HORIZON 2020 - ICT-14-2016-1

# AEGIS

Advanced Big Data Value Chains for Public Safety and Personal Security

## WP3 – System Requirements, User stories, Architecture and MicroServices

# D3.1 – Technical and User Requirements and Architecture v1.00

Version 1.0

**Due date:** 30.06.2017 **Delivery Date**: 30.06.2017

**Author(s):** Dimitrios Miltiadou, Konstantinos Perakis (UBITECH), Yury Glikman, Andreas Schramm (Fraunhofer), Elisa Rossi, Cinzia Rubattino, Giulio Piemontese, Andrea Martini, Michele Caira (GFT), Spyros Mouzakitis, Evmorfia Biliri (NTUA), Alexandru - Adrian Ormenisan, Mahmoud Ismail (KTH), Sotiris Kousouris (SUITE5), Alessandro Testa(HDI)

**Editor**: Dimitrios Miltiadou (UBITECH)
**Lead Beneficiary of Deliverable**: UBITECH

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

**Internal Reviewers:** Yury Glikman, Andreas Schramm (Fraunhofer), Elisa Rossi, Cinzia Rubattino, Maurizio Megliola (GFT), Spyros Mouzakitis, Evmorfia Biliri (NTUA)

## EXPLANATIONS FOR FRONTPAGE

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

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

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

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

**AEGIS KEY FACTS**

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

**AEGIS PARTNERS**

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

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

## EXECUTIVE SUMMARY

The scope of D3.1 is to document the preliminary efforts undertaken within the context of Tasks 3.1, 3.2 and 3.3. Towards this end, the scope of the current deliverable is to collect and analyse the functional and non-functional requirements of the platform, following the requirements extraction techniques of agile software development, adopting the principles of user stories and user cards. In addition to the functional and non-functional requirements, the scope is to identify and document the preliminary version of the technical requirements of the platform as well as of the components to be integrated in the holistic platform. In addition to the aforementioned, the deliverable aims at designing and documenting the preliminary version of the conceptual architecture of the integrated platform, including the necessary components which will address all the technical requirements identified. The architecture has been designed in a modular way facilitating easy maintenance, modifiability and extensibility. The scope of the architecture is to reach the goals and expectations of the users and the stakeholders. In this context, the key components of the AEGIS platform and their specific functionalities and interactions have been described. The complete architecture was compiled by the list of components making sure that every requirement was addressed towards the required AEGIS complete required functionality.

In the next steps the platform architecture compiled in this deliverable along with the list of components specified will be utilized in order to design the components, the different micro-services that will drive the AEGIS execution flow along with the appropriate API interfaces that will be implemented. The current deliverable comprises the preliminary version of the technical and user requirements and also delivers the conceptual architecture of the AEGIS platform. The forthcoming versions of this deliverable comprise updated iterations and will include updates on the platform and components' architecture, as well as the APIs interfaces, based on the feedback received by the project's demonstrators. Thus, the future versions of this deliverable will include the more detailed technical architecture of the platform, updating the requirements backlogs provided in the annexes, as well as additional information including for example where all data sources are located, where the processing power resides, what analytics are supported and what services are delivered to the users etc. in full detail.

# Table of Contents

## LIST OF FIGURES



## LIST OF TABLES

## ABBREVIATIONS

| | |
|---|---|
| AEGIS | Advanced Big Data Value Chains for Public Safety and Personal Security |
| API | Application Programming Interface |
| CPU | Central Processing Unit |
| CSV | Comma-separate values |
| D | Deliverable |
| DoA | Description of Actions |
| GPS | Global Positioning System |
| GUI | Graphical User Interface |
| HVAC | Heating, Ventilation and Air Conditioning |
| H2020 | Horizon 2020 Programme |
| IAQ | Indoor Air Quality |
| IEEE | Institute of Electrical and Electronics Engineers |
| I/O | Input / Output |
| IOPS | Input Output operations per second |
| IoT | Internet of Things |
| IPR | Intellectual Property Rights |
| IPsec | Internet Protocol Security |
| JSON | JavaScript Object Notation |
| MVP | Minimum Viable Product |
| NLP | Natural Language Processing |
| PSPS | Public Safety and Personal Security |
| RDF | Resource Description Framework |
| REST | Representational State Transfer |
| RSS | Rich Site Summary |
| SLOD | Security Linked Open Data |
| T | Task |

| | |
|---|---|
| TLS | Transfer Layer Security |
| URI | Uniform Resource Identifier |
| URL | Uniform Resource Locator |
| WP | Work Package |
| XML | Extensible Markup Language |

# 1. INTRODUCTION

The scope of the current section is to introduce the deliverable and familiarize the user with its contents. Towards this end, the current section summarises the objective of the current deliverable, its relation to the other work packages and corresponding deliverables, and analyses its structure.

## 1.1. OBJECTIVE OF THE DELIVERABLE

The scope of D3.1 is to document the preliminary efforts undertaken within the context of Tasks 3.1, 3.2 and 3.3. Towards this end, the scope of the current deliverable is to collect and analyse the functional and non-functional requirements of the platform, following the requirements extraction techniques of agile software development, adopting the principles of user stories and user cards. The user stories (and subsequently the elicited requirements) will stem directly from the demonstrator partners and from the overall analysis of all stakeholders involved, leading to the collection and analysis of both core as well as demonstrator specific requirements. The user stories, containing both functional and non-functional requirements, will include a complete high-level description of the expected behaviour of all sub-systems that are going to be specified and developed and will build the basis for the implementation of the platform's MVP. In addition to the functional and non-functional requirements, the scope is to identify and document the preliminary version of the technical requirements of the platform as well as of the components to be integrated in the holistic platform. The identification and analysis of the functional and non-functional requirements, and their translation into technical requirements is a living process that will last until M25, and will constantly be updated and documented in the subsequent versions of this deliverable. The updated technical requirements will also feed the platform architecture which will be customized, modified and extended accordingly in order to meet these requirements. In addition to the aforementioned, the deliverable aims at designing and documenting the preliminary version of the conceptual architecture of the integrated platform, including the necessary components which will address all the technical requirements identified. The architecture will be designed in a modular way facilitating easy maintenance, modifiability and extensibility. The platform will be designed with the scope to facilitate interoperability and to allow for extensibility while preserving a considerable degree of implementation-platform-independence. Last but not least, the deliverable will maintain the requirements backlog (split into the functional, the non-functional, and the technical requirements backlogs) regarding the components under integration during the project implementation in order to guide all development tasks.

## 1.2. INSIGHTS FROM OTHER TASKS AND DELIVERABLES

Work package 3 receives as input mainly the early reports of WP1 and WP2. Towards this end, the elicitation of the functional and non-functional requirements of the platform, their translation into technical requirements, and the design of the preliminary version of the platform architecture, constituting the conceptual rather than the purely technical architecture, receive as input mainly the results of deliverable D1.2, documenting the high-level usage scenarios, serving as the basis upon which the user stories were built. The reported user stories and usage scenarios will complement (and will be complemented by) the outputs of Task 1.5.

## 1.3. STRUCTURE

Deliverable 3.1 is organized in five main sections as indicated in the table of contents.

- The first section introduces the deliverable. It documents the scope of the deliverable and briefly describes how the document is structured. It also documents the relation of the current deliverable with the other deliverables, and how the knowledge produced in the other deliverables and work-packages served as input to the current deliverable.
- Following the introductory section, section 2 describes the methodology that was followed for the identification of the requirements. The deliverable documents the various steps followed for the engineering of the requirements in the context of the project, and how the user stories collected were translated into functional and non-functional requirements, both generic, platform related, as well as demonstrator specific ones. Section 2 also documents the actors identified and their interaction with the AEGIS platform, as opposed to the Value Chain designed and documented in D1.1.
- Following the analytical documentation of the functional and non-functional requirements, section 3 translates these requirements into the technical requirements that will actually drive the design and implementation of the AEGIS platform. The technical requirements are mapped to the corresponding functional and/or non-functional requirements, and are also prioritised (High – Medium – Low) so that different technical requirements are progressively addressed as the project and the platform mature.
- Section 4 delivers the design of the first version of the AEGIS framework architecture. It delivers a schematic providing an overview of the architecture, broken down into a series of components each one undertaking a specific role in the platform operation, and then documents the role of each of those components. The technical components are also mapped to the technical requirements to safeguard that each requirement is addressed by a specific component (or set of components) in the architecture.
- Section 5 concludes the deliverable. It outlines the main findings of the deliverable which will guide the future research and technological efforts of the consortium.

Deliverable D3.1 includes also a series of Annexes, which include the series of requirements backlogs, and more specifically:

- A stakeholders and literature requirements backlog, namely Appendix A, which summarises the more generic requirements collected and analysed within the context of WP1.
- A Functional requirements backlog, namely Appendix B, which will comprise a living backlog and will be populated with additional functional requirements and edited to remove requirements which may be considered obsolete as the project progresses and evolves.
- A Non-Functional requirements backlog, namely Appendix C, which will also comprise a living backlog and will also be populated with additional non-functional requirements and edited to remove requirements which may be considered obsolete as the project progresses and evolves.
- A Technical requirements backlog, namely Appendix D, which will comprise a living backlog as well, and will also be populated with additional technical requirements and edited to remove requirements which may be considered obsolete as the project progresses and evolves.

- A list of the User Stories, namely Appendix E, which will show the user stories collected from each demonstrator.

# 2. METHODOLOGY OF REQUIREMENTS IDENTIFICATION

In this section, we describe both of the 'standard' Agile methodology and the Agile methodologies adapted to the AEGIS needs (paragraph 2.1) reporting the templates used for user stories and requirements collection. The actors of the AEGIS solution are depicted in paragraphs 2.2 and 2.3. In paragraph 2.4 we show the results of the whole Agile methodology in AEGIS. In this particular context, the techniques of Agile software development lead to finalise the preliminary version of the user requirements through the collection and processing of the user stories (paragraph 2.4).

The overall objective of this review is to identify the user requirements of the AEGIS demonstrators in order to develop the best-suited solution for the pilots. The AEGIS user requirements have been split in 'core' and 'demonstrator' requirements. The 'core' requirements are those essential for every demonstrator, while the 'demonstrator' ones are requirements of each specific demonstrator.

Within this assessment, the purpose is to collect, analyse and synthesize the requirements of the pilot partners, following the requirements extraction techniques of Agile software development, adopting the principles of user stories and user cards.

The work that had been done in WP1, hence the high level requirements identified in D1.1, is the basis of initiating this task. By the collaboration between technical partners and pilots both functional and non-functional requirements were pointed out from the user stories.

The idea is to consolidate both AEGIS core user requirements and demonstrator user requirements in a concrete and logical list of user requirements which will be also used in section 3 to extract more technical requirements. In this section the preliminary user requirements, already delivered from Task 1.2 in D1.1, were made more solid and more descriptive.

The preliminary user requirements evidenced in D1.1, referred to all the possible AEGIS stakeholders, are shown in Appendix (A). The main part of these requirements, reminding that they come both from literature and from the questionnaire, is in agreement with the Demonstrator User Requirements and the Core User Requirements that came from the extraction techniques of Agile software development.

