HORIZON 2020 - ICT-14-2016-1

AEGIS

Advanced Big Data Value Chains for Public Safety and Personal Security

WP9 – Ethics Requirements

D9.1 OEI - Other Ethics Issues. Requirement No. 1

Due date: 31.03.2018	**Delivery Date**: 07.06.2018
Author(s): Dustin Stadtkewitz, Yury Glikman (Fraunhofer), Marina Da Bormida (AEGIS Ethics Board)	
Editor: Yury Glikman (Fraunhofer)	
Lead Beneficiary of Deliverable: Fraunhofer	
Dissemination level: Public	**Nature of the Deliverable:** Report
Internal Reviewers: Gert G. Wagner, George Karagiannopoulos (AEGIS Ethics Board)	

Explanations for Front page

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 defines a strategy for assessment of data protection and ethical issues in AEGIS applies it to the three demonstrators developed in the project.

The first step of the methodology definition is to outline relevant ethical questions arising from the consequences of Big Data analysis for the society as whole and specific societal groups as well as the individual alone. Examining ethical issues includes the definition of the respective protected interest in the first step, followed by representing the inferences of the outcome of the analysis from massive data in Big Data applications respectively in AEGIS.

The second important aspect addressed by the proposed methodology is the assessment of risks related to compliance of AEGIS to the legal regulations and especially to data protection. The article 35 sec. 1 GDPR, the central regulation for providing an data protection impact assessment, explicitly obligates the data controller to "take *into account the nature, scope, context and purposes of the processing*" whenever it "*is likely to result in a high risk to the rights and freedoms of natural persons*" which means for the controller to "*carry out an assessment of the impact of the envisaged processing operations on the protection of personal data.*" In order to enhance the purpose of a data protection impact assessment for each data processor using AEGIS in his/her data workflow, this document provides a short but comprehensive overview of AEGIS workflow, specifically its application in the chosen demonstrator cases.

The third section of the document presents the applying the assessment methodology to the three project demonstrators. The outcomes of the assessment are summarised in the conclusion section of the document.

Table of Contents

1. METHODOLOGY AND ASSESSMENT STRATEGY

The enormous amount and the continuous production of data represents an ideal source for complex data analysing technologies. What is seen as a new method for an improved collecting and processing of personal and non-personal data to reveal useful patterns for mainly industry and large market-leading companies is also regarded as extensive interference in serious privacy concerns.

The digital transformation of the business environment and the improvement of storage capacities based on distributed file-systems, enhanced analysis methods and better network capability result in the creation of new fields of application for industry and society. With the optimization and automation in the industry[1], individualized online –services[2] or all kinds of the smart technologies, privacy concerns have to be adequately considered. The interest in privacy issues and the static growth of technologically and economy should not compete with each other, but ensure complementary, in a way that fundamental data privacy principles are connected with and implemented in Big Data applications.

Identifying potential risks within Big Data technologies can avoid a number of unintended and highly impactful consequences. The research and advisory firm Gartner predicted that, "b*y 2018, 50 percent of business ethics violations will occur through improper use of big data analytics.*" In order to prove this prediction as incorrect the first step towards an effective implementation of data protection issues is to make aware of the responsibility deriving from the use of Big Data. The own advantage and purpose should not dominate the data processing in Big Data, thereby disregarding the users or costumers interest. Much more a fair Big Data application should handle privacy issues together by taking preventative steps regarding privacy design strategies in Big Data in order to avoid miserable effects. A special focus for preventive steps should be taken into account in the design phase with methods like anonymisation, encrypted search, privacy preserving computations, access control mechanisms and intern policy enforcements.

1.1. Assessment strategy of data protection and ethical issues in AEGIS

1.1.1. Considerations of Big Data

With the realization of data storages capable to store an excessive amount of data for analysis purposes, it was made possible to utilize and exploit much more information than in a common way. The increasing extraction of knowledge from information enables new possibilities and application opportunities. One of the main abilities of Big Data is the capability of mining huge datasets, which lead to reveal new and surprising insights.[3]

The wide application of Big Data to examine and predict all kind of human activities influences all fields of society. The main field of application are: e.g. online-shopping, which includes collecting and analysing consumer behaviour[4], customer behaviour of streaming platforms[5] to recommend content on demand, patient data for individual treatment methods, all kinds of

[1] e.g. predictive maintenance, automated procedures.
[2] e.g. advertisements for food deliverable services or customer preferences in amazon.
[3] Ethics and Big Data
[4] E.g. Amazon
[5] E.g. Spotify, Netflix

industry data for education purposes, or industry data from huge machines for better performance and efficiency or predictive maintenance, government issues and so on.[6]

The source of Big Data applications come from all kinds of internet accessible devices and platforms that the majority of people use like social networks[7], Blogs and comments on digital newspapers, internet search engines, the content of social media communications[8] and so on.[9] The success of most online services implies the acceptance of common business practices concerning the collection and process of all kind of data. Whereas concerns about Big Data application should not be underestimated, as the abuse can result in heavy privacy problems with impact on the individual´s life. Big Data analytics and predictions represent a great source of power to realize one-sided interest of businesses and companies at the expense of those of the consumer.

The great imbalance caused by the use of the power in knowledge requires the institutional compensation to protect the individual´s rights. It was mentioned that the "Power of Knowing" leads to the obligation to use this knowledge responsibly. In this way, developers and scientist should be aware of their responsibility as the creator of Big Data applications, to implement appropriate privacy enhancing and security methods in order to benefit of analytic procedures without weighing privacy protection issues. Beyond, the legal framework should proportionally develop in a way that govern data use and realize ethical values and fundamental principles of privacy and security in the technology age.

1.1.2. Necessity and criteria of a Data Privacy Impact Assessment

The necessity of performing a DPIA, with reference of Article 35 sec. 3, is obligated in cases, where "large-scale processing operations which aim to process a considerable amount of personal data at regional, national or supranational level and which could affect a large number of data subjects and which are likely to result in a high risk […] to the rights and freedoms of data subjects".[10] Target is the identification of possible risks during the data processing as well as the implementation of appropriate measures.

A data protection impact assessment shall ensure compliance of current data protection regulations, legal certainty for the public as well as the data subject itself. Due to the risk-based approach of the GDPR and the complexity as well as fuzziness of article 35 GDPR, required is a relation between the intensity of the data processing and the interference in the data subject´s privacy. For this, the data processing operations as well as the purpose of the data process have to be described in a way that interferences in the data subject´s interest and rights are visible. Furthermore, the data controller has to prove the implementation of proportional safeguards capable for security and the safety of personal data. In this way, the data controller is obligated to prove compliance before data protection infringements can occur.

As result, an assessment generally includes the following steps:

[6] https://www.simplilearn.com/big-data-applications-in-industries-article
[7] E.g. Facebook, Twitter, Tumblr
[8] E.g. motion profiles from messages, log-files etc.
[9] https://www.omicsonline.org/open-access/impact-of-big-data-analytics-on-healthcare-and-society-2155-6180-1000300.php?aid=75499
[10] Recitals 91 GDPR

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

1.1.3. Summary of ethical principles and risks

Ethical principle	Description
Privacy	Big Data brings new challenges in terms of the privacy. Privacy issues may be potentially boosted by the variety, volume and wide area deployment of the system infrastructure to support AEGIS services, which are based on Big Data, coming from various sources. These sources produce large amounts of data, which are transformed into valuable information, and may include, for instance, social media, sensors and devices connected to the "Internet of Things" (IoT), scientific applications, surveillance, video and image archives, Internet search indexing, medical records, business transactions and system logs. Privacy is a pertinent aspect to be addressed because more and more personal data and content are captured, shared and processed. So, a secure framework for privacy management is a very hot topic research in AEGIS. The privacy challenges cover the entire spectrum of the Big Data lifecycle, ranging from the sources of data production (e.g. devices), to the data itself, data processing, data storage, data transport and data usage. As acknowledged by the Cloud Security Alliance, despite Big Data present interesting opportunities for users and businesses, such opportunities are countered by enormous challenges in terms of privacy and traditional privacy mechanisms need to be rethought to be able to provide a capable answer to those challenges, moving forward the basic and more common solutions. In case of Big Data solutions like AEGIS, where data processing get faster, then encryption, masking and tokenization are critical elements for protecting sensitive data and an holistic vision at privacy need to be taken, lingering over the identification of the different data sources, the origin and creators of data, and who is allowed to access the data.
Control	The massive amount of data generated by everyone represents an ideal source for Big Data analysis and prediction. An interactive live-stat application showed that after one minute, about half a million tweets have been sent, 80.000 new photos published on Instagram and about 175.000.000 E-Mails have been sent.[11] Besides, there are additional data

[11] Interactive statistic for social media: http://www.internetlivestats.com/one-second/

	sources like wearables or sensor data, GPS data from messenger services and other kinds of data feeding the data storage in Big Data for various data analytic methods. One important aspect for the individual is to have an effective influence of these data. This, first of all, requires the individual to be aware of what data is collected. Provided that the individuals are perfectly aware of these procedures, with regard to the amount of data alone the fact, that the individuals stand against this massive amount of data can trigger the feeling of powerlessness. Even with the request from the user to erase all data on a business companies storage, a feeling of uncertainty could be left with regard to the right to be forgotten and considerations like for what purpose data has been used, the users cannot be absolutely sure, whether their data was not sold and transmitted to third parties in advance leading to a loss of control. Furthermore, due to the interlinking of data across all kinds of online platforms, it is nearly impossible to have an effective influence of own data. Especially with regard to the exchange of anonymised data, it is impossible to know where own data is stored.
Transparency	Taking into account that analytics in Big Data are in many cases not based on information that individual but on data observed or inferred from online activities, locations, smart devices, etc., individuals should be adequately informed. This should include the logic of a data process and the features applied in the context of big data, especially in automated decision-making processes or profiling cases. What would happen if user read all privacy policies for [Websites, Wearables, Apps, Online-Services etc.] collecting and processing personal data? A researcher´s paper has shown that reading all of the privacy policies an average Internet user encounters in a year would take 76 work days.[12]
Digital Identity	With use of digital services, the user´s generate tonnes of information, often representing personal information. The combination of these information result in digital identities or digital footprints. It is obvious that this digital identity is a section of the *"real"* identity of a person. Considering a job application, the interviewer assesses applicants according to various information, either from the application itself or the digital footprint of this person. For reasons of efficiency the digital identity can fundamentally differ from the real identity which results in miserable effects for the applicant assumed the digital identity is based on inaccurate information. Much more are those cases where automated application processes filter applicants according to unsuitable or even discriminatory criteria not transparent for the applicant. The meaning of the digital identity plays an important role it is widely used for many cases. This can be harmful in cases of imbalance affecting

