. In particular, the Expert will perform the work as follows upon demand of GFT:

    a. Provide his/her expertise in specific ethics and privacy areas (as instructed by the Consortium and the EC) during the whole duration of the project. The Expert will contribute to provide independent opinions and thoughts and to advise both the technical and the research partners on issues regarding the AEGIS methodology, the development of the platform and its components and the piloting operation, by providing expertise in specific ethics and privacy areas during the whole duration of the project. The Expert will contribute to propose the Assessment Methodology to be described in D9.1 and followed in WP1 and WP5, including, if opportune, the provision of Templates at an early stage and the coherence with the Ethical Risk Table already named in the AEGIS Annex I;
    b. Participate and/or contribute to AEGIS workshops or meetings, which will be conducted during the project;
    c. Co-create and/or review selected parts of the ethics and privacy related deliverables (e.g. Deliverable D1.2 - Aegis Methodology and High Level Usage Scenarios Aegis Methodology and High Level Usage Scenarios, Deliverable D6.3 - Data Management Handling Plan);
    d. Periodically report to the commission on the implementation of the ethical issues in project and compliance with applicable national and EU regulations. The Board Member shall contribute (in writing) to the Ethics Advisory Board's Report that summarizes the evaluation activities of the Ethics Advisory Board and contains the Ethics Advisory Board's recommendations. The reports will be

based on a common assessment methodology as introduced in D9.1of the AEGIS GA. The report shall be submitted as AEGIS Deliverable 9.3, as attachment to the AEGIS Periodical Reporting in Project Month 18. At the end of the Project, the Ethics Advisory Board shall update its report. The updated report as attachment to the AEGIS Periodical Reporting shall be submitted in Project Month 30.
The Experts are expected to collaborate collectively in the generation of these two reports. The length of each of thesse reports should be adequate and not shorter than 4 pages.

GFT will instruct the Expert in due time as to the dates of operation.

The Expert is responsible for ensuring that the research work meets scientific care, complies with accepted technical, scientific and professional standards, is undertaken by appropriate personnel and carried out in accordance with the the financial provisions laid down in Article 3.

3. The duration of the Grant Agreement is 30 months commencing on 01-01-2017 and terminating on 30-06-2019.The Expert shall commence to perform its part of the work on **00-00-0000** and shall have completed it on 31.05.2019 at the latest to enable GFT to consider the Expert's contributions in the final project report. At the latest by that date all results and reports shall have been delivered to GFT.

   The Expert shall notify GFT in writing without undue delay if it becomes apparent that it might be unable to keep the schedule.

3. The remuneration to be paid to the Expert under this agreement amounts to a lump sum of **5.000,00 Euro,** including VAT, if applicable, and costs for office supplies, communication, insurance, visa, travels, taxes, accommodation, subsistence, telecommunications and any other project related costs if not agreed otherwise herein, and is payable as follows:

   50% (contribution to WP1, WP5 and WP9) on 01.03.2018;
   50% upon completion of this Agreement and acceptance of all deliverables, reports and results.

   The Expert will be liable himself for paying taxes and making his own contributions to social security out of this sum.

Unless otherwise agreed in written form, the Expert shall personally bear travel and subsistence costs incurred by the Expert in connection with providing the services under this Agreement.

GFT shall make the payments to the bank account stated in Appendix 2 upon delivery and acceptance of the performed work and receipt of invoices from the Expert. Such invoices shall quote a reference to the Grant Agreement Number of the European Commission and shall provide a detailed description of the work/deliverables concerned.

4. The Board Member will get access to deliverables and Results generated by the Project Partners as well as to intended publications from the Project whose drafts need to be treated in confidence. The Expert is bound by the Confidentiality Obligation as well as the other obligations set forth in §§ 2, 3, 4 and 5 of the NDA.

5. Should any results generated by the Expert – in particular from evaluating the Project results or deliverables – be eligible for protection under intellectual or industrial property laws or be protected under copyright law, the rights to use such results shall be assigned by the Expert to GFT and the other partners in the AEGIS project. This shall especially apply to the Expert's contributions to the documents and reports mentioned under Section 1. Additionally, GFT and the other AEGIS partners shall be entitled to use the contributions and recommendations generated by the Expert unrestrictedly in time, place and content. For the avoidance of doubt, this right of use contains also the right to implement and develop such contributions and recommendations.