These first results will build the basis for the implementation of the platform's MVP.

## 2.1. DESCRIPTION OF THE METHODOLOGY

The methodology that seemed to be more suitable with the AEGIS purposes and needs is the Agile development one since it provides opportunities to assess the direction of a project throughout the development lifecycle. Moreover, it is people-focused and communications-oriented under the light that requirements cannot be fully collected at the beginning of the software development cycle; every aspect of development — requirements, design, etc. — is continually revisited to have time to steer it in another direction. The Agile planning is adaptive, focused on quick responses to change through continuous development and improvement. In the figure below (Figure 2.1-1) it is shown the general Agile methodology development process.

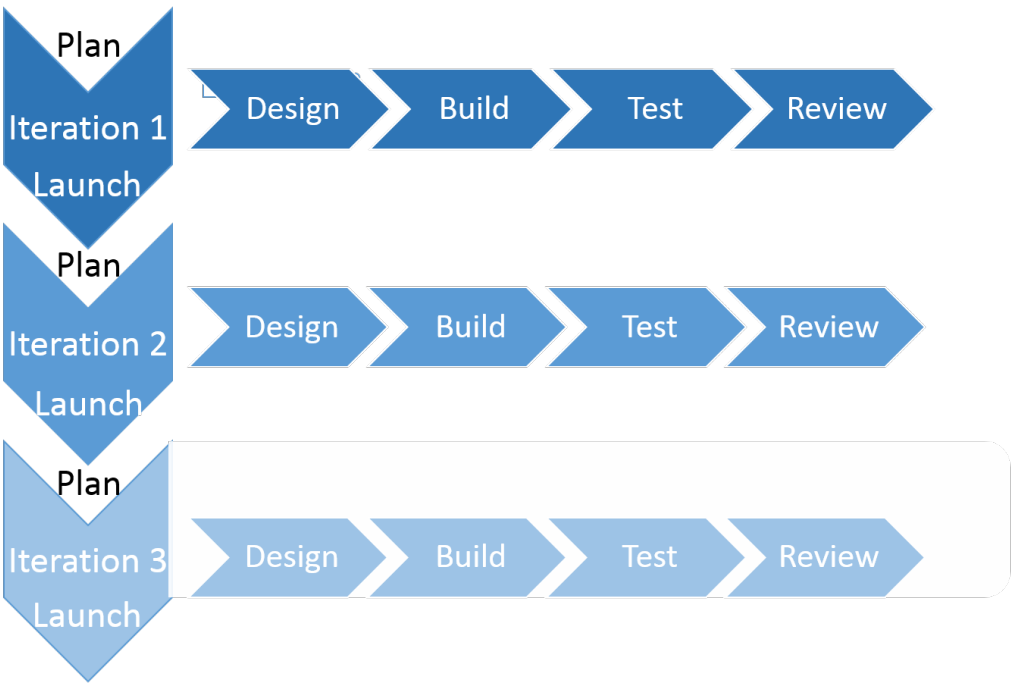



**Figure 2-1: Agile methodology**

As seen in Figure 2.1-1, the whole product in Agile methodology is broken into small increments that minimize overall risk and allow the product to adapt to changes quickly. Iterations are short time frames (called timeboxes), each involving a cross-functional team working in all functions: planning, analysis, design, coding, unit testing, and acceptance testing. The requirements/tasks to be done in the timebox are defined at the beginning of iteration and agreed upon by the team.

As mentioned previously another key concept in Agile development is the close collaboration between the development team and the business stakeholders. In accordance with that, the requirements elicitation was achieved adopting the principles of user stories, coming directly from the demonstrator partners.

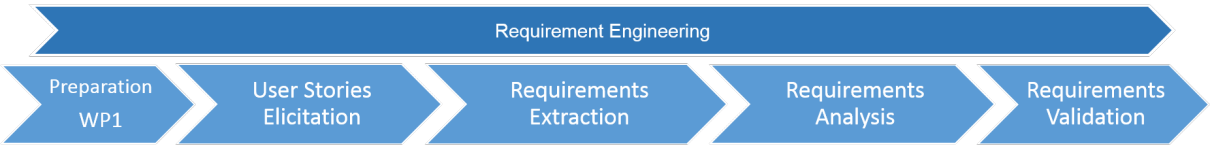



**Figure 2-2: Requirement Engineering in AEGIS**

The picture above represents the phases for the definition of the AEGIS user requirements. During the activities of WP1 the business needs of AEGIS stakeholders have been identified and this has taken to the definition of the first high level requirements that have been then revised. They define the high-level goals, objectives and needs of the organization. Afterwards the User Stories elicitation has been performed. At this stage workshops and interviews with

the pilot team have been performed by the technical partners in order to obtain the relevant user stories that AEGIS solution should implement.

The user stories have been collected and analysed by the technical partners to extracts the user requirements. The requirements collected using the structure defined in the following sections have been analysed for consistency and finally a specific workshop has been held to validate them and obtain the final list reported in paragraph 2.3.

### *2.1.1. User Stories and Cards elicitation*

User stories are a very high-level definition of requirements, describing a feature told from the end user's perspective (i.e. who desires the new capability), usually a user or customer of the system.

A user story is short, generally one-sentence, but it contains enough information to describe the requirement, so that the developers can produce a reasonable estimation of the effort to implement it.

A user story typically follows a simple template:

*As a < user-type>, I want to <user-requirement> so that <reason>.*

User stories are written throughout the Agile project. Usually a story-writing workshop is held close to the start of the Agile project. Additionally, new stories can be written and added every iteration. The template for the collection of them is shown in Table 2.1.

**Table 2-1: AEGIS User Stories Template with some examples of user stories**

| Id | Epic | User story | | | Priority | Value | Acceptance |
|---|---|---|---|---|---|---|---|
| | | As a <user type> | I want to <user requirement> | So that <reason> | | | |
| HYP1 | Connection/Integration | Data Consumer (Developer that want to use API) | get indoors environmental data from installed physical devices (sensors) | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can receive and process indoor environmental data (Humidity, Temperature and Luminance) from sensors and conduct analysis or correlation with other data |
| VIF1 | Browse services | Data Provider | select broken road indicator service | use broken road indicator service | 2 Med | 2 Med | Users should be able to select and use the broken road indicator service. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HDI1 | Services/Micro-services | Data Consumer (Data Analyst) | mark each analysis result in a structured folder (organized tree) | facilitate further traceability | 2 Med | 2 Med | AEGIS needs to implement a system for information traceability queries |

The User Stories Excel sheet is structured as follows:

- ID: an arbitrary ID composed by the pilot partner name and a number;
- EPIC: it can be defined as a large undefined topic that needs to be broken down in smaller stories to find out the requirements. Following the results of WP1, some epics were identified from the Scenarios (e.g. Analytic Tools, Browse services, Connection/Integration, Dashboards/Report/Notifications, Data Collection, Data Management, Data Protection, Data Storage, Data Upload, Linked Data, Multilingual Tools, Open Data, Personal/Public Safety, Query Tools, Semantically Enrichment, Services/Micro-services, Sharing/Visibility), but each technical partner could add epics if necessary. Since Epics were a way to think about the different functionalities, they were very useful for the user stories elicitation;
- USER STORY
    - *As a <role>*: through a dropdown list, it is the role of the subject of the story; the roles are those identified in WP1 and described in Paragraph 2.2;
    - *I want to <do a thing>*: the functionality to be added to the AEGIS solution; this field is free;
    - *So that I can <reason>*: the added value of the development suggested; this field is free;
- PRIORITY: it is a priority value (high, medium or low) given from who defined the user story;
- VALUE: it is the added value (high, medium or low) to the AEGIS solution from the defined story (from the pilot-user point of view);
- ACCEPTANCE: the conditions of satisfaction of the story.

The user stories were pointed out through workshops between a technical partner and a pilot; for each pilot a technical partner was assigned. All the user stories collected are presented in Appendix (E).

### *2.1.2. User Requirements Definition*

The user stories are useful to extract the user requirements. Once the user stories have been collected, the technical partners retrieved the requirements from them. The template provided for the requirements definition is shown in **Table 2-2**. During this step, the requirements were classified as ‘Core’ or ‘Demonstrator’ and ‘Functional’ or ‘Non-Functional’.

**Table 2-2: AEGIS Requirement Definition Template**

| Id | Requirement type | Functional or not | Requirement description | REFERENCES |
|---|---|---|---|---|
| UBI1 | Core | Functional | AEGIS should be able to process sensor data, including environmental (indoor and outdoor), occupancy sensing and Air Quality monitoring, from installed physical devices. | HYP1, HYP2, HYP3 |
| KTH1 | Demonstrator | Functional | Users should be able to choose a service to work on their data | VIF1,VIF13,VIF26 |
| GFT2 | Core | Functional | AEGIS needs to have a data/knowledge base to handle traceability and submitted issues status | GFT1 |

The Requirements Excel sheet is structured as follows:

- ID: an arbitrary ID composed by the technical partner name and a number;
- REQUIREMENT TYPE: each technical partner indicates if the requirement is a Core or a Demonstrator requirement;
- FUNCTIONAL OR NOT: each technical partner indicates if the requirement is a functional or a non-functional requirement;
- REQUIREMENT DESCRIPTION: the description of the requirement;
- REFERENCES: the user stories ID from which the requirement comes from.

A technical workshop was organized at the end of the requirements definition in order to refine the whole requirements, validate them and clarify the expected functionalities of AEGIS. The final requirements were identified in respect of several rules, summarized in **Table 2-3**.

**Table 2-3: AEGIS Requirement General Characteristics**

| Characteristic | Meaning for the requirements |
|---|---|
| Unambiguity | The requirements cannot be interpreted in different ways. There is a clear understanding and open issues have been indicated. |
| Understandability | The requirements are understandably described for all the possible stakeholders. Notations and models are adapted to the target group. |

| | |
|---|---|
| Completeness | The requirements describe the functionality completely. |
| Consistency | The requirements must be consistent according to the content, the degree of abstraction and the description. |
| Verifiability | The requirements are weighted by their importance. The requirements are testable and can be measured unambiguously |
| Traceability | The requirements are clearly identified and include information on status, author, version, etc. They are linked at least with one test case and if it is necessary, they are linked horizontally and vertically to other requirements of comparable abstractions. |
| Relevancy | The requirements are necessary to implement the properties, which have at least a concrete benefit for one target group. |
| Feasibility | The requirements can be implemented within the resources and the technology to be used. The complexity of the implementation is estimated and arranged. |

## 2.2. IDENTIFIED ACTORS

The actors of the AEGIS solution were identified in agreement with the consortium and are deeply described in Table 2-4.