[12] http://techland.time.com/2012/03/06/youd-need-76-work-days-to-read-all-your-privacy-policies-each-year/

	in important sections of the individual´s life. When the digital identity is opposed to the real individual identity, which is known as "*dictatorship of data*", important decisions are not based on the individual´s real actions but all the data the individuals only indicating what the personal probably may be whereas the personal interactions take place after the precomputing the individual from the digital identity.
Accountability	In Big Data domain, accountability may be referred above all to algorithms: several kind of algorithms are used, both on the one hand, using and producing data or information about things and non-human entities (e.g. meteorological circumstances) and, on the other hand, processing or producing data and information directly or indirectly referring to persons, either as individuals or groups. Both of them can affect the lives of humans significantly and bring about important transformations in societies, cultures and societal practices. Therefore, special mechanisms of accountability concerning the making and deployment of algorithms in Big Data setting are becoming gradually more urgent, in line with the GDPR recognition of the importance of accountability mechanisms and closely related concepts, such as transparency, as guiding protection principles. It is difficult to appropriately apply the accountability paradigm, as stated in GDPR, to algorithms operating on Big Data. This is due to the facts that GDPR is fuzzy and such algorithms are complex and often operate on a random group-level, that may imply additional difficulties when interpreting and articulating the risks of algorithmic decision-making processes. In Big Data environment, in light of the possible significance of the impact on human beings, accountability mechanisms now inherent in the regulation need to be interpreted and applied considering the complexities and the broader scope of algorithms.
Fairness, Trustworthiness	Trust affects the relation between the user and the data processor. In this relation the user expects a service a performance by the data processor, whereas in return the data processor will be capable of collect and process his personal and non-personal data. One aspect to be considered is the ethical issue of imbalance between the individual and the data controller. Having the knowledge and control over the data process and used technology, the user cannot fully overview the process of his data. Generally, there is a divergent or opposing field of interest between the user and the data controller. The different interests and roles in this relation lead to an unequal balance of power in favour of - in most cases - the data process. Insofar, the less powerful party has to be properly respected by the opposite party, as in return the user can trust in the data processor. For example, using wearables generates amounts of health data monitoring the physical condition, tracing the runs, speed etc., representing the source for individual services. Insofar, between the

	relation of the individual and the data controller exists a contract to provide this service in particular limits. Both the individual and the data controller have an interest in this data. On basis of this relation, it would be highly problematic if the data controllers sells health data, even anonymised, to a thirds person, like an insurance company or a credit institute, for further processing.
Awareness	The possibilities and implications of digital services in the life of the individual is clearly noticeable. This ranges from social media platforms like *Youtube* or *Facebook*, online shopping services like *Amazon* or *eBay* and all kinds of platform to exchange and trade all kinds of goods. During the use of all these services, there is a huge business around for companies that dependent of the user´s interaction which generates data, a fundamental resource nowadays. The necessity of collecting and storing data for realizing certain kinds of business models overlaps with the privacy concern of individuals. Insofar, data controller in business and companies should be aware of this intersection in a way that respects the user´s privacy concerns. She should not be degraded to the pure object of the business´s purpose, but interact with the user´s privacy concerns. Much more, the data controller should not only be aware of his own responsibility, but make the user aware of what the use of particular online services and the disclosure of personal information means for his privacy. Concretely, the data controller must enhance the individual´s power and freedom by comprehensively explaining which data is collected and how it is purposed. This results for the data controller in being aware of his power he gains when collecting all kinds of information about the user and in accepting the duty of making the user aware what his presence in the digital world and the use of online services means for his privacy. Maybe helpful for the awareness principle is to formulate a respective intern policy about the handling of data, the analysis of data and the consequences with respect to harm, the responsibility, the legal borders etc. These points intend to protect the individual from: ● Disclosure sensitive or harmful information. ● Lack of information about the collection, process and use/purpose of personal data. ● Including sensitive information in legal decisions that are not wanted to be revealed by the person concerned. ● not making use of his/her rights. ● arbitrariness of the data processor.

Ethic Risks and Ethic risk table

Ethical and societal risk	Description	Ethic Risk Management
Personal data misuse	Sensitive personal data (credit card transactions, call detail records) shall only be used for the purpose of the project and not for any other reason, despite any interest from third parties	⇨ Access to data strictly denied to any other than the authorised key persons ⇨ Confidentiality statement to be signed by all members involved in data gathering, analysing and reporting
Discrimination and Profiling	Considering that automated decision making has been explicitly mentioned and regulated in the General Data Protection Regulation, it is obvious that its impact - predictions that have been made with mathematical-stochastic methods having significant legal effect on the data subject - for the individual is disproportional. With regard to the principle of accuracy, data has to be kept up to date which requires the data controller to take every reasonable step to delete inaccurate data if necessary for the respective purpose. The relevant question concerns the practical implementation of data quality standards. A common misunderstanding is based on the assumption "Quantity over Quality" meaning that the mass of data equalizes a bad quality data.[13] In fact, using a mass of wrong data in machine learning procedures result in the worst case in wrong conclusions and predictions.[14] What this means for procedures whose prediction values are dependency	⇨ Development of transparent machine learning[17,18]

[13] http://blog.iese.fraunhofer.de/quantity-over-quality-mistake/

[14] http://blog.iese.fraunhofer.de/quantity-over-quality-mistake/

[17] Gerd Gigerenzer, Klaus-Robert Müller und Gert G. Wagner "Algorithmen müssen anhand von Beispielen transparent gemacht werden"

[18] Grégoire Montavona, Wojciech Samekb, Klaus-Robert Müller "Methods for interpreting and understanding deep neural networks", Digital Signal Processing 73, 2018, pp. 1–15

	for granting a credit or calculating insurance fees are obvious. Insofar, it is highly recommended, not only to comply with the GDPR and avoid civil claims, to establish data quality standards and data management policies for all data processed. Considering the issues of automated decision-making, where the individual is faced with high complex procedures, whose outcome in form of decisions can represent a high relevance for the data subject. What is relevant in such cases and analogue cases, is the fact of requesting information about such procedures. The advantage of efficiency from automated procedures is the low expense of costs and saving in time. However, efficiency issues are no argument for interfering with individual´s rights and interest, which requires rights and obligations as compensation. In cases of profiling (and scoring) respectively automated decision findings, article 13 sec. 1 lit. (f) explicitly prescribes the data controller to present sufficient information, This includes the logic of how decisions have been made, in order to enable the data subject to influence in this decision process, e.g. by identifying incorrect data or questioning certain procedure as discrimination.[15] Challenging is the way of understanding what algorithmic systems really do (in the case of neural networks not even the programmers can easily understand whats going on in the network) SEE ATTACHT PAPER !	

[15] E.g. discrimination of skin colour for choosing a job candidates.

	and presenting information about this procedure, as the complexity has to be reduced in a comprehensive language and appropriate content level (granularity). Intention shall be that the majority of people is able to understand the procedure to make use of the rights provided by this regulation. As indication of the extent of what information have to be provided can be taken from the recital 63 of GDPR. It is stated that information should include "*the period for which the personal data are processed, the recipients of the personal data, the logic involved in any automatic personal data processing and, at least when based on profiling, the consequences of such processing"*.[16] The scope of *involved logic* concerns the underlying algorithm but is restricted to the protection of business and trade secrets, so that it is questionable how this right contains in specific individual cases. Referencing the judgement of the federal supreme court concerning the German right to information (§ 34 BDSG), subject of the case was the scope of information that have to be provided by the SCHUFA – a German credit investigation company. The disclosure of information about the calculation and used algorithms was denied for reason of higher interest. None the less, an obligation to disclosure information comprises the values of data, additionally creation of those. The data controller shall provide as	

[16] https://gdpr-info.eu/recitals/no-63/

	much information as the data subject is able to understand the reasons for using which data as calculation basis in order to explain relevant data and circumstances. By this, the data controller is forced to personally deal with individual case and integrate human valuation. The extent of information to be provided is dependent on the individual case, whereas weighing the interest of the data subject and the data controller is decisive relevant for the outcome.	
De-Anonymisation	Anonymising personal data is not the overall goal of data protection, as re-identification becomes more problematic with stronger algorithm and extremely wider data storages. With combining various non-personal respectively anonymised data with advanced analytics, there is the possibility to infer information related to a person or a group. 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-anonymisation 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*". Legally problematic is the assignment of data as anonymised data, as data is held as anonymised, if due to the personal,	The AEGIS project uses an anonymisation tool supporting data publishers in removing all personal identifiers in the datasets. It is important to remember that 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[19], even if all identifiers have been removed.

[19] Which means more types of data is present in the data storage.

	temporal, technological and financial cost are that much that de-anonymisation can no longer be anticipated. Insofar, the aforementioned criteria are, due to the rapid technological progress, decreasingly devalued, that de-anonymisation is likely to be expected and realizes procedures like profiling and automated decision-making. Consequently, relevant criteria for the determination of the presence of personal data must be the potential circumstance of de-anonymisation in Big Data. In particular, it is not appropriate to consider the effects of de-anonymisation only with the presence of the harm.	
Arbitrariness	Big Data analysis stand for seeking and revealing interest and conspicuous patterns. The intention is to make unnoticed characteristics visible. When using Big Data technologies, the purpose is generally worded, as the target of Big Data analysis is to generate as most results as possible. For instance, when analysing customer data, it is not defined in which specific domain a better strategy can lead to improve customer satisfaction or an increasing sales number. The results must be interpreted by experts to determine appropriate conclusions. Insofar, there is a span between the interest of the data subject to know for what purpose of his own data are used and the data controller to gain results helpful to enhance and improve his business model/strategy. Overall, the tension lays in the "*predictability and legal certainty regarding the purposes of the data*	1. The data controller describes and evaluates specific risks concerning the processing of personal data regarding the intention of the purpose limitation in advance in order to identify dangers and negative consequences, e.g. personal data from wearables are sold to insurance companies or monitoring employees for performance analysis instead for safety reasons. 2. The data controller identifies and describes technical and organization measures, e.g. Anonymisation/Pseudonymization, separation of databases/data categories, supervision of data mining processes, separation of role and functions of those employees evaluating the data mining results according to the purpose of the data/result (access-control) 3. In accordance to the risk assessment, the data controller has to assess the effect of technical and

	process on one hand, and the pragmatic need for some flexibility on the other". The main problem in Big Data is to identify a purpose for the process of those data which are necessary to determine a legitimate purpose, resulting in a circular reasoning. The purpose determination has to be determined before the data process is performed. The purpose limitation principle does not provide the possibility to determine an appropriate purpose afterwards, with the intention to avoid arbitrariness. This leads to a reduction of the Big Data capability, although a consent could be gained with subsequent approval.	organizational measures regarding the risks for the person concerned, additionally certain provisions for guaranteeing their compliance. 4. Machine learning should be done in transparent way.