5. This Agreement shall come into force upon the date of its signature and shall thereafter be valid until complete fulfilment of all obligations undertaken by the Expert under this Agreement.

6. Any and all disputes that will arise in connection to this Agreement will be governed by the laws of Italy. Any disputes arising out of the present Agreement which cannot be solved amicably, shall be finally settled under the Rules of Arbitration of the International Chamber of Commerce by one or more arbitrators appointed in accordance with the said Rules. The place of arbitration shall be Milan, Italy if not otherwise agreed by the conflicting Parties. The award of the arbitration will be final and binding upon the Parties. Nothing in this Agreement shall limit the Parties' right to seek injunctive relief or to enforce an arbitration award in any applicable competent court of law.

7. If any provision of this agreement is determined to be illegal or in conflict with the applicable law, the validity of the remaining provisions shall not be affected. The

ineffective provision shall be replaced by an effective provision, which is economically equivalent. The same shall apply in case of a gap.

Signed for the Expert

Place, Date

Signature

Signed for and on behalf of GFT

Place, Date

Signature

APPENDIX D: LITERATURE

Acquisto, G. and Domingo-Ferrer, J. and Kikiras, P.and Torra, V. and de Montjoye, Y. and Bourka, A. (2015) Privacy by design in big data, European Union Agency for Network and Information Security: Enisa – Privacy in Big Data, p. XX

Albrecht, J.-P. (2016) *Das neue EU-Datenschutzrecht - von der Richtlinie zur Verordnung*: CR 2016, 88ff

ARTICLE 29 DATA PROTECTION WORKING PARTY (2013) *Opinion 03/2013 on purpose limitation:* WP-29 purpose limitation p. xx

ARTICLE 29 DATA PROTECTION WORKING PARTY (2017) *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679:* WP-29 decision making p. xx

ARTICLE 29 DATA PROTECTION WORKING PARTY (2017) *Guidelines on Data Protection Impact Assessment (DPIA):* WP-29 DPIA p. xx

Auer-Reinsdorff, A. and Conrad, I. (2016) *Handbuch IT- und Datenschutztrecht,* 2nd Edition, München*:* Auer-Reinsdorff/Conrad, IT- und Datenschutzrecht, § xx *Chapter* Rn. x

Bharosa, N., Janssen, M., Klievink, B., & Tan, Y. (2013). Developing Multi-sided Platforms for Public-Private Information Sharing: Design Observations from Two Case Studies. *Proceedings of the 14th Annual International Conference on Digital Government Research*. New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2479724.2479747

Boehme-Neßler, V. (2016) *Das Ender der Anonymität – Wie Big Data das Datenschutzrecht verändert, Oldenburg*: DuD, 2016, 419ff

Callies, C. and Ruffert, M. (2016) *EUV/AEUV – Das Verfassungsrecht der Europäischen Union mit Europäischer Grundrechtcharte Kommentar*, 5. Edition, München: Calliess/Ruffert/Kingreen EU-GRCharta Article xx Rn. x

Culik, N. and Döpke, C. (2017) *Zweckbindungsgrundsatz gegen unkontrollierten Einsatz von Big Data-Anwendungen,* München: ZD 2017, 226ff

Curry, E. (2016). The big data value chain: definitions, concepts, and theoretical approaches. In New Horizons for a Data-Driven Economy (pp. 29-37). Springer, Cham.