**Table 2-4: AEGIS actors**

| Actor | Description |
|---|---|
| Administrator | The Administrator is both the Tools Administrator and the System Administrator. He is responsible for: 1) setting up the proper configuration of the AEGIS tools to be easily adopted by the end users; 2) setting up the AEGIS solution and the monitoring of the proper functionality of the overall execution environment. He is the supervisor of the AEGIS Core. |
| Service provider | The service provider is the mid-actor between the stakeholders and the AEGIS tools. He promotes the long-term interests of end users or services supplied, the efficiency and competitiveness. |

| | |
|---|---|
| Data provider | The data provider is the provider of datasets that may be available on a specific format or even raw data. He is responsible for: 1) data source identification; 2) type of data provided; 3) objective of the data; and 4) data format. |
| Data curator | The data curator is responsible of the processes and activities related to the organization and integration of data collected, annotation, publication, presentation and maintenance of the data. The data curator is responsible for: status / maintenance; 2) improvement of data accessibility and quality; and 3) data reliability, reusability. |
| Data Analyst | The data analyst is the person that takes advantage of the available data for the realization of an analysis based on his business needs. |
| Developer | The developer is: 1) the developer of components of the AEGIS tools, responsible for the development of the desired functionalities as well as the proper interfaces; 2) the Demonstrator developer of the API/tools that interact with the AEGIS Core Services |
| Non-expert user | The non-expert user is the end-user that visualize the results of an analysis. |

## 2.3. ACTOR INTERACTION

The interaction between the identified actors of AEGIS is described in the Figure below.

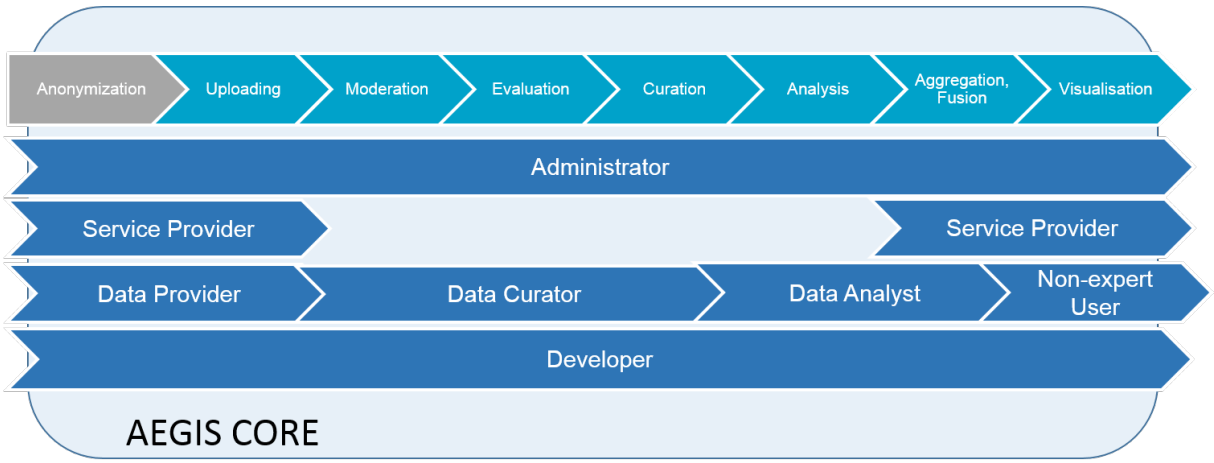



**Figure 2-3: AEGIS actor interaction**

The first row of the Figure concerns the high-level Big Data Value Chain adopted by AEGIS and pointed out in WP1. Each step of the chain is related with the actor(s) involved in the step.

## 2.4. USER REQUIREMENTS

As aforementioned in 2.1.2 the user requirements extracted by user stories have been analysed and validated during a technical workshop where the partners have discussed about each requirement on the view of the whole AEGIS project and of the needs of each pilot.

The requirements were split following two main features:

- Functional or non-functional requirement;
- Core or Demonstrator requirement.

**Table 2-5** reports an overview of the main features of each characteristic.

**Table 2-5: Requirement characteristics: an overview**

| Characteristic | Description |
|---|---|
| Functional Requirement | Functional requirements are one of the most well-known types of requirements, defining the required behaviour of the system to be built, as reported by a hypothetical observer envisioning the inputs that the future system will accept and the outputs it will produce in response to those inputs, e.g., they define the capabilities that a product must provide to its users. Functional requirements are based on system objectives and respond to the critical task of ensuring the right implementation of the expected functionality in the final software. One of the main tools to achieve this goal is system testing, i.e., a mechanism to verify that the system performs the behaviour expressed in its requirements. |
| Non-Functional Requirement | Non-functional requirements specify additional properties of AEGIS, other than functionality. These requirements can be subcategorized into categories such as performance, design constraints (that can also be categorized under external interface requirements), logical database requirements, and “characteristics” (termed “attributes” in IEEE Std. 830) that don’t fit neatly into any of the other categories. The Non-Functional Requirement can also describe quality attributes, design and implementation constraints that the product must have, thus they are more qualitative and may require a different approach for their elicitation. |
| Core Requirement | All processing tasks refer to Core Requirements, since they are requirements that AEGIS should offer. Such processes will be available for each pilot/stakeholder. |

| | |
|---|---|
| Demonstrator Requirement | Demonstrator requirements are the requirements of each specific demonstrator. Demonstrator requirements refer to actions performed by the users or to processes supported by the applications to be developed on top of AEGIS. |

The idea is to consolidate both AEGIS core user requirements and demonstrator user requirements in a concrete list of user requirements which will be also used in section 3 to extract more technical requirements.

*2.4.1. Functional Requirements*

The Functional requirements validated were firstly split into Core User Requirements (paragraph 2.4.1.1) and Demonstrator User Requirements (paragraph 2.4.1.2), then grouped following a similarity criteria.

The identified criteria which are based on the type of macro-functionality the requirement relate to are:

- Analytic: they were categorized as analytic the requirements that refer to processing, uploading, extraction and implementation activities;
- Correlation: processing of different data sources in search of generated insights, including interlinking of datasets;
- Dashboard/Report/Alert;
- Input/Output: I/O requirements are those referred to import and export desired capabilities;
- Privacy: requirements concerning the data policy and anonymization;
- Security: requirements that describe the need to control access to the data. This includes controlling who may view and alter application data;
- Specific Services;
- Rules/Management: requirements including query rules and AEGIS management;

Other: were grouped as ‘Other’ the requirements that were not included in the aforementioned categories.

2.4.1.1. Core User Requirements

**Table 2-6** shows the collection of the Core User Requirements.

**Table 2-6: Core Functional Requirements**

| Id | Description (Detailed description of the requirement) | References |
|---|---|---|
| Analytic | | |
| CFR17 | AEGIS has to implement an advanced clustering algorithm | GFT26 (HDI26) |
| CFR18 | AEGIS provides advanced analytics to compare different results | GFT28(HDI28, HDI29, HDI43, HDI55, HDI56) |
| CFR22 | AEGIS should be able to process daily routines as self-reported from users or automatically extracted by wearables | UBI6 (HYP9) |
| CFR23 | AEGIS should be able to process external weather data | UBI7 (HYP10) |
| CFR24 | AEGIS should be able to process Public Health Information data and statistics | UBI8 (HYP11) |
| CFR25 | AEGIS should be able to process energy (electricity) retailer prices data | UBI9 (HYP12) |
| CFR26 | AEGIS should be able to process CO2 emissions Footprint data | UBI10 (HYP13) |
| CFR27 | AEGIS should be able to process crime and accident raw data | UBI11 (HYP14) |
| CFR39 | AEGIS should allow uploading of users data to be processed for example trips data | KTH2 (VIF2) |
| CFR4 | AEGIS has to support many analysis types (e.g. estimation of correlations between variables, linear regression, predictive analysis, clustering algorithms, simulations) | GFT4 (HDI3,HDI3, HDI6) |
| CFR41 | AEGIS should be able to process sensor data (including environmental (temperature/humidity/luminance), occupancy sensing and Air Quality monitoring) from installed physical devices | UBI1 (HYP1, HYP2, HYP3) |
| CFR42 | AEGIS should be able to process data (energy consumption and set points) from in-house physically installed devices through interfaces | UBI2 (HYP4, HYP5) |

| | | |
|---|---|---|
| CFR43 | AEGIS should be able to process positioning information and location data from mobile phones and wearables | UBI3 (HYP6) |
| CFR44 | AEGIS should be able to process health information and activity data from wearable devices | UBI5 (HYP8) |
| CFR45 | AEGIS should be able to process real-time data on PSPS related events from social media and RSS channels | UBI12 (HYP15, HYP16, HYP17) |
| CFR46 | AEGIS should be able to get weather data and correlate it with users' imported data | KTH10 (VIF15, VIF16) |
| CFR5 | The user should be able to perform the same kind of analysis in different ways (e.g. in critical situation analysis through a simplified algorithm) | GFT5 (HDI4, HDI5) |
| Correlation | | |
| CFR10 | AEGIS has to correlate datasets to geospatial data with their description | GFT12 (HDI14, HDI23) |
| CFR11 | AEGIS should work simultaneously with public and private (customers) data | GFT17 (HDI18) |
| CFR12 | AEGIS has to correlate customer events, habits and policy data | GFT19 (HDI19) |
| CFR15 | AEGIS has to support interlinking of datasets from different sources, also correlating unstructured data sources with internal/IoT data | GFT24 (HDI22, HDI34) |
| CFR28 | AEGIS should be able to correlate positional information (and additional information from wearables) with Public Health Information data and announcements, taking into account also time | UBI16 (HYP22, HYP23) |
| CFR29 | AEGIS should be able to correlate selected historical data from specific sources (indoor/outdoor environmental weather data) with HVAC and lighting device data towards the extraction of user behavioural patterns that interpret the behaviour of building occupants at specific devices under different contextual conditions | UBI17 (HYP24) |
| CFR3 | AEGIS should be able to cross-check data (structured data) | GFT3 (HDI2, HDI19, HDI25) |

| | | |
|---|---|---|
| CFR30 | AEGIS should be able to correlate streams of HVAC, lighting devices data, indoor environmental data and operational patterns towards the implementation of control strategies on Smart Home Environment Devices under specific building contextual conditions | UBI19 (HYP25) |
| CFR31 | AEGIS should be able to correlate Indoor Air quality monitoring data with data for the usage of HVAC devices towards facilitating the implementation of control strategies on HVAC devices by taking into account information about IAQ conditions | UBI21 (HYP26) |
| CFR32 | AEGIS should be able to correlate energy consumption data for HVAC and lighting devices with Energy (electricity) Retailer Prices towards the implementation of control strategies on specific devices by taking into account retailer prices | UBI22 (HYP27) |
| CFR33 | AEGIS should be able to correlate crime related data with HVAC devices usage data towards the implementation of control strategies on devices under specific critical conditions | UBI23 (HYP28) |
| CFR34 | AEGIS should be able to correlate occupancy data and with usage of HVAC and lighting devices data towards the implementation of an occupancy based automation framework in premises | UBI24 (HYP29) |
| CFR35 | AEGIS should be able to correlate public data (weather, public safety, etc..) with personal health vulnerability data, Smartphone/Wearables data, taking into account locality and time | UBI26 (HYP31) |
| CFR36 | AEGIS should be able to correlate real-time data that come either from third party sources or from individuals, based on their location or on the time of their occurrence, with risk patterns | UBI28 (HYP32) |
| CFR37 | AEGIS should be able to correlate personal safety profiles with smartphone data, based on location, activity type and time of execution | UBI29 (HYP33) |
| CFR38 | AEGIS should be able to correlate personal safety profiles with real time data from various sources (smartphones/wearables/sensors), based on their location and time of occurrence | UBI30 (HYP34) |