1.1.4. Approach of the data protection impact assessment

The approach of this deliverable is to explore the compliance with the regulatory and ethical framework and economic and societal impact of AEGIS within the three demonstrator cases. The overall regulatory framework, taking into account the ethical and data protection requirements, provides an overview about AEGIS vision with regard to the economic and societal impact including AEGIS components and workflow necessary for the understanding of AEGIS functionality as well as critical stages that are attached to the compliance of data protection regulations. A practical view is given with the demonstrator cases in the following subsection, representing the overview, goals and approach.

The ethical framework provides ethical issues that are likely to address the privacy of individuals. Insofar, the concept and integral components of *privacy* are the central part of the ethical chapter. Besides the ethical concerns, the advantages of AEGIS are mentioned in a comprehensive manner consisting of the solutions AEGIS provides and the value for the society by AEGIS application. With this, the ethical concerns and the welfare represent the two poles of AEGIS, so that the impact assessment has to provide a balancing assessment, contrasting the privacy concerns with the societal welfare. Moreover, this document should suggest strategies or safeguard – solutions by which harm for privacy can be avoided while simultaneously profit from AEGIS value or even bypass particular contradictions.

The legal chapter provides a data privacy impact assessment comprising the legal standard that AEGIS has to comply with. This assessment is necessarily restricted to a general legal data impact assessment focusing on AEGIS as Big Data application as whole. Concrete statements

shall be elaborated with means of the three demonstrator cases are used as reference point for the application of the specific legal requirements and obligation. A questionnaire containing relevant questions of the regulatory framework is used as an interface to prove compliance of the demonstrator cases and its related data processing operations with the ethical and data protection requirements set out in D1.2/D1.3. This should be the first assessment verifying the compliance of the demonstrator cases concerning the data protection principles, in which way the rights of the person concerned are implemented and practicably applicable, as well as the compliance with obligations addressing the data controller. It should be noted that this document does not represent a generic data protection impact assessment for all use cases of AEIGS to comply – when required – with article 35 GDPR, but a data protection impact assessment in general. In accordance to Article 82 GDPR liable for violations against of the data protection regulation is the data controller conducting data processing. As stated in the concept declaration, AEGIS "will support the complete range of data processes", responsible for the compliance of the data protection regulations are those who conduct the overall data process and the process of personal data.

These suggestions have taken into account AEGIS specific issues and are intended to serve as reference or guidance for a specific data protection impact assessment for a particular use case. For example, when the data controller consults a data protection officer, especially when according to article 37 GDPR, the designation of a data protection officer is prescribed, this assessment may be relevant.

The assessment guidance is described in the following subsection of this document and is used as basis for the application of the assessment for the demonstrator cases.

1.1.5. Questionnaire for the ethical and data impact assessment

Requirement	**Description**
1. Explain your compliance with processing data with legitimate aim and the purpose limitation principle! 1.1. For every purpose, explain how your data process, based on one specific purpose complies with the requirements listed below and what part AEGIS services play! 1.2. Do you plan to process personal data for purposes other than the original purpose? If yes, explain the compatibility of both purposes! 1.3. Explain in which way your implemented technical and organizational safeguards enhance the purpose limitation principle!	Required is that every data process is restricted to one legitimate, specific and explicit purpose. The purpose limitation principle states that processing personal data is only allowed/lawful with proof of a explicitly determined and lawful purpose adequate and relevant for one particular data process This requirement implies that: • 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.

1.4. Do you assure that no personal data collected will be sold or used for any purposes other than the current project?	The requirement of purpose limitation 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".
2. Explain your compliance with the Data Minimization principle and Proportionality requirement! 2.1. What categories and how much data do you collect in order to achieve the respective purpose(s) of your data process(es)? 2.2. Do you anonymise data for the data process? 2.3. Do you have anonymisation tools directly integrated in the data process? 2.4. How do you address the issues of de-anonymising data as linking various kinds of datasets can result in the presence of personal data and the application of the data protection regulation? 2.5. How do you secure data from getting De-Anonymised?	The data minimization principle is set forth by Article 5 I lit. (c) of the general data protection regulation, as personal data shall be: "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed". This requirement is also quoted by the DoA (Section 5.1.1): "A data minimization 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 data process shall be restricted to the minimum amount necessary to fulfil the pursuit purpose of the data process. According to this, the data process needs to be appropriate, substantial and restricted to the minimum amount necessary. 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 will be stripped of any direct identifiers and, in addition, adequate technical and organizational safeguards have to be taken for mitigating the risks of re-identifying the individuals.
3. Explain your compliance with data Storage/Retention, Minimization and guarantee integrity and confidentiality! 3.1. How do you implement data retention policies in order to keep data in storage only as long as necessary for the respective purpose?	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". This is necessary, also promoting data minimization, as binding personal data to a particular purpose guarantees transparency for the data subject and

3.2. Do you anonymise personal data after the purpose of a data process has been achieved? 3.3. Do you comply with the following requirement, stated in the DoA: "After the end of the project, all collected data that can related to individuals will be deleted from the platform"?	avoids arbitrariness of the data controller to use data afterwards to a not specified purpose. 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 effective authentication techniques. Furthermore, 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.
4. Explain your compliance with fair data process including the avoidance of discrimination, harm and social sorting! 4.1. In case of conducting automated processing applications according to article 22 GDPR 4.1.1. What is the purpose of this processing activity? 4.1.2. What (legal) relevance does result from the output for the individual? 4.1.3. What categories of data (kind of information) are used? 4.1.4. Is the application of automated decision - making necessary and adequate in the specific context? 4.1.5. In case that this kind of processing unfolds legal effect: 4.1.5.1. What is the proportion of the purpose unfolding legal effect to the individual´s impact in his rights and interests?	The Consortium has to avoid that AEGIS demonstrators or AEGIS overall system facilitate discrimination (race, gender, age, religion, disabled) especially in automated decision processing or social sorting in sense of profiling. 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*". 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

4.1.5.2. In which way do you respect and enforce the individual´s rights and interest in your particular case? 4.2. In case of creating profiles of individuals, how do you handle these artificial profiles for automated decision-making and making decisions without the use of automated processing by humans 4.3. What safeguards – technical and ethical - are implemented to avoid discrimination and stigmatization and the use of wrong information	automated processing means that the decision does not contain any human valuation or action decisive for the decision. 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. This does especially apply for decisions which unfold legal affect for the individual. Legal relevance occurs whenever the legal position of an individual changes in a way, constituting or revoking a right or legal relationship, e.g. rejection of a job-application or calculating insurance fees. 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. Thus, transparency of the algorithms is an important issue, which can be addressed by transparent machine learning approaches.
5. Explain your compliance with the assignments 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 circumstances set forth by Article 37 of the 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.
6. Explain your compliance with an informed consent and other conditions!	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

6.1. Explain in which way you fit the following conditions of gaining consent: • Clear and unambiguous consent • Individual case • Informative consent • Purpose limitation – specific, explicit and legitimate - expression must be intelligible and distinctive, referring "clearly, precisely to the scope and consequences of the data processing". • Form of consent • Comprehensibility of the consent declaration • Clear and plain language • Voluntariness and free exercise of choice and avoidance of coupling the collection of personal data: The collection of personal data shall not be coupled with the performance of contractual obligations if not necessary for the data process and any sort of intimidation, coercion or risk of negative consequences in accordance to article 7 IV GDPR • No compulsion • Possibility of repealing • Documentation of gaining consent • Timing - the consent has to be given before the starting of the processing; 6.2. How do you ensure that a broad mass of people is capable of understanding how and for what purpose their personal data is used? 6.3. How granular is your consent description in relation to the complexity of your data process?	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 general requirements of gaining consent have been formulated in article 7 GDPR. The data subject shall be capable of comprehensively understanding what amount and in which way his personal information are processed, from whom and for what purpose.
7. Explain your compliance with the 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. 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.
8. Explain how you respect the data subject´s rights and interest and facilitate the data subject to make use of his rights granted by the GDPR! 8.1. Can you provide a copy of personal data from the data subject if requested according to article 15 III GDPR? 8.2. Explain your compliance with the information obligations according to the GDPR 8.3.Do you provide the following information at time of collection: o Identity and contact details of your company o Purposes of data processing and respective legal base o The recipients or categories of recipients o Transfer of personal data to third country or international organisation o Period for which the personal data will be stored o **Reference to the data subject´s rights – aforementioned question** o Whether the provision of personal data is a statutory or contractual requirement o The use of automated decision-making o Whether processing data is intended on basis of a purpose other than that for which the personal data were collected 8.4.Do you inform the data subject about his rights granted by the GDPR: o Article 15: Right to access o Article 16: Right to rectification o Article 17: Right to erasure o Article 18: Right to restriction o Article 20: Right to data portability	The main categories of data subject's rights relevant to AEGIS can be split into two categories: 1. Rights of information 2. Rights of intervention (including rectification and erasure as well as, according to the new Regulation, data portability). The second 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 intervene if necessary. The chance to withdraw the consent can be attributed to this category. 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*". 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, as essential requirement for the data subject for being able to execute certain rights according to the GDPR, is that the data controller provides as much information as necessary to ensure that the data subject has a chance to be appropriately informed.

○ Article 21: Right to object	
9. Explain your compliance with data quality including data Accuracy and data security! 9.1. How do you monitor the correctness of data quality in your processing activities? 9.2. How do you ensure an appropriate level of data quality? 9.3. Do you use data quality policies to ensure protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures? 9.4. Do you use data quality policies for corresponding data purposes? 9.5. Describe your data security measures and other tools used to protect data against external attacks! What are your data security strategies to protect your technical infrastructure from harm and external attacks? - E.g. database encryption, encrypted transmission of personal data etc.	The 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" Accuracy according to this definition 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. 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. In case of inaccurate datasets, the data subject has the right of correction or erasure of wrong datasets, according to article 16 GDPR. 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. 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. 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". 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 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.
10. Explain your compliance with Privacy by default and privacy by design 10.1. Describe how your default settings comply with the "privacy by default" requirement 10.2. What risks have you identified that are likely to occur in your data process and what rights and interest of the data subject are affected thereby? (You can find a list of specific categories based on recital 75 of the general data protection regulation.) 10.3. If particular risks have been identified, what is the probability of its occurrence, the possible harm and the degree of harm? 10.4. What tool and techniques as used to anonymise persona data? 10.5. How do you consider and approach the problem of De-anonymisation? 10.6. What technical and organizational safeguards did you implement to protect	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 According to Article 25 of the Regulation, 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" Privacy by Design is at the core of AEGIS

personal data in accordance to the identified risks 10.7. Do you have appropriate data management strategies and/or data access strategies and/or data retention policies for the handling of personal data? 10.8. Do you use automated policies monitoring the compliance of a data process regarding e.g. the purpose exists, the minimum amount of data has been collected, interferences of request from the data subject etc.	approach for the elicitation of privacy and data protection requirements, whilst data protection by default is coherent with data minimisation requirement. The GDPR integrated an "*harm-based approach"* which means, 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. This requirement also implies adopting anonymisation as much as possible. 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 of the art encryption methods". 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 minimizing 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.
11. Explain your compliance with the 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
12. Explain your compliance with the application scrutiny to local/national boards if required by national legislation concerned!	As regards the demonstrator, "authorisation or notification by the National Data Protection Authority must be submitted, where applicable". 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). 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
13. Explain your compliance with 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 will 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 authorized persons, contractually obliged to keep confidentiality and meet data security rules. It is recommended an authentication and authorization infrastructure in AEGIS. 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.
14. Explain your compliance with the 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. iv) Volunteer have the possibility to interfere with the – to contact or complain and repeal the - participation processing activities.
15. Explain your compliance with use of 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. 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?)