Danezis, G. and Domingo-Ferrer, J. and Hansen, M. and Hoepman, J. and Métayer, D. and Tirtea, R. and Schiffner S. (2014) *Privacy and Data Protection by Design – from policy to engineering,* European Union Agency for Network and Information Security: Enisa – Privacy by design, p. XX

Ehmann, E. and Selmayr, M. (2017) *Datenschutz-Grundverordnung Kommentar,* 1st Edition, München: Ehmann/Selmayr, DS-GVO Art. xx Rn. x

Eisenmann, T. R., Parker, G., & Van Alstyne, M. W. (2006). Strategies for Two-Sided Markets. *Harvard Business Review, October*, 1–12. https://doi.org/10.1007/s00199-006-0114-6

Fung, B. and Wang, K. and Chen, R. and S. Yu, P. (2010) *Privacy-Preserving Data Publishing: A Survey of Recent Developments*, ACM Computing Surveys, Vol. 42, No. 4, Article 14: Privacy Preserving, p. XX

Gründwald, A. and Hack, J. (2017) *Das neue umsatzbezogene Sanktionsregime der DS-GVO - Bußgeldbemessung nach kartellrechtlichen Maßstäben?*, München: ZD 2017, 556ff

Hagiu, A., & Wright, J. (2015). Multi-sided platforms. *International Journal of Industrial Organization*, *43*, 162–174. http://dx.doi.org/10.1016/j.ijindorg.2015.03.003

Hoeren, T. (2016) *Big Data und Datenqualität –ein Blick auf die DS-GVO*, München: ZD 2016, 459ff

Hoffmann, M. and Johannes, P. (2017) *DS-GVO: Anleitung zur autonomen Auslegung des Personenbezugs*, München: ZD 2017, 221ff
Kartheuser, I. and Gilsdorf, F. (2017) *Dynamische IP-Adressen können personenbezogene Daten sein:* MMR-Aktuell 2016, 382533

Katko, P. and Babaei-Beigi, A. (2014) *Accountability statt Einwilligung? - Führt Big Data zum Paradigmenwechsel im Datenschutz?*, München: MMR 2014, 360ff

Kindhäuser, U. and Neumann, U. and Paeffgen, H.-U. (2017) *Strafgesetzbuch Kommentar*, 5. Edition, München: Kindhäuser/Neumann/Paeffgen, Strafgesetzbuch, StGB § xx Rn. X

Kühling, J. and Buchner, B. (2017) *Datenschutz-Grundverordnung Kommentar*, 1st Edition, München: Kühling/Buchner Art. xx Rn. x

Maunz, T. and Dürig, G. (2016) *Grundgesetz Kommentar*, 79. Edition, München: Maunz/Dürig/Grzeszick, 79. EL Dezember 2016, II. Rn. 61

Monreal, M. (2016) *Weiterverarbeitung nach einer Zweckänderung in der DS-GVO*, München: ZD 2016, 507ff

Nebel, M. (2015) *Schutz der Persönlichkeit – Privatheit oder Selbstbestimmung? - Verfassungsrechtliche Zielsetzungen im deutschen und europäischen Recht*, München: ZD 2015, 517ff

Ohrtmann, J. (2014) *Big Data und Datenschutz – Rechtliche Herausforderungen und Lösungsansätze:* NJW 2014, 2984ff

Sachs, K and Sachs, M. (2016) *Europäische Grundrechtcharta: GRCh Kommentar*, 1. Edition, München: Stern/Sachs GRCh p. 213

Schaar, K. (2017) *Anpassung von Einwilligungserklärungen für wissenschaftliche Forschungsprojekte*, München: ZD 2017, 213ff

Schmitz, B. and von Dall'Armi, J. (2017) *Datenschutz-Folgenabschätzung – verstehen und anwenden*, München: ZD 2017, 57ff

Sydow, G. (2017) Europäische Datenschutz-*Grundverordnung Handkomment*, 1st Edition, Baden-Baden: Sydow, DS-GVO Art. xx Rn. x

Veil, W. (2015) *DS-GVO: Risikobasierter Ansatz statt rigides Verbotsprinzip*, München: ZD 2015, 347ff

Werkmeister, C. and Brandt, E. (2016) *Datenschutzrechtliche Herausforderungen für Big Data*, München: CR 2016, 233ff

Wybitul, T. (2017) *EU-Datenschutz-Grundverordnung Handbuch*, 1st Edition, Frankfurt a.M.: Wybitul, Handbuch DS-GVO Art. xx Rn. x

Zhao, F. and Grumbling E. (2014) *BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE*, PCAST: Big Data and Privacy p. xx