| | | |
|---|---|---|
| CFR47 | AEGIS should be able to correlate environmental data with Public Health Information data, based on their location and time | UBI15 (HYP21) |
| CFR49 | AEGIS should be able to correlate indoor environmental with outdoor environmental conditions towards the extraction of patterns that affect the operation over specific devices (how the external environmental conditions affect internal environmental conditions) | UBI37 |
| CFR50 | AEGIS should be able to correlate indoors IAQ with outdoors IAQ towards the extraction of patterns that will further facilitate the implementation of control strategies | UBI38 |
| CFR51 | AEGIS should be able to proceed with correlation analysis over indoor and outdoor environmental conditions towards the extraction of seasonal patterns - time based correlation | UBI39 |
| CFR52 | AEGIS should be able to proceed with correlation analysis over indoor and outdoor environmental conditions towards the extraction of seasonal patterns - time based correlation | UBI40 |
| CFR53 | AEGIS should be able to proceed with correlation analysis over retailer prices and energy consumption towards the extraction of patterns of energy usage affected by external factors | UBI41 |
| CFR54 | AEGIS should be able to proceed with correlation analysis over retailer prices towards the extraction of seasonal patterns - time based correlation | UBI42 |
| Input / Output | | |
| CFR16 | AEGIS shall provide capability to manage data sources (import, select, upload and download) from/to different formats (e.g. Excel, download in local storage, USB pen) | GFT25 (HDI24) |
| CFR40 | Users should be able to share and export their analysed data into different systems | KTH7 (VIF11) |
| Privacy | | |
| CFR21 | AEGIS should ask for user's approval to use their data | KTH13 (VIF25) |
| CFR48 | AEGIS should be able to anonymise personal data | UBI32 (HYP38) |
| CFR6 | AEGIS has to implement security mechanisms as well as proper handling of privacy issues (e.g. in case of using private datasets) | GFT6 (HDI7, HDI11, HDI16, HDI52, HDI66, HDI71) |
| Security | | |

| | | |
|---|---|---|
| CFR20 | AEGIS should allow moderators to review and approve uploaded user's code | KTH17 |
| CFR7 | AEGIS needs to display different levels of information depending on who is accessing to the data | GFT7 (HDI8) |
| CFR8 | AEGIS should allow the creation of the different users/groups and access rights for authorized system user | GFT8 (HDI9, HDI61, HDI68) |
| Other | | |
| CFR1 | AEGIS has to provide an organized (online-offline/in private cloud) storage for the data analysis | GFT1 (HDI1) |
| CFR13 | AEGIS provides a function to upload in-house datasets to AEGIS cloud | GFT20 (HDI20) |
| CFR14 | AEGIS provides a private cloud area for dataset storage | GFT21 (HDI21) |
| CFR19 | AEGIS should allow uploading of compatible user's code to process their data | KTH4 (VIF4) |
| CFR2 | AEGIS should allow to load previous results (if available), select data sources, type of analytics and input parameters | GFT2 (HDI1) |
| CFR9 | AEGIS should provide the ability to handle queries spanning multiple datasets/data sources and the possibility to search in a query storage | GFT9 (HDI12, HDI34, HDI40, HDI41, HDI42, HDI45, HDI49, HDI50, HDI62, HDI63, HDI64) |

#### 2.4.1.2. Demonstrator User Requirements

**Table 2-7** shows the collection of the Demonstrator User Requirements.

**Table 2-7: Demonstrator Functional requirements**

| Id | Requirement description | References |
|---|---|---|
| | Analytic | |
| DFR1 | AEGIS needs to have the ability to process data gathered from recording sensors and dedicated devices in real time (e.g. geo-localized data - data events and customers) | GFT10 (HDI13, HDI59, HDI60) |
| DFR23 | AEGIS should be able to collect social activity data of end-users in order to perform analyses on them | UBI13 (HYP18, HYP37) |
| DFR27 | AEGIS should be able to extract accurate personal safety profiles / personas to classify elderly people to different vulnerability levels | UBI25 (HYP30, HYP38) |
| DFR28 | AEGIS should be able to extract accurate risk patterns | UBI27 (HYP31) |
| | Correlation | |
| DFR17 | AEGIS should be able to get traffic data and correlate it with users imported data | KTH8 (VIF12, VIF16) |
| DFR18 | AEGIS should be able to get accidents hotspots data and correlate it with users imported data | KTH9 (VIF14, VIF16) |
| DFR19 | AEGIS should be able to crawl social networks for traffic conditions and correlate such data with users imported data | KTH14 (VIF27,VIF29) |
| DFR20 | AEGIS should be able to crawl newspapers for traffic conditions and correlate such data with users imported data | KTH15 (VIF28,VIF29) |
| DFR21 | Different Stakeholders should be able to generate regional driving safety risk metric report and take actions accordingly | KTH16 (VIF30,VIF31,VIF32,VIF33,VIF34) |
| DFR29 | AEGIS should be able to correlate positioning information data from mobile phones and wearable devices with identified PSPS events | UBI31 (HYP35) |
| DFR31 | AEGIS should be able to identify trends and outliners on end users | UBI34 (HYP41) |
| DFR6 | AEGIS has to correlate events' geospatial references with their description and provide feedback concerning the risk | GFT16 (HDI17, HDI67, HDI69) |

| | | |
|---|---|---|
| DFR9 | User can enrich service output with extra information from a list of preconfigured datasets (e.g. Institutional Open Data) | GFT23 (HDI21) |
| | Dashboard/Report/Alert | |
| DFR13 | Users should be able to specify departure and destination points on map and then choose the best route according to the roads conditions | KTH5 (VIF5, VIF6) |
| DFR14 | Different Stakeholders should be able to view and group the list of damaged roads | KTH6 (VIF7, VIF8, VIF9, VIF10) |
| DFR15 | Different Stakeholders should be able to generate driver's safety style report and take actions accordingly | KTH11 (VIF17, VIF18, VIF19, VIF20, VIF21, VIF22) |
| DFR30 | AEGIS should be able to create rules for notifications | UBI33 (HYP39) |
| DFR32 | AEGIS should provide a dashboard with advanced analytics and intuitive visualisations regarding risks for the health and well-being of individuals, potential cognitive deterioration and physical wellbeing deterioration and frailty status | UBI35 (HYP42, HYP43, HYP44) |
| DFR33 | AEGIS should identify and provide recommendations and/or notifications and/or simplified alerts in cases of identified risks or when certain thresholds have been exceeded (e.g. risks for health and well-being or potential cognitive / physical wellbeing deterioration of individuals) | UBI36 (HYP33, HYP34, HYP35, HYP45, HYP46, HYP47, HYP48, HYP49, HYP50, HYP51) |
| DFR5 | AEGIS app should send an alarm notification to the customer following rules specified by the user | GFT15 (HDI16, HDI65) |
| | Input / Output | |
| DFR22 | AEGIS should be able to determine the location of the user from his device | UBI4 (HYP7) |
| | Privacy | |
| DFR8 | AEGIS should mask datasets before uploading | GFT22 (HDI21) |
| | Rules/Management | |
| DFR10 | User can configure algorithms parameters, input dimensions to analyse his datasets | GFT27 (HDI27) |
| DFR11 | Users should be able to choose a service to work on their data | KTH1 (VIF1, VIF13, VIF26) |
| DFR2 | AEGIS has to upload and filter (through semantic rules) interesting events as soon as they appear in a predefined list of web pages and record them following semantic rules (hence data enabled for further filtering) | GFT11 (HDI13, HDI16) |

| | | |
|---|---|---|
| DFR24 | User should be able to select the data to be made available to the AEGIS apps / for processing by the AEGIS platform. | UBI14 (HYP19, HYP20, HYP36) |
| DFR3 | Automate (real-time) upload from the customer AEGIS app of the customer information and rules to evaluate possible losses | GFT13 (HDI15, HDI44, HDI51) |
| DFR4 | The Data Analyst should create his own rules (query) to identify the data to analyse | GFT14 (HDI16, HDI33, HDI58) |
| Specific Services | | |
| DFR12 | AEGIS should provide predefined algorithms for some tasks such as road detection | KTH3 (VIF3) |
| DFR16 | Users should be able to quantify their driving style and compare it to other drivers | KTH12 (VIF23) |
| DFR25 | AEGIS should be able to extract thermal and visual comfort profiles | UBI18 (HYP24) |
| DFR26 | Provide control actions on HVAC and lighting devices towards a home automation environment | UBI20 (HYP25, HYP26, HYP27, HYP28) |
| DFR7 | AEGIS has to implement non-intrusive predictive models to profile customers and identify a personalized offering for individuals to best fit their habits and needs | GFT18 (HDI19) |

*2.4.2. Non - Functional Requirements*

To identify the not functional requirements, the model proposed by ISO/IEC 25010:2011 was adopted. Following that model there are eight quality characteristics contributing to software product quality (**Figure 2-4**).

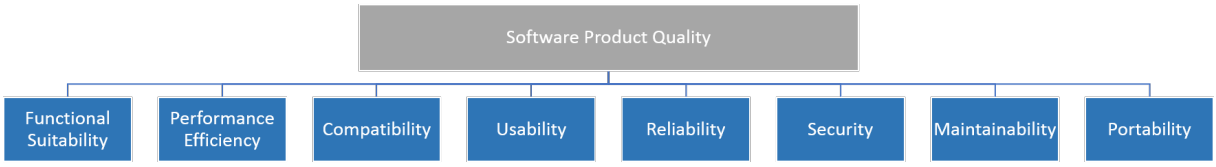



**Figure 2-4: ISO/IEC 25010:2011 Software Product Quality model**

Each characteristic deal with a specific framework within the software product quality; to have a better understanding of them, they can be well detailed (ISO/IEC 25010:2011) in sub-categories as reported in **Table 2-8**.

**Table 2-8: ISO/IEC 25010:2011 Software Product Quality Model Sub-Categories**

| Quality characteristic | Sub-categories |
|---|---|
| Functional Suitability | • Functional Completeness<br>• Functional Correctness<br>• Functional Appropriateness |
| Performance Efficiency | • Time Behaviour<br>• Resource Utilization<br>• Capacity |
| Compatibility | • Co-existence<br>• Interoperability |
| Usability | • Appropriateness Recognisability<br>• Learnability<br>• Operability<br>• User Error Protection<br>• User Interface Aesthetics<br>• Accessibility |

| | |
|---|---|
| Reliability | • Maturity<br>• Availability<br>• Fault Tolerance<br>• Recoverability |
| Security | • Confidentiality<br>• Integrity<br>• Non-repudiation<br>• Authenticity<br>• Accountability |
| Maintainability | • Modularity<br>• Reusability<br>• Analysability<br>• Modifiability<br>• Testability |
| Portability: | • Adaptability<br>• Installability<br>• Replaceability |

**Table 2-9** shows the not functional requirements pointed out for the AEGIS solution, taking into account the eight qualities aforementioned.