2. AEGIS OVERVIEW - ARCHITECTURE AND DEMONSTRATORS

The following sections presents a short introduction of AEGIS architecture as well as a short demonstrator description. Thereby, it is necessary to explain how AEGIS is designed. Then with presenting the three demonstrator cases, which are the first application examples of AEGIS and so the first key reference for the legal and ethical analysis, the social impact and ethical relevant questions are going to be outlined, as the facts from the previous sections are used for a comprehensive application of legal standards and ethical issues of AEGIS. This exploration shall identify where the critical stages in the design layer are and what meaning this has for the demonstrator cases, which means for issues and questions the data controller is confronted with.

2.1. Intention of AEGIS

The Horizon2020 project AEGIS on Advanced Big Data Value Chains for Public Safety and Personal Security, brings together the data, the network & the technologies to create a curated, semantically enhanced, interlinked & multilingual repository for public & personal safety-related big data. It delivers a data-driven innovation that expands over multiple business sectors & takes into consideration structured, unstructured & multilingual datasets, rejuvenates existing models and facilitates organisations in the Public Safety & Personal Security linked sectors to provide better & personalised services to their users. In early the stage of this project, relevant stakeholder groups have been identified, namely Smart Insurance, Smart Home, Smart Automotive, Health, Public Safety/Law Enforcement, Research Communities, Road Construction Companies, Public Sector, IT Industry, Smart City and End Users. The use-cases ranges from assisted and personalized driving, improved GPS navigation (including road-conditions, weather situation, social events, traffic situation and safety risks in nearby location), predictive maintenance to autonomous driving systems. Focusing on the welfare and protection of the general and the public, AEGIS intends to enhance the prevention and protection from dangers affecting safety such as accidents and disasters.

AEGIS will provide a central platform and a set of additional tools to realise the integration and combined analysis of a plethora of diverse data source, cross-domain and cross-lingual, having different formats and conforming to various standards.

2.2. AEGIS implementation

The AEGIS platform is a modular system, composed of several components. Each component is designed for a specific task and offers a well-defined set of features. The functionalities of the components align with the AEGIS Data Value Chain (Figure 1), which defines the high-level workflow of the AEGIS platform.

Figure 1 - AEGIS Data Value Chain

This workflow is described in detail in Deliverable 1.1 (AEGIS Data Value Chain Definition and Project Methodology) and can be seen as basic guideline and ideal application of the AEGIS components. The users are not required to follow every step. A minimal data lifecycle may only include Data Acquisition, Data Storage and Data Usage.

However, in the following the particular components are shortly described with an emphasis on data control, data management and security. In general, the AEGIS platform provides two distinct types of data processing:

- Offline Data Processing is executed by on-premises software. I.e. AEGIS components which are installed within the infrastructure of the organization and only the outcome of any processing may leave the premises. This applies for very sensitive pre-processing and ensures that organisational data protection mechanisms can be applied.
- Online Data Processing is executed on the AEGIS platform, which is available as a service and runs outside the organisational premises.

It is to mention that the AEGIS platform only offers tools and software components. How these artefacts are effectively used in an overall data process architecture, is decision of the data processer. Consequently, the data processor is responsible of the effect his data process can or will have. Though, the AEGIS platform will always ensure the complete control of user-generated and user-provided data to the respective owner. This specifically includes the complete removal of data and derived data in any moment.

2.2.1. Offline Data Processing

The following tools will be available for on-premises installation and use.

Anonymisation Tool

The anonymisation tool is an extensible, schema-agnostic component that allows real-time efficient data anonymisation. The purpose of the anonymisation is to enable the potential value of raw data in the system by accounting for privacy concerns and legal limitations. The tool will help the user generate an anonymised dataset as an output, making sure that that the individual sensitive records or subjects of the data cannot be re-identified. The tool is connected to the AEGIS online platform and the anonymised data can be directly provided to the platform. The tool assists in ensuring an adequate anonymisation, but relies on human supervision in interaction to ensure compliance.

Data Storage	The component does not store any data.
Data Processing	The component allows the manipulation and editing of the input data.
Access Control	The component can only be accessed by privileged users set by the defied administrator of the component.

2.2.2. Online Data Processing

The following tools will be available on the AEGIS online platform.

Data Harvester

Data Harvester is the component enabling the import of data from heterogeneous sources and their transformation to the required AEGIS data format and structure. The generated data is stored in the Data Store.

Data Storage	The component does not store any raw data, but short-living processing log, interface data about the sources and transformation rules.
Data Processing	The component allows the manipulation and editing of the input data.
Access Control	The component can be accessed by registered users of the AEGIS platform. Harvesting pipelines can only be accessed by their original creator. This access right can be granted to other users by the original creator.

Data Annotator

The Data Annotator is the component in the AEGIS platform that is responsible for interactively equipping input data with suitable metadata. The generated metadata is stored in the Linked Data Store.

Data Storage	The component does not store any raw data.
Data Processing	The component allows the creation, editing and deletion of metadata. For the purpose of assisting in the creation of the metadata the tool will analyse the raw data associated with the metadata.
Access Control	The component can be accessed by registered users of the AEGIS platform. Creators of metadata can grant access to it to other users, by granting access to the associated data.

Data Storage

The AEGIS data storage component is responsible for storing data that was collected and curated by the Harvester. A distributed file system approach was chosen for flexibility, reliability, and scalability. The distributed file system allows the storage of large amounts of data while enabling access to the file from other AEGIS supported services such as the Query Builder and the Visualizer.

Data Storage	This component stores raw data in the form of datasets, as well as logs and other results of computations.
Data Processing	This component does not perform any data processing.

Access Control	The component can be accessed by the users of the AEGIS platform. Users can access their own data as well as data that is shared with them. Within the Aegis platform data is aggregated in datasets and then datasets are aggregated in projects. Users can be added to projects and will thus have access to all the datasets present within the project. Datasets can be shared with other projects without creating new copies of the data. Any user that has access to a dataset can further share it with other projects.

Linked Data Storage

The Linked Data Storage is responsible for storing the metadata associated with a particular dataset within the AEGIS platform. This metadata poses the foundation of the processing of the data within the AEGIS platform, since it offers detailed information about the semantic and syntax of the data itself.

Data Storage	The component does store the metadata of the AEGIS platform.
Data Processing	The component allows the creation, editing and deletion of data.
Access Control	The component can be accessed by registered users of the AEGIS platform. The metadata is private and can by default only be accessed by the data owner, who can grant access to other users.

Query Builder

Query Builder is the component that provides the capability to interactively define and execute queries on data available in the AEGIS system.

Data Storage	The component reads data out of the data store and allows users to save processed datasets back to the data store.
Data Processing	The component allows the manipulation and editing of the input data.
Access Control	The component can be accessed by registered users of the AEGIS platform. Query Builder can only access datasets available under the AEGIS project within which the user is using the component. This access right can be granted to other users by the original creator of the project.

Cleansing Tool

Data cleansing is an umbrella term for tasks that span from simple data pre-processing, like restructuring, predefined value substitutions and reformatting of fields (e.g. dates) to more advanced processes, such as outliers' detection and elimination from a dataset.

Data Storage	The component reads data out of the data store and allows users to save processed datasets back to the data store.
Data Processing	The component allows the manipulation and editing of the input data.
Access Control	The component can be accessed by registered users of the AEGIS platform. Query Builder can only access datasets available under the AEGIS project within which the user is using the component. This access right can be granted to other users by the original creator of the project.

Algorithm Execution Container

The Algorithm Execution Container is an adapted version of the Zeppelin notebook that is configured to run on top of the AEGIS infrastructure, used with preloaded MLlib library that is able to provide a set of algorithms for data analysis. The module is provided as pre-compiled paragraphs of code in Zeppelin in order to serve users with a graphical interface for running their analyses.

Data Storage	This component stores outputs of analyses performed over the AEGIS platform (analysed datasets and analysis performance metrics)
Data Processing	This component processes data according to the analytics algorithm that is selected.
Access Control	The component can be accessed by the users of the AEGIS platform, who are able to load to the component datasets to which they have access (that belong to their projects).

Visualizer

Visualiser is the component enabling the visualisation capabilities of the AEGIS platform for the output of the analysis results produced by the Algorithm Execution Container as well as for the output of the querying and filtering results coming from the Query Builder. The produced visualisations are presented to the user via the AEGIS Front-End.

Data Storage	The component does not store any raw data. The generated visualisations can be stored as an image.

Data Processing	The component does not allow manipulation or editing of the input data.
Access Control	The component can be accessed only by registered users of the AEGIS platform. Additionally, the component can only access and generate visualisations of the results of either the Algorithm Execution Container or the Query Builder.

Brokerage Engine

The Brokerage Engine is a component of the AEGIS platform that is responsible for recording transactions performed over the platform in a blockchain ledger, and also checks artefacts of the platform against their attributes coming out of the Data Policy Framework to resolve if certain operations are allowed.

Data Storage	This component stores details regarding transactions performed over AEGIS in a distributed ledger.
Data Processing	This component processes data coming out of the blockchain ledger to resolve transactions' feasibility.
Access Control	The component is inaccessible from AEGIS users for direct interaction, and works on the background, facilitating in an automatic manner the logging of transactions and the exchange of data and other related artefacts as described in D2.1.

2.3. AEGIS demonstrator cases

The bandwidth of stakeholders and AEGIS intention of a better public safety and personal security is covered by the following three demonstrators.

In order to classify the following demonstrator cases in an appropriate scheme, it is necessary to define criteria corresponding with the legal and social issues of this document. This includes to formulating accurate questions relevant for a legal assessment. Thereby critical stages of AEGIS infrastructure should be outlined.

2.3.1. Demonstrator Case 1: Smart Automotive and Road Safety

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.

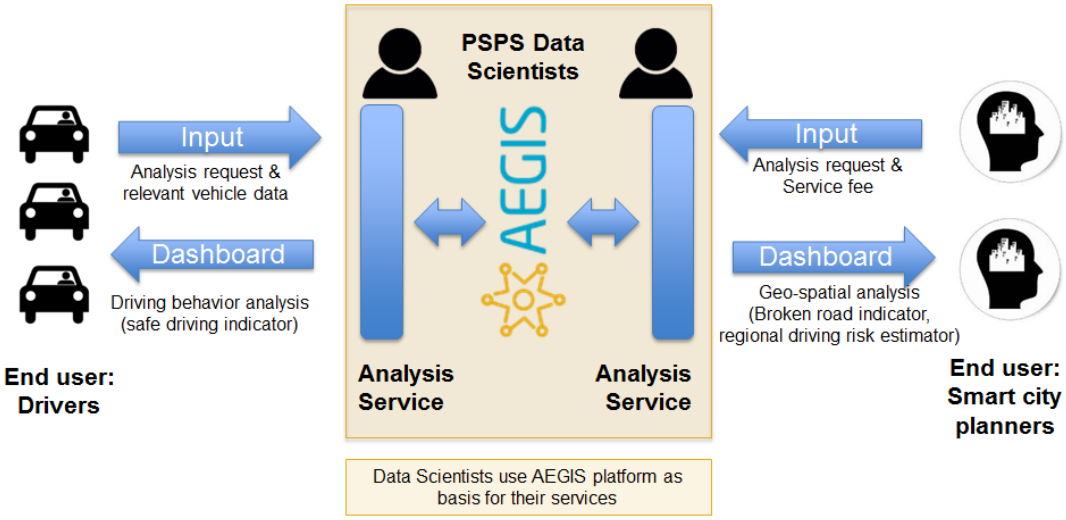

Figure 2: Actors of the automotive and road safety demonstrator

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[20], 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.

[20] https://www.data-protection-authority.gv.at/

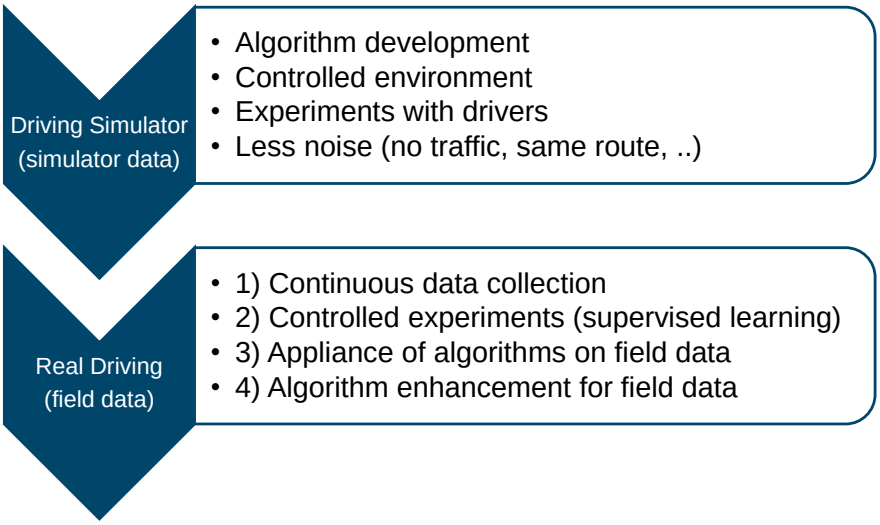

Figure 3: Simulator data and field data

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.

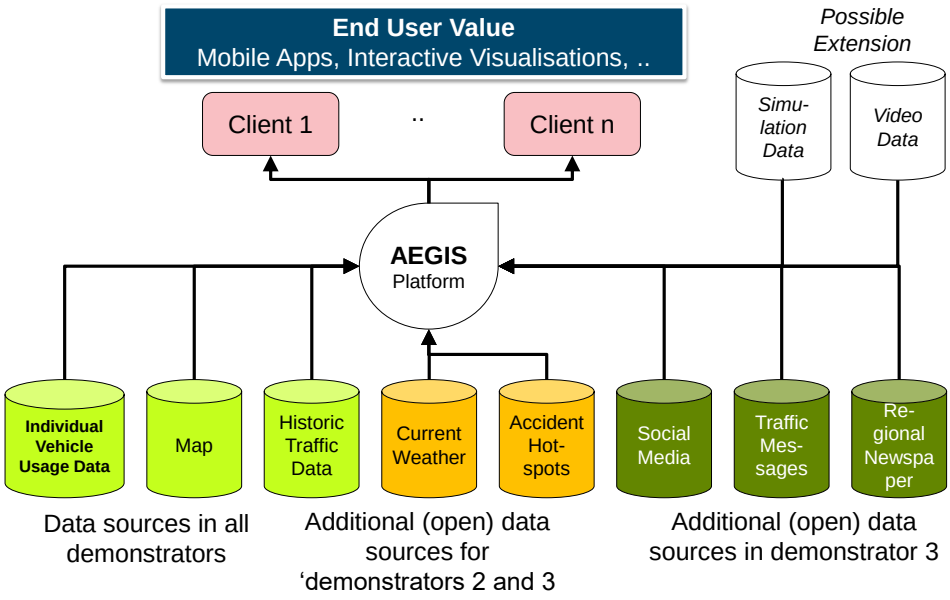

Figure 4: Data sources relevant to the automotive demonstrator

Data lifecycle

Collection • At VIF	In the automotive demonstrator vehicle operation data is collected by a number of voluntary people employed at VIF, using a data logger developed at VIF connected to the OBD2 interface of their

	vehicle, which is operated in Greater Graz area. All voluntary participants signed an informed consent and are well informed about the envisaged data collection process, the purpose of the data collection, the nature of the collected data, and the AEGIS project. Collected vehicle data is then manually exported from the data logger by a member of the AEGIS project team and stored in an access-restricted AEGIS project folder located on a file share serviced by VIF's IT department. Only the AEGIS project team has access to this folder. The exported vehicle data collected is time series data and consists of four comma-separated value (CSV) files per export to be stored into a folder (obd.csv, acc.csv, gyro.csv, trip.csv). Driver and vehicle information are anonymised (and they are not stored in the same place)
Preparation • At VIF for algorithm research • At the AEGIS platform (later)	All further data processes will be conducted on the AEGIS platform. However, a technical feasibility study including the development and evaluation on the vehicle data-specific algorithms is conducted offline at VIF. For this purpose, a particular data preparation process has been designed at VIF to be implemented on the AEGIS platform in the course of the project. In a nutshell this transformation process is to transform exported vehicle data from tall to wide format, join the four tables, split the results (all measurement data) into single trips, interpolate the measurements onto a regularly spaced time grid, rotate acceleration and gyroscope data to align it with the coordinate system of the vehicle, and finally save transformed data into one file per trip (trip 1 – trip x). Vehicle data will be uploaded as raw data to the AEGIS platform and there is no (offline) preparation process of vehicle data envisaged in the first step. In particular the same process as described above will be conducted on the vehicle platform in the data processing step.
Input • At the AEGIS platform	A demonstrator project is created on the platform and raw vehicle data is uploaded manually to the AEGIS platform by an AEGIS project team member in bulk into a demonstrator specific sub folder. The uploaded vehicle data is exactly the vehicle data exported and consists of four different comma separated value (CSV) files per export (obd.csv, acc.csv, gyro.csv, trip.csv) containing the sensor measurements as time series data.

Processing • At the AEGIS platform	While the data processing procedure has been also explored offline at VIF in a feasibility study to develop and evaluate the vehicle-data specific algorithms, the main data processing will obviously been done on the AEGIS platform. The first step data processing procedure includes to transform raw data from tall to wide format, join of the four tables, split the results (all measurement data) into single trips, interpolate the measurements onto a regularly spaced time grid, rotate acceleration and gyroscope data to align it with the coordinate system of the vehicle, ant finally save transformed data into one file per trip (trip 1 – trip x). The second step data processing procedure includes detecting the events hidden within vehicle data (which is the core of the demonstrator) as well as including data from further sources (weather data and map data). These events are related to identify road damage, identify patterns related to driving safety, and to driving risk.
Output • At the AEGIS platform	The detected events are visualized by using mainly a heatmap (with/without bubbles on the geo-location of a detected event) as well as tables including the events with meta information (e.g. speed, time, …) to be finally shown to drivers (e.g. the voluntary participants) and other stakeholders (e.g. road maintenance, city planners).
Storage • At the AEGIS platform	The detected events are further stored within further datasets (road damage data, safe driving data, and risk score data) in the demonstrator folder on the AEGIS platform.

2.3.2. Demonstrator Case 2: Smart Home and Assisted Living Demonstrator

The combined objective of the Smart Home and Assisted Living (SHAL) Demonstrator, as was presented in D5.1, is to illustrate and implement a services bundle towards advanced holistic monitoring and assisted living management, aiming to improve everyday living and enhance the wellbeing of people belonging to vulnerable groups. In summary, a social care service provider, for example a care centre for elderly individuals or a nursing home, desires to exploit big data-driven insights, in order to provide added value services to vulnerable individuals. The services pertain proactive and reactive security and protection through smart notifications and personalised recommendations, as well as indoor comfort and quality preservation. Proactivity and reactivity of the aspired services aim at prolonging self-sufficiency and independence of the at-risk individuals, boosting safety, and facilitating informed decision making, either by the individuals themselves, or by their (in)formal carers. The demonstrator is developed in Athens

by Hypertech, UBITECH and Suite5, Information and Technology (IT) companies which adopt the AEGIS roles of the service developers and data scientists.

The SHAL 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:

a) 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.
b) 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.

Data Process Lifecycle

Collection	The data sources used in the demonstrator are the following: 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. The size of data depends on the time granularity of data acquisition process. It should be in the order of 10-100 MB/day. Personal data are required only for login/identification purposes. The data are collected for the purpose of demonstrating the SHAL services, as these are described above. Any personal data are stored locally and are not uploaded to the AEGIS platform.
Preparation	The collected data are to be subjected to an offline preprocessing workflow, with the primary aim of data anonymisation and dissociation with any sensitive information.
Input	The preprocessed data are to be uploaded to the AEGIS platform for further processing/cleaning and further managing of data (quality assessment, outlier detection, fill missing values, visualization). Secondly, some of the specific da
Processing	The processing step entails the training of regression and clustering models, so as to develop the required event identification engine, which is then employed by the demonstrator.
Output	The result of the processing pertains the provision of a notification service, which will identify potentially harmful events and inform the end-user.

Storage	The datasets are stored in the AEGIS platform for as long as needed in order to perform the requested actions. Data in the local server are stored for a certain time period (e.g. 1 month) and are then deleted.

2.3.3. Demonstrator Case 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 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 exemptions. 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). 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,…

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".