**Table 2-9: Non-Functional requirements**

| Requirement Sub-category | Id | Description (Detailed description of the requirement) |
|---|---|---|
| Functional Suitability | NFR1 | AEGIS should be able to perform a great variety of analysis on selected multiple datasets (private or public) in order to provide state of the art analytics |
| Performance efficiency | NFR2 | AEGIS should be able to perform analytics in a timely and efficient manner |
| | NFR3 | AEGIS should be able to guarantee the full optimization of the response time to ensure a functional and flexible navigation through the AEGIS solution |
| Compatibility | NFR4 | AEGIS should be able to interact and exchange information with other products / IoT devices in a secure |

| | | |
|---|---|---|
| | | way (for example secure REST API) |
| | NFR5 | AEGIS should have the information distributed in a cloud system and be able to merge the heterogeneous information of the cloud |
| | NFR6 | AEGIS has to provide communication capabilities to allow other applications to interact with the AEGIS platform |
| Usability | NFR7 | AEGIS shall feature a user-friendly interface, provide an overview of supported kind of analytics and visualization through both online and offline user guides (i.e. demo guide, manuals and documentation database) |
| | NFR8 | AEGIS should have a Multi-language user interface |
| | NFR9 | AEGIS should provide an attractive application icon that makes the application easily distinguishable. |
| | NFR10 | AEGIS should enable to track logs about evolution and faults history and periodically send debug reports |
| Reliability | NFR11 | AEGIS should be able to securely store uploaded private datasets |
| | NFR12 | AEGIS should provide a strong algorithm to recognize aliases whenever there is a duplicate resolution |
| | NFR13 | AEGIS should ensure high availability of the system and the stored datasets |
| Security | NFR14 | AEGIS should offer login with user credentials |
| | NFR15 | AEGIS should handle the geo-referencing data only at the application level |
| | NFR16 | AEGIS should take into account privacy and security rules according to national legislation |
| | NFR17 | AEGIS should verify the authenticity of its data sources |
| Maintainability | NFR18 | AEGIS should provide a periodically data maintenance and/or replacement during the period 8p.m to 8a.m. |
| | NFR19 | AEGIS should be easily maintainable (e.g. upgrades, bug fixing) and the user has to be informed on time |
| | NFR20 | AEGIS should be able to raise alarms about hardware/software failures of the solution |
| Portability | NFR21 | AEGIS should be able to be deployed on various Linux distributions |

| | | |
|---|---|---|
| | NFR22 | AEGIS should be able to be deployed in a timely and efficient manner. |

# 3. TECHNICAL REQUIREMENTS

## 3.1. LIST OF TECHNICAL REQUIREMENTS

The scope of this section is to define the technical requirements which will drive the definition of the AEGIS architecture. The technical requirements were driven from the list of user requirements as defined in the previous section and with the mission to utilize the goals and expectations of the users and the stakeholders.

The following list defines the technical requirements that will pave the way for the design and implementation of AEGIS.

**Table 3-1: Technical Requirements**

| ID | Correlates to functional/ non-functional requirement: | Need | Description | Priority | Component |
|---|---|---|---|---|---|
| TR1 | CFR15, CFR22, CFR29, CFR30, CFR31, CFR32, CFR33, CFR34, CFR37, CFR41, CFR42, CFR43, CFR44, CFR49, CFR50, NFR4 | Process measurements from "sensor nodes" | AEGIS should be able to connect to, receive and process data from sensors and IoT devices, such as from smart home sensors and/or wearables. | High | Data Harvester, Data Annotator, Algorithm Execution Container |
| TR2 | CFR45 | Process measurements from social media sources | AEGIS should be able to connect to, receive and process data from social media sources such as Facebook and Twitter. | High | Data Harvester, Data Annotator, Algorithm Execution Container |
| TR3 | CFR10, CFR35 CFR37, CFR38, CFR43, NFR4 | Process data from mobile devices | AEGIS should be able to connect to, receive and process data from mobile devices, including GPS signal, acceleration / gyroscope information etc. | High | Data Harvester, Data Annotator, Algorithm Execution Container |
| TR4 | CFR23, CFR24, CFR25, CFR26, CFR27, CFR28, CFR35, CFR36, CFR46, CFR47 | Process data from other open sources | AEGIS should be able to connect to, receive and process data from open sources including for example weather data etc. | High | Data Harvester, Data Annotator, Algorithm Execution Container |

| | | | | | |
|---|---|---|---|---|---|
| TR5 | CFR9, CFR11, CFR13, CFR15, CFR16 | Process data from relational databases | AEGIS should be able to connect to, receive and process data from relational databases. | High | Data Harvester,<br>Data Annotator,<br>Algorithm Execution Container |
| TR6 | CFR9, CFR11, CFR13, CFR15, CFR16 | Process data from NoSQL databases | AEGIS should be able to connect to, receive and process data from NoSQL databases. | High | Data Harvester,<br>Data Annotator,<br>Algorithm Execution Container |
| TR7 | CFR9, CFR11, CFR13, CFR15, CFR16 | Process data from structured sources | AEGIS should be able to connect to, receive and process documents in structured format (e.g. JSON, XML). | High | Data Harvester,<br>Data Annotator,<br>Algorithm Execution Container |
| TR8 | CFR9, CFR11, CFR13, CFR15, CFR16 | Process data from semi-structured sources | AEGIS should be able to connect to, receive and process documents in semi-structured format (e.g. csv files). | High | Data Harvester,<br>Data Annotator,<br>Algorithm Execution Container |
| TR9 | CFR15, CFR16, NFR4 | Connect to data sources through well-defined APIs. | AEGIS should be able to consume and store data through well-defined APIs. | High | Data Harvester,<br>Data Annotator |
| TR10 | CFR39, CFR51, CFR52, CFR53, CFR54 | Configure data sampling / update period | AEGIS should be able to allow for the configuration of the data sampling / update period. | High | Data Harvester,<br>Data Annotator |
| TR11 | CFR1, CFR9, CFR13, CFR14, CFR16 | Store big data | AEGIS should be able to store large datasets (Big Data Infrastructure). | High | Data Harvester,<br>Data Annotator, |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | Data Store |
| TR12 | CFR9, CFR18 | Produce RDF Triples | AEGIS should be able to produce RDF Triples out of the big datasets or out of the metadata of these big datasets. | High | Data Harvester,<br>Data Annotator |
| TR13 | CFR1 | Store RDF Triples | AEGIS should be able to allow for the storage of the produced RDF triples. | High | Data Store |
| TR14 | NFR5 | Handle big data scalability | AEGIS should be able to handle (query and process) large datasets (Big Data) and be able to scale horizontally. | High | Data Store, Query Builder,<br>Algorithm Execution Container, Big Data Processing Cluster |
| TR15 | NFR6 | Ensure data availability | AEGIS should be able to ensure data availability, so that data is available for data consumers at a required level of performance during normal or extraordinary operation. | High | Data Store |
| TR16 | NFR13 | Ensure data persistence | AEGIS should be able to ensure data persistence so that users can be confident that their own data, and the data used by their processes/services, is always available. | High | Data Store |
| TR17 | CFR1, CFR3 | Ensure data consistency | AEGIS should be able to ensure consistency, so that data storage is always conducted according to specific rules, and it is therefore easier to consume the required data. | High | Data Store,<br>Data Harvester,<br>Data Annotator |
| TR18 | CFR1, CFR9, CFR13, CFR14, CFR16, NFR5 | Offer distributed storage | AEGIS should be able to offer distributed storage. | High | Data Store |
| TR19 | CFR1, CFR8 | Offer public storage | AEGIS should be able to provide public storage in order to allow access to all (non-malicious) third parties. | High | Data Store |

| | | | | | |
|---|---|---|---|---|---|
| TR20 | CFR1, CFR6, CFR7, CFR8, CFR11, CFR14, | Offer private storage | AEGIS should be able to provide private storage in order to prevent access from unauthorised third parties. | High | Data Store |
| TR21 | CFR1, CFR14, NFR11, NFR17 | Provide ability to users to securely upload datasets | AEGIS should provide the ability to its users to upload datasets. | High | Data Store, Data Harvester, Data Annotator |
| TR22 | CFR6, CFR21, CFR48, NFR16 | Provide ability to declare IPR on uploaded datasets and set access restrictions | AEGIS should provide the ability to its users to set access restrictions to the datasets they upload to serve as the pre-conditions for the formulation of micro-contracts. | High | Brokerage Engine |
| TR23 | CFR6, CFR7, CFR8, CFR21, CFR48 | Access data | AEGIS should be able to provide to its (authorised) users data access capabilities from original and/from originating datasets, allowing them to export the required data from the available knowledge bases (e.g. based on SPARQL queries). | High | Data Store, Query Builder |
| TR24 | CFR39 | Select data | AEGIS should be able to provide to its users data selection capabilities from original and/from originating datasets. | High | Data Store |
| TR25 | CFR7, CFR8, | Implement a data management policy | AEGIS should define and implement a data management policy to restrict unauthorised access to restricted datasets. | High | Data Store,<br>Brokerage Engine |
| TR26 | CFR6, CFR7, CFR8, NFR14 | Support authentication mechanisms for access to data | AEGIS should be able to support authentication mechanisms for restriction of malicious-driven access to data. | High | Data Store |

| | | | | | |
|---|---|---|---|---|---|
| TR27 | CFR6, CFR7, CFR8 | Secure datasets | AEGIS should be able to offer security and encryption of the datasets so as to prevent unauthorised access to and tampering with stored datasets. | High | Data Store |
| TR28 | CFR3, NFR12 | Clean datasets | AEGIS should be able to offer (offline) tools for data cleansing. | Medium | Data Harvester, Data Annotator |
| TR29 | CFR48 | Anonymise datasets | AEGIS should be able to offer (offline) tool for (sensitive / personal) data anonymization. | Medium | Anonymization Tool |
| TR30 | CFR9, CFR11, CFR15 | Support interlinking of AEGIS triplestore with SLOD | AEGIS should be able to support the interlinking of the AEGIS triplestore with the SLOD to provide open access to data. | Medium | Data Store |
| TR31 | CFR9 | Provide SPARQL Endpoints | AEGIS should be able to provide SPARQL Endpoints for supporting federated queries over distributed datasets. | High | Data Store |
| TR32 | CFR3, CFR16, | Support conversion of multiple source formats to RDF | AEGIS should be able to provide the means for transforming data to RDF and support the conversion of multiple source formats –including structured and unstructured data. | High | Data Harvester, Data Annotator |
| TR33 | CFR3, CFR16, | Support conversion of multiple source formats to multiple formats | AEGIS should be able to provide the means for transforming data in one format to another format (selection amongst a close list of formats) and support the conversion of multiple source formats –including structured and unstructured data. | Medium | Data Harvester, Data Annotator |