Data Lifecycle

The Insurance Demonstrator implements all of the six steps of the data processing lifecycle; before starting with a brief description of each of these steps, it is important to remark which data are involved in the scenarios proposed for the Insurance Demonstrator (for further information about the scenarios, ref. D5.1 and D5.2).

The data used within the Insurance Demonstrator can be split in two main categories:

- External data, coming as open data from defined trusted websites.
  That kind of data are mainly related to events (e.g. weather, riots etc.) or to stats (e.g. flood risk distribution map, classification of seismic risk areas etc.).
- Internal data, named as in-house datasets, coming from the HDI databases.
  That kind of data are mainly related to customers (e.g. customer data, policy data etc.) or to business data (e.g. agencies, marketing data, claims data etc.).

Hereinafter each step of the data processing lifecycle will be described, highlighting the identified (if any) differences between the two aforementioned data categories.

- Data collection:
  - Within the first two scenarios, external data are searched/generated and uploaded after the detection of an event of interest by the Event Detection tool. In the third scenario, external data are searched and uploaded after a request of an HDI Business Unit. In both of the cases, the HDI Data Scientists refer to a predefined list of trusted online data sources. These sources have been chosen because they are recognized as the Italian official news/stats websites.
  - HDI in-house datasets are collected each time a customer buys/changes a policy or changes his/her data. During the collection phase, the customers shall allow the use of their data for analysis purpose to receive personalised offers, warnings and support for claims.
- Preparation:
  - For what concerns external data the preparation phase could vary depending mainly on the data source; the data could not need any preparation phase, or they could need a preprocessing (filtering, conversion to another format).
  - The preparation phase of the in-house datasets could be summarized as follows:
    1. Selection of the data (columns) of interest for the analysis to be performed.
    2. Filtering of the customers that allowed the use of their data in order to receive personalised offers, alerts and notifications.
    3. Anonymisation: an offline Anonymiser is used to remove sensitive data, and to replace some fields with random strings (e.g. a random ID will be assigned to each customer) or with similar but less explicit data (e.g. the addresses will be replaced with the geospatial coordinates).
- Input:
  The Data Scientist creates a project on his/her workspace of the AEGIS platform and then uploads the data/datasets of interest for that project. The creator of the project that in this case is also the owner of the datasets, i.e. the HDI Data Scientist, sets the visibility features of each dataset/project. In this phase the Data Scientist sets as private the in-house datasets and could allow the access to the project to his/her colleagues.
- Processing (Main part):
  The HDI Data Scientists exploit the Analyser and the Query Builder provided by the AEGIS platform, analysing cross-domain data. Different algorithms are used depending on the scenario (e.g. predictive algorithms, simulations etc.). Once performed an analysis, the outcome, depending on the needs, could be further processed through the Visualizer.

- Output:
  At the end of the processing phase, the Data Scientist fills a form setting some features of the report (e.g. priority, business area etc.) and then pushes the "Share" button. The outputs of the analysis are in three main formats (tabular, textual, graph), depending on the processing phase. Their use is managed by the Engine, a component of the Insurance Demonstrator Architecture (ref. D5.2), that reads the features defined by the Data Scientist and sends them to the proper HDI Operator/Department Web App.
  Once the report has been received and downloaded, it should be deanonymised, allowing the operators to contact the customers.
- Storage:
  The datasets and the final report associated to a project are stored within the HDI Data Scientist workspace of the AEGIS platform.

3. ASSESSMENT OF THE DEMONSTRATOR CASES

3.1. Demonstrator Case 1: Smart Automotive and Road Safe

3.1.1. Compliance with Data Protection

Main aspect of the Smart Automotive and Road Safety is that no personal data is collected and processed additionally to the use of anonymisation procedures. The General Data Protection Regulation can only be applied with the presence of personal data, which is mostly not the case here. Nevertheless, relevant issues and questions have to be taken appropriately into account order to ensure data protection and privacy issues in the same manner. The following assessment refers to the data processing procedures, taking into account how vehicle driving data and other road safety related data is processed with the purpose of valuable insights into driving styles and driving behaviour as purpose of this application.

This includes the anonymisation procedures and implemented design services enhancing privacy and data security.

3.1.1.1. Data protection principles

3.1.1.1.1. Data minimization – avoidance of personal data if not necessary

No personal data is collected and processed. The collection of data is reduced to technical data, which is used to detect the particular events relevant for the analysis. For the experiment, multiple loggers are used on a voluntary base in order to record the participant´s trips, whereby the amount of data collected depends on the length of the trip and the measurement rates.

The anonymisation is conducted while the data is exported from the data logger: The SD card is inserted into the computer. An export script is started asking the exporter to manually provide a *driver_ID* and a *vehicle_ID*. Both strings are then save in the exported CSV files as an additional column. The real names and vehicle names can be cross-checked on a sheet of paper, but they are not stored in a database.

As result, both the drivers name as well as the data of the vehicle are anonymised. A de-anonymisation due to additional information is considered as implausible, only possible with exclusive data e.g. owned by the Republic of Austria and high effort. The participants names are only available to three researchers of the institute working in the AEGIS projects and are locked away from the others.

3.1.1.1.2. Purpose limitation principle

The restriction of collecting data only with a legitimate and precise defined purpose does only apply for personal data. In the Smart Automotive demonstrator, only technical data are collected from data logger connected to the OBD2 interface of a passenger car. These logger files are automatically collected whenever the car is driven. The data of the logger files includes information like speed, data from gyroscope, acceleration data but also GPS and time data from the GPS sensor. Even though GPS data can be used to create motion profiles of a person, the relation of such a profile has to be assigned to a particular person. As the driver as well as the vehicle information are anonymised, the re-identification of this person is only possible with additional data and disproportional effort. It is unlikely that these anonymised data are re-identified in other procedures, as the data workflow consists of the following steps:

1. Manually upload data
2. Transform data using the AEGIS-Platform
3. Save transformed data on the AEGIS-Platform
4. Identify events in the saved data
5. Save identified events to a new datasets
6. Visualize new dataset as an overlay on a geographic map

The demonstrator includes three scenarios:

1. Demonstrator Scenario 1: Broken Road Indicator
   a. Upload technical data (OBD, GPS, ACC, GYRO)
   b. Transform data from tall to wide format, join of the four tables, split the results (all measurement data) into single trips, interpolate the measurements onto a regularly spaced time grid, rotate acceleration and gyroscope data to align it with the coordinate system of the vehicle, save transformed data into one file per trip (trip 1 – trip x)
   c. Detect events in the saved trips (road damage) by applying an algorithm developed by VIF (using RPM, ACC and GYRO data) and save detected events in a new dataset
   d. Visualize detected events related to broken roads on a geographic map (as a heatmap or with bubbles on the position where the event has been detected)
2. Demonstrator Scenario 2: Safe driving indicator
   a. Similar procedure for data transformation
   b. Detect safety critical events in the transformed data (using RPM, ACC and GYRO data), additionally weather data is queried to identify the weather condition at the date/time/location of the detected events, save detected events including the weather information as a new dataset
   c. Calculate a driving risk score for a trip using the saved data and save the risk score in a risk score dataset.
   d. Visualize the detected events related to safe driving on a geographic map (as a heatmap or with bubbles on the position where the event has been detected) and show the risk score.
3. Demonstrator Scenario 3: Regional Driving Safety Risk Estimator
   a. Use data and events from scenario 1 and 2
   b. Use save driving dataset, broken road dataset and risk score dataset
   c. Calculate a two-dimensional density estimate of driving risk and visualize it with a heatmap as overlay to a geographic map

3.1.1.1.3. Storage retention – Restriction of storage

Raw data and technical data are stored in an AEGIS project folder on a file share at VIF. The access is only restricted to (three) selected employees. For the upload to the AEGIS platform, no personal data is included. Data saved on the local file system is ***not*** going to be deleted, as no need requires such an operation. Furthermore, this data shall be used for further data processing which does not interfere with data protection. The need for keeping this technical data on storage is that further processing of this data enhances the purpose of the Smart Automotive project and improves and refines the re-running algorithms.

3.1.1.1.4. Data accuracy, integrity and confidentiality

Data quality depends on the sensor settings and the granularity of the values. The chosen settings satisfy the requirements of the analysis, which is shown in test examples. Outliers and

measurement errors are and will be eliminated during the data transformation, in order to provide "clean" data for the analysis stage. Before analysis, data is transformed and interpolated as preparation of the data analysis.

3.1.1.1.5. Fair data process - Use of automated processing

Taking into account the interest of the participants, no discrimination can be determined. Neither the data process includes discriminative criteria for the Smart Automotive purpose nor does the data process disrespect privacy and other interest of the users. Besides the offered service only produces a service to improve the driving behaviour, no data or data analysis result with personal reference is transmitted to any third party. The purpose as well as the test results are very well explained to the participants and the results are provided, whenever requested. Beyond, no automated decision making or profiling is performed and planned to be part of the offered service in the future.

3.1.1.2. Data processing

3.1.1.2.1. Lawful – legal basis and its conditions

As no personal data is collected, only a project explanation satisfying the general information requirement is given to the participants. The relevance of an informed consent does not apply in this demonstrator case.

3.1.1.2.2. Comply with obligations, especially information obligations

The obligations of the GDPR have no relevance here. Considerations for obligations can refer to respect and protect privacy and fairness issues which means to restrict the data process to the purpose of enhancing the driving behaviour and security of the end-user. Further use of the collected technical data are – at the moment – intended to improve the offered service – faster, more accurate and more precise. This interest belong are assigned to the end-user´s sphere as well as VIF as scientific institution.

3.1.1.2.3. Respect data subject rights and facilitate its execution

Analogous to the aforementioned section, only the technical execution including aspects of the AEGIS platform have been communicated.

3.1.1.3. Technical and organization measures

3.1.1.3.1. Technical and organizational safeguards – privacy by design

The storage solution of the AEGIS platform adopts HopsFS[21], a new implementation of the Hadoop Filesystem (HDFS), offering advanced security with a plethora of authentication mechanisms as well as data access control, data integrity and data consistency mechanisms. HopsFS is making use of checksum to ensure security and integrity control of the data in storage. The encryption of data is not used to avoid the efficiency problems within the data analysis process in the AEGIS big data ecosystem.

[21] http://hops.readthedocs.io/en/latest/user_guide/hopsfs.html

Concerning the security of data in transit or data in motion, which includes data transfer between the AEGIS services and clients either within the internal network or through the internet, AEGIS is providing data encryption via Secure Sockets Layer (SSL) and Transport Layer Security (TLS) at the RPC layer. The third aspect of the holistic security approach is related to security of "data in use" refers to data at-rest state, residing on one particular node of the network (for example, in resident memory, swap, processor cache or disk cache). Although the AEGIS consortium has already identified a list of candidate technologies, such as Homomorphic Encryption and Verifiable Computation, it was decided that the evaluation and adoption of such technologies will be included in the upcoming releases of the AEGIS platform.