| | | | | | |
|---|---|---|---|---|---|
| TR34 | CFR3, CFR9, CFR11, CFR15, CFR16, | Support interlinking of datasets from different sources. | AEGIS should be able to support automatic or semi-automatic interlinking of datasets from different sources. AEGIS should be able to configure the interlinking task by selecting meta level items (properties, classes) or instance level items (individuals) to be interlinked. | High | Data Harvester, Data Annotator |
| TR35 | | Support the reuse of available canonical standard vocabularies and ontologies. | AEGIS should be able to support the automatic or semi-automatic construction and derivation of vocabularies and ontologies from the data sources. | High | Data Harvester, Data Annotator, AEGIS ontologies and vocabularies |
| TR36 | CFR3 | Deal with corrupted, inconsistent data or misconfiguration of mapping rules | AEGIS should be able to provide measures to deal with corrupted, inconsistent data or misconfiguration of mapping rules and means for maintenance of data, in order to ensure the quality of the checked in and/or the transformed RDF data. | Medium | Data Harvester, Data Annotator |
| TR37 | CFR1, CFR13, CFR14 | Support management of data sources | AEGIS should be able to support the management of data sources through various functionalities: e.g. add/import, select and upload functionalities. | High | Data Harvester, Data Annotator, Data Store |
| TR38 | | Add metadata, enriching the existing data sources | AEGIS should provide the means for adding metadata/enriching the existing data sources. | Medium | Data Harvester, Data Annotator, AEGIS ontologies and vocabularies |
| TR39 | | Select / Search for / Define vocabularies | AEGIS should provide the means for selecting/ searching/ defining (Linked Data) vocabularies. | Medium | Data Annotator, AEGIS ontologies and vocabularies |

| | | | | | |
|---|---|---|---|---|---|
| TR40 | CFR1, NFR17 | Inspect data sources | AEGIS should allow for the inspection of data sources (and for the tentative mapping of the data sources to the available vocabularies). | Medium | Data Harvester, Data Annotator, AEGIS ontologies and vocabularies |
| TR41 | CFR9, NFR1 | Ability to handle queries spanning multiple datasets | AEGIS should be able to handle queries spanning multiple datasets. | High | Algorithm Execution Container, Query Builder |
| TR42 | CFR16 | Support format conversion | AEGIS should be able to support conversion of RDF to any format as well as various export formats including CSV, JSON, XML. | High | Data Harvester |
| TR43 | | Manually balance transformation automation and semantic soundness | AEGIS should provide the means to the user to fine-tune the balance of transformation automation and semantic soundness, whenever the automatization of some transformation is potentially speculative (e.g., guessing link types from column headers). | Low | Data Harvester, Data Annotator |
| TR44 | CFR4, CFR5, CFR17, CFR18, | Support data analytics on distributed big data | AEGIS should be able to support a set of different types of robust algorithms for data analytics on top of big data based on the needs of end users. Ideally, AEGIS should be able to reuse open source frameworks for machine learning algorithms. | High | Algorithm Execution Container, Query Builder, Big Data Processing Cluster |
| TR45 | CFR4, CFR5, CFR17, CFR18, | Support data analytics on distributed big data | AEGIS should be able to support a set of different types of robust algorithms for data analytics on top of distributed big data based on the needs of end users. Ideally, AEGIS should be able to reuse open source frameworks for machine learning algorithms. | High | Algorithm Execution Container, Big Data Processing Cluster |

| | | | | | |
|---|---|---|---|---|---|
| TR46 | CFR2 | Support initialization of algorithms | AEGIS should be able to support the initialisation of available algorithms supporting big data analytics, by defining the available set of algorithm-specific parameters prior to the execution of the algorithms. | High | Algorithm Execution Container |
| TR47 | CFR4, CFR17, CFR18, NFR2 | Support the processing of data in multiple formats. | AEGIS should be able to support the execution of robust algorithms for data analytics on top of (distributed) big data made available in various formats, whether structured or semi-structured. | High | Algorithm Execution Container |
| TR48 | CFR2, CFR40, NFR2 | Support a set of different types of output formats. | AEGIS should be able to support the execution of robust algorithms for data analytics on top of (distributed) big data making the results (data output) available in various formats. | High | Algorithm Execution Container |
| TR49 | CFR2, NFR2 | Customise the analytics process | AEGIS should be able to support functionalities for customization of the analytics process. AEGIS should be able to load previous results (optional and if available), select data sources select type of analytics and input parameters. | High | Algorithm Execution Container |
| TR50 | CFR45 | Support real-time analysis | AEGIS should be able to support analysis based on real time data feeds and predefined processes. | | Algorithm Execution Container |
| TR51 | CFR2, CFR40 | Manage the results of the big data enabled analytics | AEGIS should be able to support the management of the results of the big data enabled analytics, including ability to export the analytics results / output, to save and/or to visualize the produced results. | High | Algorithm Execution Container<br>Visualizer |
| TR52 | CFR2, NFR2 | Support overview of resources | AEGIS should be able to provide an interface to the users to overview the resources to be used during analytics, including the resources | High | Data Management<br>Visualizer |

| | | | | | |
|---|---|---|---|---|---|
| | | | description, their preliminary visualization, the view of sample of the whole dataset etc. | | |
| TR53 | CFR18, NFR3 | Preview a small selection of the results of the generated query | AEGIS should provide the means to the user to preview a small selection of the results of the generated query so as to extract some initial insights out of the foreseen analytics | Medium | Visualizer |
| TR54 | CFR5, CFR9, CFR18 | Support a set of different types of visualizations | AEGIS should be able to support a set of different types of visualizations based on different types of input datasets formats. Provide means for visualizing different data modalities (e. g. special, temporal, statistical) and provide an overview of the supported kinds of visualization | High | Visualizer |
| TR55 | CFR5, CFR9, CFR18, | Create advanced graphs based upon queries spanning multiple datasets | AEGIS should be able to provide to its users the ability to create advanced graphs based upon queries spanning multiple datasets | High | Visualizer |
| TR56 | CFR6, CFR7 | Support secure communication among the AEGIS components | AEGIS should be able to guarantee security and trustworthiness in the communication among the various platform components, providing secure technical interfaces between them | High | Data Harvester,<br>Data Annotator,<br>Data Indexing Engine,<br>Algorithm Execution Container,<br>Query Builder,<br>Visualizer,<br>Big Data Processing Cluster, |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | Brokerage Engine,<br>Data Store |
| TR57 | CFR2, CFR40 | Save analytics results | AEGIS should be able to provide to the users the ability to save their projects / results of the executed analysis for further reuse or analysis | High | Visualizer, Data Store |
| TR58 | CFR40 | Support authorized sharing of saved results | AEGIS should be able to support the "distribution" of projects / results saved by the users, with configurable sharing visibilities, namely making the results available to a subset of users authorised to access them. | Medium | Visualizer, Brokerage Engine |
| TR59 | CFR13, CFR48 | Support local deployment of the anonymization tool | AEGIS should be able to support and provide deployment options for the anonymization tool on a local platform, and provide the interfaces for the management of the tools and its execution options | Low | Anonymization Tool |
| TR60 | CFR3 | Support local deployment of the cleansing tool | AEGIS should be able to support and provide deployment options for the cleansing tool on a local platform, and provide the interfaces for the management of the tools and its execution options | Low | Cleansing Tool |
| TR61 | CFR3 | Support local deployment of the harvesting tool | AEGIS should be able to support and provide deployment options for the data harvesting tool on a local platform, and provide the interfaces for the management of the tools and its execution options | Low | Data Harvester |
| TR62 | CFR3 | Support local deployment of the | AEGIS should be able to support and provide deployment options for the transformation engine on a local platform, and provide the | Low | Data Harvester |

| | | | | | |
|---|---|---|---|---|---|
| | | transformation tool | interfaces for the management of the tools and its execution options | | |
| TR63 | CFR8 | Provide push notifications | AEGIS should be able to support the ability to provide notifications to end users (e.g. in the case an analytics task has been finalised, a dataset has been updated etc.) | High | Data Harvester |
| TR64 | CFR2, CFR16, CFR21, CFR40, NFR3, NFR7, NFR8 | Provide secure user-friendly interface and flexible navigation. | AEGIS should be able to provide a user-friendly interface so that the users can easily navigate between the functionalities offered by the platform in a secure way. | High | Visualizer, Query Builder, Data Harvester, Data Annotator |
| TR65 | NFR10, NFR20 | Provide system management dashboard | AEGIS should be able to offer to the Administrator a dashboard with system overview like log reports, transaction and activity report, system alarms, storage and general health report of the system. | High | AEGIS Web GUI |
| TR66 | CFR10, NFR15 | Support geospatial data analysis | AEGIS should be able to support a set of different types of algorithms for data analytics on top of geospatial data analysis | High | Algorithm Execution Container |
| TR67 | NFR18, NFR19 | Provide sustainability of system | AEGIS should be easily maintainable (upgrade and update) with a consistent and robust procedure. | High | |
| TR68 | NFR21, NFR22 | Provide a solid deployment workflow of the system | AEGIS should be able to be deployed in a timely and efficient manner. | Medium | |
| TR69 | CFR19, CFR20 | Support uploading user’s | AEGIS should provide a secure way for the user upload code implementing algorithms for data analysis. The uploaded code should be reviewed | Low | Algorithm Execution Container |

| | | | | | |
|---|---|---|---|---|---|
| | | custom algorithm implementation | and approved by the moderators before it can be used. | | |

## 3.2. MAPPING OF ACTORS AND TECHNICAL REQUIREMENTS

In order to better visualize the relationship between the technical requirements listed above with the AEGIS actors the mindmap[1] approach has been adopted and for every actor a mindmap has been created.

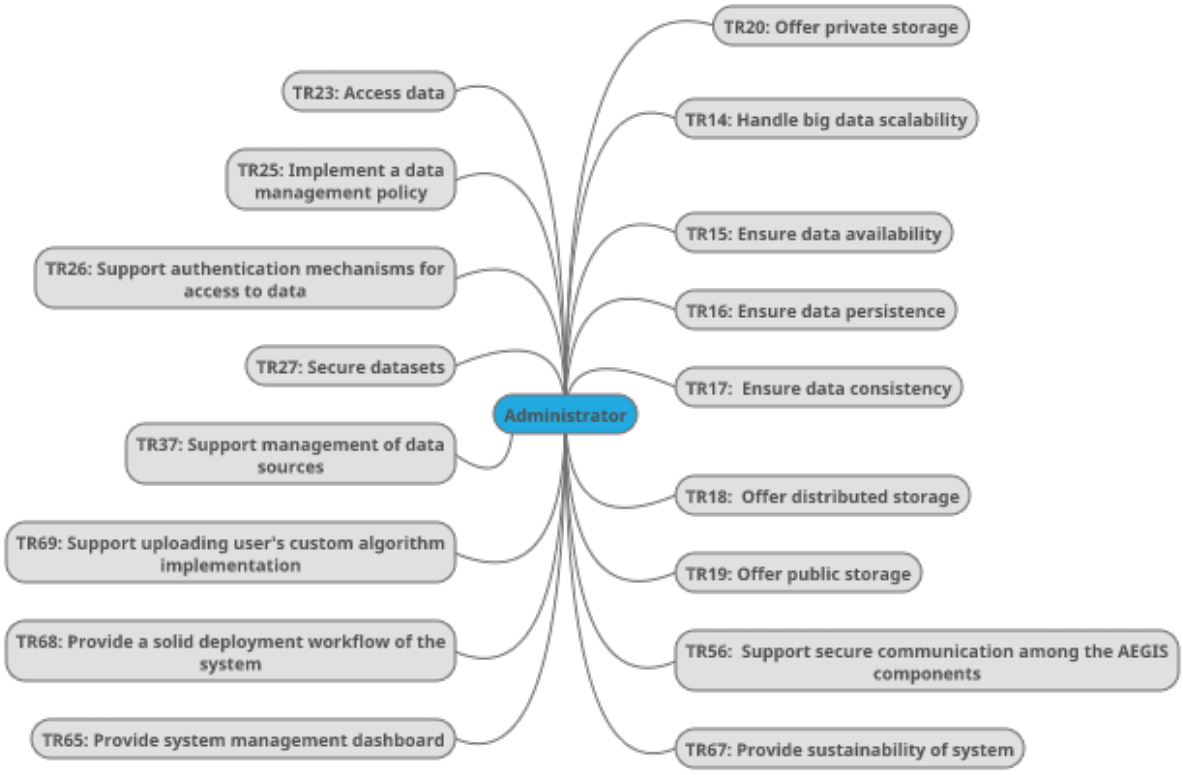



**Figure 3-1: Administrator**

[1] Please refer to http://www.mindmapping.com/

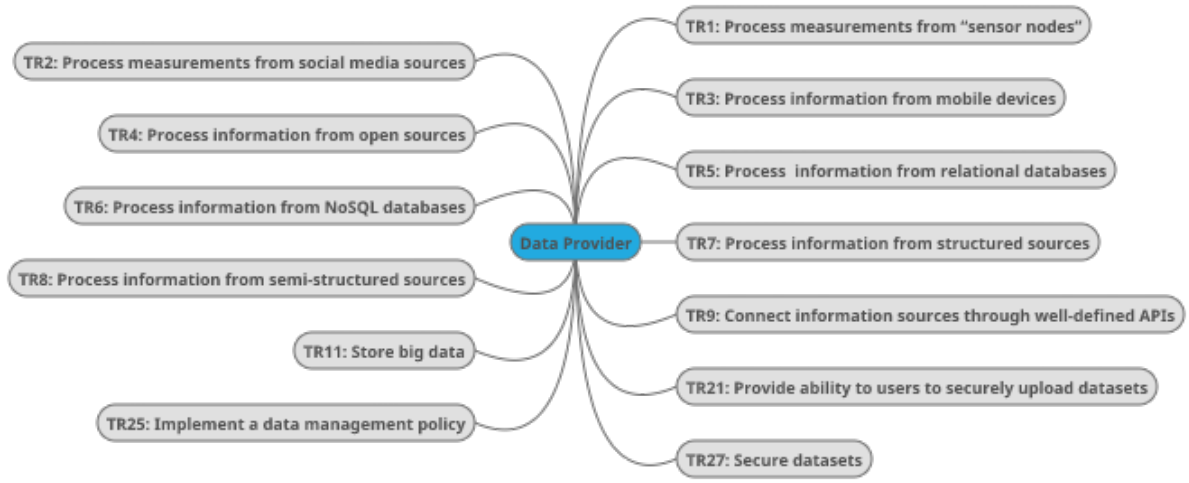



**Figure 3-2: Data Provider**

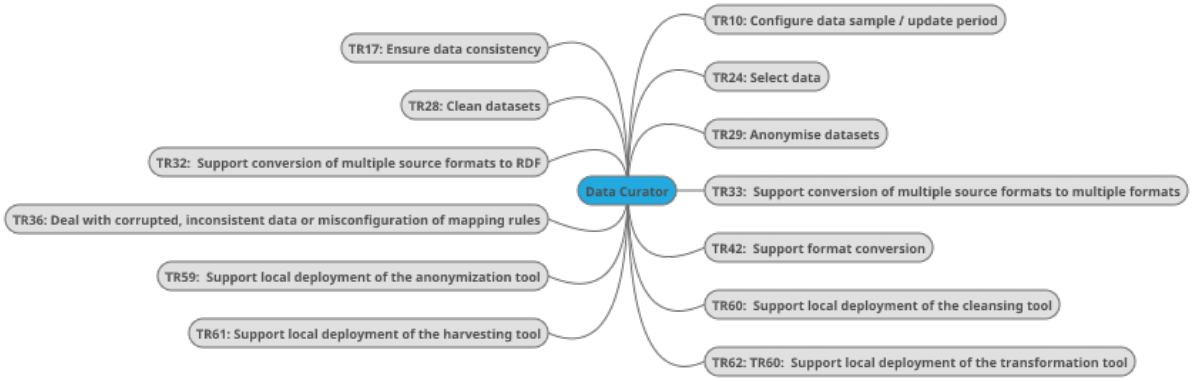



**Figure 3-3: Data Curator**

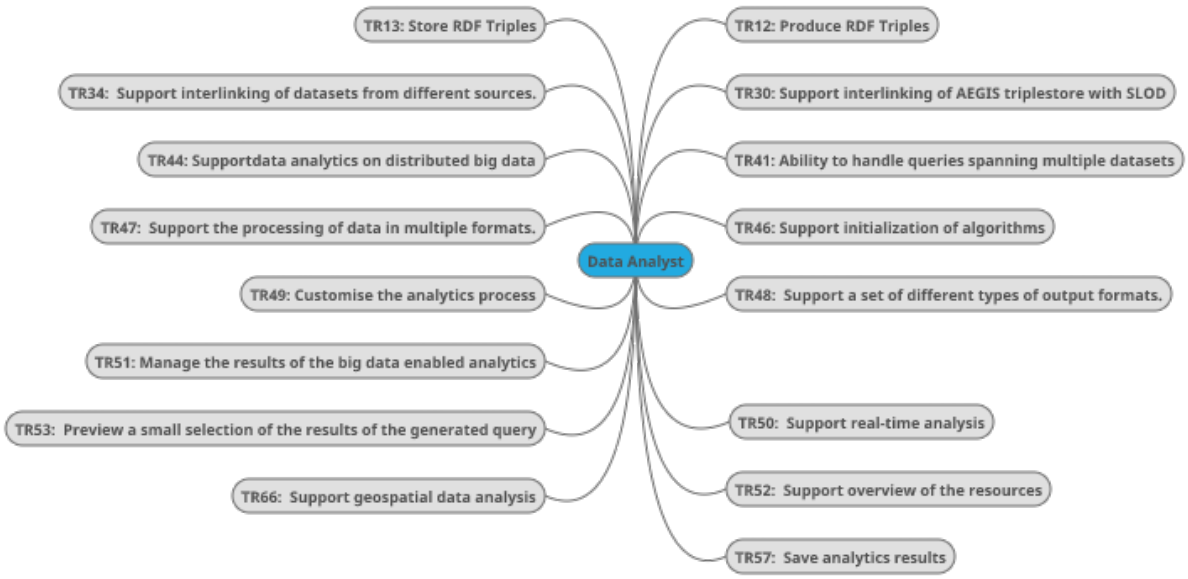



**Figure 3-4: Data Analyst**

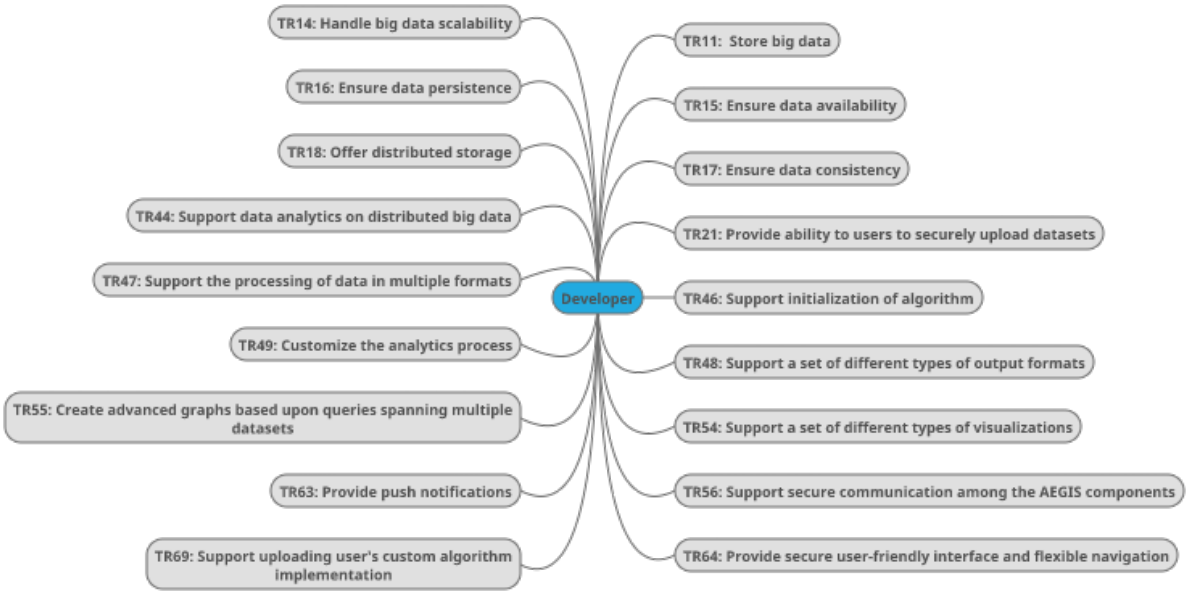



**Figure 3-5: Developer**

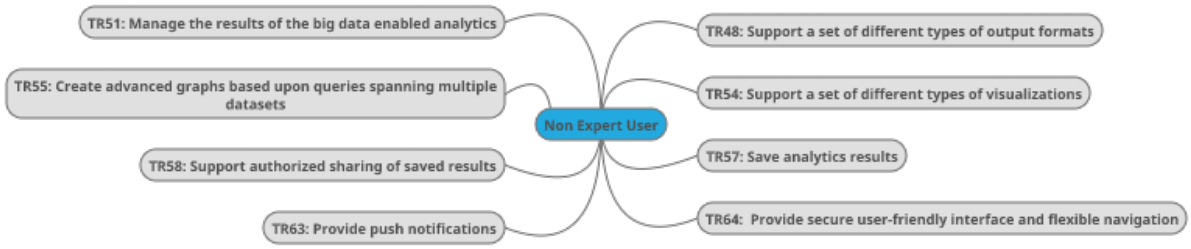



**Figure 3-6: Non-Expert User**

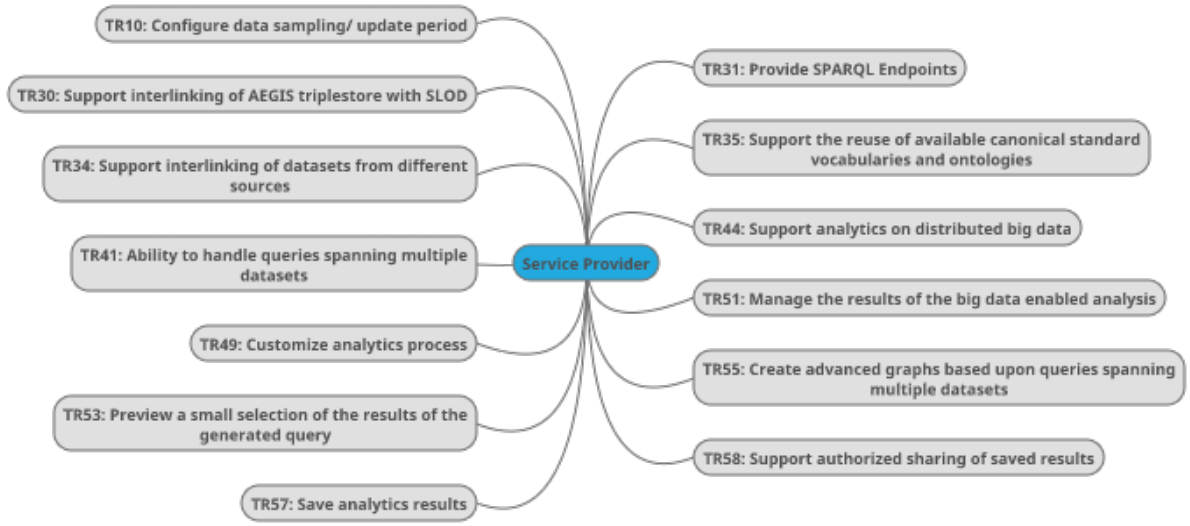



**Figure 3-7: Service Provider**

# 4. AEGIS FRAMEWORK ARCHITECTURE

## 4.1. ADDRESSING STAKEHOLDER'S REQUIREMENTS

In a previous section a concrete analysis of AEGIS user requirements has been described from which the technical requirements of AEGIS derived. After a careful analysis of these requirements, the scope of this section is to summarize the most critical technical requirements which will hold the most significant role in the design of the high-level architecture of AEGIS. Following this approach, it will guide us to the definition of the high-level description of the components that will form the architecture of AEGIS. The purpose of AEGIS is to provide a novel secure data handling and management framework by exploiting advances in big data management, linked data, data harmonisation mechanisms and data analytics technologies.

To reach our goal, a "bottom-up" approach was adapted, that will progressively guide us to the high-level architecture. In the course of following this approach, several challenges have been identified and have been addressed by the components.

First of all, there is a need that AEGIS can connect, receive and consume data from various sources, stored or received in various heterogeneous formats like different databases, structured or semi-structured documents, mobile phones or wearables. These data can be equipped with some metadata (correct or even partially correct) or with no metadata. For AEGIS, a component or a chain of components that will be able to import and transform the aforementioned data offering an effective way of harmonization in parallel is highly needed. Overcoming the difficulty of heterogeneous data will offer great potential to AEGIS and the data analysis that will be performed to address the stakeholder's requirements, along with growing interest on the AEGIS services offered. Furthermore, by annotating the aforementioned data sets with domain-specific upper ontologies and vocabularies will provide common ground for the semantic linking among a variety of data types and will further increase the value of the data analysis performed.

In every Big Data ecosystem, the main component of the system is the data storage component. There is no doubt that it has a key role in the system and this is the reason why the requirements linked to this component are more than crucial for the success of the AEGIS. However, due to this key role, there is an increased level of difficulty on addressing those requirements. In general characteristics like scalability, high availability, consistency, security, flexibility and efficiency are those that describe this component. Besides handling large amounts of data, it should be able also to provide input/output operations per second (IOPS) necessary to deliver data to analytics tools that will be used.

Additionally, another crucial component of the system is the one responsible for executing the data analysis. This could be broken in several parts like the one initiating and monitoring the execution of the algorithms, the one actually performing the execution and analysis, and the low-level part which is controlling the environment and the resources needed so that the execution can take place. The first one should offer a user-friendly interface providing the list of the available datasets, a great variety of the algorithms that can be performed on these datasets and the possibility to parameterize the algorithm upon need. The second should be responsible for organizing and scheduling the execution of the algorithm and will be also responsible for orchestrating the parts that need to be executed. The data analysis must be

performed on highly efficient cluster-computing environment. The last one should be responsible for the efficient resource management of this cluster-environment.

Another critical requirement is the offering of querying capabilities of the system. AEGIS should be offering a user-friendly interface to perform queries supplemented with advance tools and easy-to-use technics. This component will give the insight of the datasets available in AEGIS to the users and should also perform in a timely and efficient manner.