The AEGIS holistic security approach also covers the security aspects for the technical interfaces provided by the platform. This includes the interfaces provided by the components of the platform in regards to the authorisation, authentication and access approval mechanisms. AEGIS adopts a token based authentication with JSON Web Token (JWT)[22]. JWT is an open standard (RFC 7519) that defines a compact and self-contained way for securely transmitting information between parties as a JSON object.

The AEGIS project provides an offline tool for anonymisation of data, which can be used by a data provider in own IT infrastructure to anonymise data before uploading it to AEGIS platform.

The data management in AEGIS is organised on the basis of computational projects. As starting activity in working with the AEGIS platform a user has to create computational project. The user becomes the owner of the project. The owner of the project can provide access to it to other users. After uploading the data to the platform and adding them to the own project, only the members of the computational project and the platform administrators can see and access the data. The owner of the project can allow selected users to find and make a copy of the selected by the owner project dataset(s).

AEGIS does not provide any special tools for data aggregation. However, the users can write and execute a data analytics script implementing certain aggregation algorithm.

In the AEGIS demonstrators is not planned to store any personal data in the AEGIS platform. The data providers involved in the demonstrators implementation anonymise the data before uploading it.

In the Smart Automotive and Road Safety demonstrator only technical data (anonymised) will be uploaded to the platform. As GPS data is also uploaded, one might – as a result of a clever combination of dataset – be able to identify the location of a driver (start/end) of a trip. The risk mitigation measure is to cut 1-2 minutes from the time series data.

3.1.1.3.2. Anonymisation and approach of de-anonymisation

The only critical data are GPS/location data which can be merged with other datasets containing identical identifiers like in relation to time and location. The required effort is very high and it

[22] JSON Web Tokens, https://jwt.io/

is unlikely to find and use suitable datasets in order to re-identify a person. As result, De-anonymisation is considered to be very unlikely.

3.1.1.3.3. Privacy by default

By default, all data uploaded to AEGIS by a user are accessible and visible only for this user.

3.1.2. Ethical awareness

3.1.2.1. Administrative requirements

3.1.2.1.1. National legislation – data protection officer

For VIF, the responsible national data protection authority in Austria is Austrian Data Protection Authority. [23].

3.1.2.1.2. Confidentiality and access restriction

The organizational structure of VIF only allows the AEGIS project team to have access to all data stored at VIF, which is a closer group. It exists separate tables containing the link of a *driver_id* to a person, which is separated from the actual raw data store. The access to both is only granted to the AEGIS project team as well.

3.1.2.1.3. Requirements referring to the voluntary participation to AEGIS demonstrators!

VIF ensures that all participants are employed at VIF and voluntarily take part in the Smart Automotive demonstrator case.

3.2. Demonstrator Case 2: Smart Home and Assisted Living Demonstrator

3.2.1. Compliance with Data Protection

3.2.1.1. Data protection principles

3.2.1.1.1. Data minimization – avoidance of personal data if not necessary

Main purpose of the Smart assisting demonstrator is to illustrate and implement a services bundle towards advanced holistic monitoring and assisted living management, aiming to improve everyday living and enhance the wellbeing of people belonging to vulnerable groups.

Data is collected from the following sources: Smartphone and wearable data are exported from the respective devices in supported, for each particular device, formats (mainly xml, using specified schemas), and uploaded to the SHAL backbone server/database through a dedicated mobile app and API that is offered to the at-risk individuals. Only people that have explicitly declared their agreement in supplying their data and in turn receive personalized notifications and warnings will be considered. Some additional clarifications pertaining specific types of data arriving from these devices are discussed; geo-locational data will be processed after acquisition to keep only the minimum necessary information (city, area) and not the exact latitude and longitude, biometrical information (height, weight) which represent sensitive data pursuant to

[23] https://www.data-protection-authority.gv.at/

article 9 GDPR are entered voluntarily and with explicit consent by the user, vital signs (blood pressure, heart rate) are recorded only for pre-specified purposes.

The demonstrator is built around the availability of the minimum set of data required for the provision of the following services: a) Display Notifications on potentially harmful conditions/events, b) inform the end-users of identified risks. In all cases and scenarios, any recorded information leading to direct correlation of the person to the data is kept offline (in the database server of the demonstrator), which can be accessed only by the designated data processor. Data uploaded to Aegis are subject to strict anonymisation processes, as offered by the Aegis anonymiser tool.

To realize the bundle services, the data sources have been defined at the beginning of the project in order to estimate which data is necessary for the development of the demonstrator application and single services. The amount of data collected depends of the granularity of the data acquisition process.

Personal data collected are intended to be subjected to an offline preprocessing workflow which includes the data anonymisation and dissociation with any sensitive information. Additionally personal data for the login and registration process are locally stored on the HYP database and are not uploaded to the AEGIS platform.

Concerning the issues of de-anonymisation resulting from association and linking different datasets, the actions to be taken are the following: No metadata pointing from one dataset to the other will be openly available and stored online. Any combined processing will not result in the saving of the joint dataset and will be performed by an assigned data processor. Insofar no data saved on the AEGIS platform are considered to be de-anonymised.

3.2.1.1.2. Purpose limitation principle

In relation to the aforementioned purpose of the Smart home and assisted living demonstrator, the collecting and acquisition of data depends on the respective service. Data processing within the demonstrator aims at the provision of a monitoring and event notification service to individuals. No extraction of further information than necessary is performed. To that extent, the derived data process workflows are limited to utilization of either publicly-available, simulated or anonymised data. Regarding the last category, any de-anonymisation information is only stored locally and is not uploaded to the Aegis platform.

3.2.1.1.3. Storage retention – Restriction of storage

All data that have been collected and processed for the demonstrator processed are only be kept as long as needed for the demonstrator test cases and will be completely destructed and removed the test case finalisation. This especially includes all personal data which are required for the data process in all stages. The datasets to be deleted are marked with an expiration date and will be deleted by the responsible data processor performing all actions in the data process. Due to the expiration date, the erasure of datasets will take place in relation to the purpose pursuant to article 4 number (3) GDPR. Besides, no automated deletion algorithm was implemented.

3.2.1.1.4. Data accuracy, integrity and confidentiality

The quality of data is pursued through actions in both data acquisition and data processing steps. As data is the main source in order to provide appropriate personalized services, HYP has performed the following steps:

- First careful testing and calibration of the monitoring equipment
- Second, the rigorous establishment of outlier and anomaly detection algorithmic procedures.

This procedure shall ensure that all services are running correctly, granular and precise enough to provide an assisted living management, thereby complying with the accuracy requirement of article 5 I nb. (d) GDPR.

3.2.1.1.5. Fair data process - Use of automated processing

Taking into account the demonstrator purpose, the extracted profiles are used to infer user preferences and suggest possible actions to the user. Services provided intend to pursuit the user´s interest, taking into account personal preferences in the data processing and not beyond. The final decision is up to the individual which results in avoiding discrimination and contrasting interest.

With regards to safeguarding discrimination and stigmatization, the demonstrator has considered along with the double pseudonymisation approach, the generation of personas. More specifically, personas will be created for the "grouping" of individuals with similar profiles (e.g. belonging to a similar age group and having correlatable medical profiles). This approach is not affecting at all the project outcomes, as the objective of the demonstrator is to test the technical implementation (at a research level) of the different services and applications mentioned above. Therefore, non-actual but "fake" pseudonymised individuals (thus no actual sensitive data) will be considered for testing the specific business functionalities of the project.

3.2.1.2. Data processing

3.2.1.2.1. Lawful – legal basis and its conditions

All the test subjects will be informed and given the opportunity to provide their consent to any monitoring and data acquisition process. The pilot tests supervisor will inform the participants with clarity about the procedure of the pilot tests, the system operation and the objectives, the data retrieval and storage and the exact dates the tests will be running. The consent is written in simple terms and defines exact correspondence between each data source and the respective data process and intention, for which it is used. No data will be collected without the explicit informed consent of the individuals under observation and their legal guardian where applicable. This involves being open with participants about what they are involving themselves in and ensuring that they have agreed fully to the procedures/research being undertaken by giving their explicit consent. Before the experiments start, an informed consent procedure will be applied. Additionally, the subjects will be strictly volunteers and all test volunteers receive detailed oral information.

3.2.1.2.2. Compliance of obligations and respect to data subject rights

All the test subjects will be informed and given the opportunity to provide their consent to any monitoring and data acquisition process. The pilot tests supervisor will inform the participants with clarity about the procedure of the pilot tests, the system operation and the objectives, the

data retrieval and storage and the exact dates the tests will be running. Moreover, every participant has the Right to obtain from the pilot controller without constraint at reasonable intervals and without excessive delay or expense:

- confirmation as to whether or not data relating to him are being processed and information at least as to the purposes of the processing, the categories of data concerned, and the recipients or categories of recipients to whom the data are disclosed,
- communication to him in an intelligible form of the data undergoing processing and of any available information as to their source,
- knowledge of the logic involved in any automatic processing of data concerning him.

3.2.1.3. Technical and organization measures

3.2.1.3.1. Technical and organizational safeguards – privacy by design

Only de-identified and anonymised data will be uploaded to the Aegis platform and a cloud storage for the demonstrator. Data pertaining identification and de-anonymisation are stored encrypted in a local server under strict firewall access and are handled only by demonstrator involved personnel. By default, any identification data are removed before any processing and stored locally for only as long as necessary. The anonymisation process is performed through the respective dedicated component, which is installed independently of the AEGIS platform as local standalone software. All signup and login information is protected using standardized encryption protocols. The same is true for any sensitive information that is stored locally. Data process activity logs are kept in a local server under strict access policies.

3.2.1.3.2. Privacy by default

By default, all data uploaded to AEGIS by a user are accessible and visible only for this user.

3.2.1.4. Administrative requirements

3.2.1.4.1. National legislation – data protection officer

Due to 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.

3.2.1.4.2. Confidentiality and access restriction

Within the demonstrator beneficiaries, only identified data processors have access and are allowed to process de-anonymised data.

3.2.1.4.3. Requirements referring to the voluntary participation to AEGIS demonstrators

For the purposes of the demonstrator, data subjects come from within the demonstrator beneficiaries and no discrimination sorting is applicable. Nevertheless all required procedures regarding consent are followed.

3.3. Demonstrator Case 3: Smart Insurance: SameHealthForAll

3.3.1. Compliance with Data Protection

3.3.1.1. Data protection principles

3.3.1.1.1. Data minimization – avoidance of personal data if not necessary

Purpose of the HDI Insurance demonstrator comprises specific areas (e.g. Metropolitan City of Rome). For this, only the customers that allow the use of their data for analysis purpose to receive personalized offers, warnings and support for claims are taken into account. Data of further customers and/or further data will not be uploaded on the platform neither anonymised. As result, the amount of personal data will be kept at a minimum amount. Data processed are the data of the policies held by the customers and geolocation data provided by the HDI Mobile App, restricted to the necessary amount, additionally to the in-house datasets after anonymisation has taken place.

Concerning the problem of De-Anonymisation, It is highly improbable since the in-house datasets are linked with external data of events (weather or social) but not related to people. In case of incidental findings the data will be deleted. Besides, the data will be managed and handled only by the HDI employees that are working on the AEGIS project.

3.3.1.1.2. Purpose limitation principle

The Insurance Demonstrator processes data in order to help its customers preventing or managing a potentially damaging event. For this provided are three scenarios, the description of the data used in each of them follows, while a detailed description of data process is presented in section 2.3.3 of the present deliverable. Presented are the use cases of the scenario including the data necessary to pursue the purpose of the respective use case.

1. Personalized early warning services for asset protection

   The aim of this use case is to provide a warning service to a customer possibly involved in an event, and in the same time to offer a policy related to the event. The data analyzed will be open data (e.g. weather forecasts from websites) and in-house datasets (customer data: policy data - kind of policy, validity of the policy, “location“ of the policy).

2. Financial impact, customer support and services

   The aim of this use case is to evaluate the financial impact for the company of an event already happened, and to contact customers possibly involved, providing them information and (if needed) claim documentation. The data analysed will be open data (e.g. news from websites) and in-house datasets (customer data: policy data, geolocation).

3. Marketing strategy and pricing support services

   The aim of this use case is to build accurate business plans and marketing strategies. The data analyzed will be open data (e.g. Italian official stats websites) and in-house datasets (customer data: policy data both actual and historical). To reach the objective of Scenario 1, Scenario 2 and Scenario 3, the in-house datasets are accurately chosen

and filtered in order to process only the data needed for the evaluation of the AEGIS platform and the demonstrator.

3.3.1.1.3. Storage retention – Restriction of storage

The data used in the project are filtered (e.g. only customers that have a policy in the metropolitan City of Rome), anonymised and then uploaded on the platform (for further information ref. 2.3.3 section of the present deliverable). These data will be available for the analysis of the HDI Data Scientists for the duration of the tests and evaluation within the timing defined in WP5. After the project, all data containing personal identifiers will be deleted from the platform and not further used for any data process.

3.3.1.1.4. Data accuracy, integrity and confidentiality

In order to provide integrity of the datasets, only reliable data sources will be considered in the analysis (e.g. for Scenario 1 and 2 only news provided by trusted sources, for Scenario 3 only official Italian stats, will be considered). Regarding the in-house data, the data management of the parties involved in HDI Assicurazioni takes place through the relationship between the insurance agencies and the final customer. Thereby, the customer can view his/her data through the "customer area" or even interact with our agents. Furthermore, in the latter case, customers can request changes to their data.

Current company and sector data quality policies will be applied to the data stored in the HDI local servers, whilst the de-identified and anonymised data will be accessible on the platform only to authorized users for the time of the analysis and then will be deleted. In case of expired policies, the agents of HDI Assicurazioni verify the correctness of the data with the customers in order to provide data quality and data correctness. In particular, HDI Assicurazioni has technical processes and procedures to prevent data loss or unauthorized access to it. In order to provide data availability at any time, adequate backup / restore and business continuity policies guarantee data loss victims. Access profiles to segregated data by type of processing performed on data are in place. Additionally, HDI will apply the current company and sector data quality policies to the data stored in the HDI local servers. These may include differences in relation to the specific purpose of the processing. Data quality policies are analyzed from time to time to verify the correspondence of processes to existing treatments. Unauthorized access to data will be avoided with the implementation of suitable measures – a detailed list of technical components can be found below in 3.3.1.3.

3.3.1.1.5. Fair data process - Use of automated processing

The Insurance demonstrator does not take into account any personal, discriminative references to race, gender, age, religion, disability or other. As HDI offers personalized services through the use of the AEGIS analytic tools, the customers will be provided with additional support and information about natural or social events of interest. This is expected to improve also their satisfaction, instead of resulting in negative implications from the individual´s perspective. There is no risk of discrimination and stigmatization due to data processing in AEGIS Insurance Demonstrator. Any information is provided by the customers on a voluntary basis, following strict informed consent procedures, including understandable explanations of project's objectives, as well as data processing main features and purposes and individual's rights.

Within the three scenarios, the use of automated decision making is not integrated. In doing so, no profiles of individuals are going to be created. In order to provide a fair data process, taking into account the person´s interest, the main focus of the company is to provide to its customer a fast and efficient service in order to improve among other things customer loyalty. To improve customer loyalty, customer rights and interests are always kept at first

3.3.1.2. Data processing

3.3.1.2.1. Lawful – legal basis and its conditions

HDI in AEGIS Insurance Demonstrator is going to adopt the same level and accuracy of informed consent procedures as it is usual to follow in its daily activities, in full compliance with European legislation – article 7 GDPR - , as well as with best practices and guidelines followed in the insurance sector. This includes, before the starting of the processing in the demonstrator, the provision of specific, complete and understandable information to the individuals and the adoption of safeguards for his/her dignity. We use informed consent form formulated in simple, clear, precise and understandable terms. Any additional information will be provided verbally before the trial takes place, when the individual can also have the chance to ask for questions and clarifications. Different explanations will refer to the different scenarios, in order to be as much concrete and specific as possible.

This information is functional to obtain a truly free given, unambiguous and specific consent, which is expression of the free exercise of choice. HDI also informs of and ensures the possibility of withdrawal without any negative effects for the individual and to avoid coupling the collecting of personal with providing the respective service. In case of secondary used, specific consent will be collected.

3.3.1.2.2. Compliance of obligations and respect to data subject rights

Due to the fact that data is anonymised for the most part, most conditions are set in the consent declaration. Beyond, the data subject is able to access his own data and can request any changes in order to provide the possibility to interfere as well as the correctness (personal) data. Besides, any privacy concerns are handled apart from AEGIS.

3.3.1.3. Technical and organization measures

3.3.1.3.1. Technical and organizational safeguards – privacy by design

The in-house datasets are stored in the HDI servers. Only the data necessary for the analysis are anonymised, filtered and uploaded on the platform by authorized users. These data are visible only to the HDI Data Scientists.

The main risks may be partially different in respect of the different scenarios of the Insurance Demonstrator. In general there is the risk of loss of confidentiality and re-identification. However, this is very unlikely, considering both the technical and organizational safeguards adopted, as described in D5.1and D5.2, and the use of a private cloud.

As for the customer’s position provided by HDI to the AEGIS platform and which may be gained from the HDI Mobile App, the individual will provide specific consent for the use of them.

The data are anonymised before the upload on the platform, only authorized users can access them. The HDI databases implement the policies (in agreement with the Italian Law) for data protection and handling. HDI use an internal monitoring system for assessing any solutions' compliance with the set of pre-identified requirements of data processing.

The anonymisation process is performed by the Anonymiser, a component adopted by the AEGIS project, which is installed as local standalone software. Data will be deanonymised only after the generation of the report, within the HDI processes through an offline tool.

HDI Assicurazioni adopts the following technical measures to prevent loss and unauthorized access to data. Below is the list of existing measures:

- Perimeter security management with Fortinet advanced firewall;
- SOC (SOC II level) for all services;
- Virtualization of users' workstations;
- Virtualized access to applications (two levels of security);
- System of backup recovery of all data;
- Business Continuity.

Further initiatives to strengthen security and data protection policies are being evaluated:

- Introduction of a cybersecurity competence center (SOC III level);
- Introduction of a system for the encryption of production databases;
- Introduction of an SIEM;
- Adoption of Identity Management solutions to centrally manage and monitor the life cycle of users and their respective authorizations;

Extension of the SSO solution for centralized access control and password policy rules.

3.3.1.3.2. Privacy by default

By default, all data uploaded to AEGIS by a user are accessible and visible only for this user.

3.3.1.4. Administrative requirements

3.3.1.4.1. National legislation – data protection officer

Considering the features of the Insurance Demonstrator and the already existing notifications/authorizations by the National Data Protection Authority in Italy (Garante della Privacy), HDI is going to contact it to eventually extend/modify such notifications/authorizations in relation to HDI's activities within the AEGIS project.

3.3.1.4.2. Confidentiality and access restriction

Being data processing a daily activity in HDI, its internal organization complies with all the relevant provisions regulating the entities involved in data handling (e.g. data controller, data processors, etc.). They are still applicable also under GDPR regime. The AEGIS platform is accessible only by a restrict number of users, that signed a specific contract about data confidentiality, in agreement with the Italian legislation. On the other hand, it is ongoing the process to integrate this consolidated architecture with the figure of the Data Protection Officer in the cases outlined in Article 37 of GDPR.

3.3.1.4.3. Requirements referring to the voluntary participation to AEGIS demonstrators

The Mobile App will be proposed to the customers of a restricted area as aforementioned in order to minimize the data collected, only volunteers will install and use it. The in-house datasets of HDI are feed while buying a policy, and in the same time the customer signs voluntarily if allow the use of his/her data for analysis purpose to receive personalized offers, warnings and support for claims.

4. CONCLUSION

As summery of this deliverable, the compliance of data protection as well as ethical issues concerning privacy are fully considered. The current progress of each demonstrator including the planning as well as the first implementation phase involve all relevant issues to ensure data protection and privacy. The continuous progress is observed by the EAB – Ethical Advisory Board – whereby reporting privacy concerns and issues as well as compliance with data protection shall be reported during the implementation phase. With regard to the current project phase, the following statements shall summarize the assessment of the demonstrators:

1. Not every demonstrator collects personal data – see Smart Automotive in 3.1. Most demonstrators only require technical data which do not contain any personal identifiers
2. If personal data are collected, the amount is restricted to minimum necessary to provide the respective, personalized service
3. The demonstrator´s purposes including their respective services are specified in a sufficient manner to provide the correct handling of data in order to only achieve the desired result of the data process. This focus means the exclusion of arbitrariness.
4. In addition to purpose limitation, most participants, who voluntarily participate in the respective demonstrator, are able to monitor and access their personal as well as non-personal data for the purpose of correctness and data quality. Therefore, the demonstrators not only prove compliance, but the participants themselves can access and verify the correctness of the data process.
5. All demonstrators are using the AEGIS offline anonymisation or similar anonymisation procedures. Even if personal data process are collected, the analysis in AEGIS is always performed without reference to an individual – the interference is considered as at least minimal.
6. The security measures – beyond AEGIS itself – protect the service from external attacks and unauthorized access. Internal management access control guarantees that access is internally organized which means that only assigned employees in the respective project have access.
7. Research work in this project is only focusing on relevant analysis procedures. Besides, most data are technical data, all researchers are aware of their ethical responsibility.