Finally, AEGIS should be able to visualize the results of the algorithm execution or the query execution. It should offer a variety of visualization option including maps, line plots and tables that can be combined in an interactive dashboard. This component should be able to offer a visual representation of the results to the user upon needs.

## 4.2. HIGH LEVEL ARCHITECTURE DESCRIPTION

A high level architecture of the first version of the AEGIS platform has been designed, taking into consideration all of the aforementioned requirements. This conceptual high-level architecture is depicted in the figure below.

The high-level architecture describes the complete lifecycle of the AEGIS supported processes, starting from the selection and extraction of the valuable datasets from the identified data sources besides the batch upload of large datasets, to their semantic annotation, to the extraction of RDF triples and their storage, and continuing with advanced analytics including state of the art big data focused algorithm execution and sophisticated processing, complemented by advanced visualizations of the analysis results.

Data Harvester consists of all applications enabling the import of the data and metadata of the original data set, along with any possible transformations required. Since the original data sets provided may be in multiple formats it is essential that a wide spectrum of possible forms of data are supported. In general data sets can be in the form of a structured (e.g. XML or JSON) or semi-structured (e.g. Excel) document and in the form of a database. Finally, one form that should be considered and properly handled is the streaming data source. Data Harvester should include the identification of the meaningful items based on predefined rules and the separation of them from the rest of the data that is not of direct value to the AEGIS stakeholders. This includes the transformation in the target format. The metadata will be always transformed to linked data. The data in some cases, depending from its kind, will be transformed to linked data as well, or kept in the original format or transformed in another selected target format.

Data Annotator component refines the output of Data Harvester with the main purpose being the enrichment of the metadata or linked data using predefined ontologies and vocabularies. Through the semantic annotation, the concepts included in the selected subset will be related to well defined semantics. Semantic annotations provide information 'about' the data, for example the meaning or what the data is about and the available semantic relationships from a domain model in which the data is defined. The purpose of semantically annotating a dataset is to create a context in terms of the content and functionality of the data so that it can be easily interpreted, combined and reused by computers.

In the context of AEGIS the format-describing annotations are required to specify subsequent data extractions. The annotations have to specify the correlatability of the data in question (e.g.,

the semantic nature of the X and Y axes). The AEGIS consortium has identified a set of upper domain-specific ontologies and core vocabularies which will be extended so that the domain specific business needs identified can be covered.

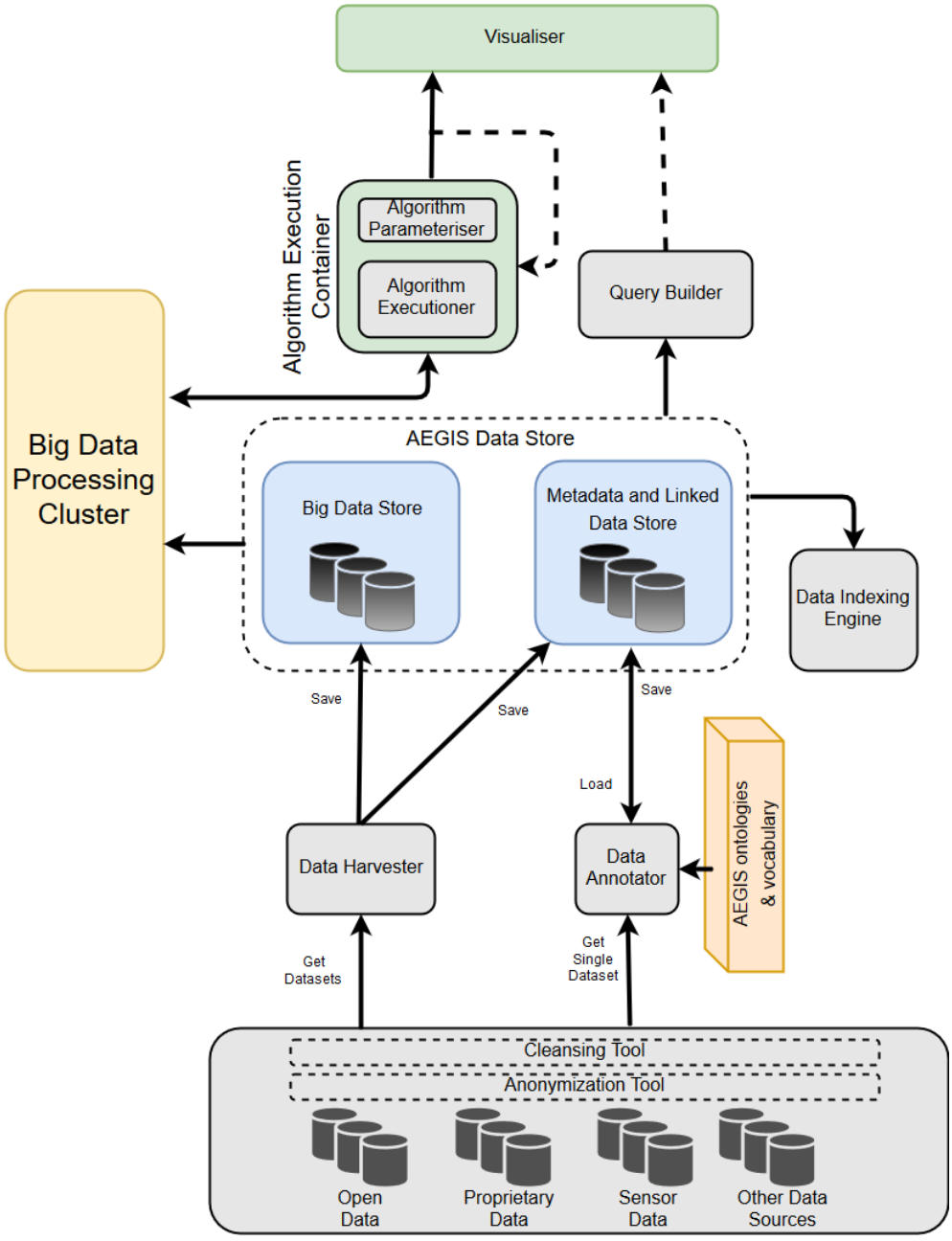



**Figure 4-1: AEGIS high level architecture**

These metadata will be presented in RDF format. RDF is a W3C recommended standard originally designed as a metadata data model. It evolved as a general method for conceptual description or modelling of information that is implemented in web resources, which can support semantics-aware Business Intelligence using a variety of syntax notations and data serialization formats. The idea of RDF is to represent everything in terms of triples, which are basically the data model of RDF which in turn has three components: 1) the Subject (S), 2) the

Predicate (P), and 3) the Object (O). The output of the Data Annotator is the enriched metadata or in some case enriched data if the harvested data was stored in AEGIS in the liked data format.

One of the core components of the AEGIS is the Data Store. The Data Store is the heart of the AEGIS since it provides the storage capabilities of the platform. It consists of (at least at the logical level) the Big Data Store containing all the large datasets (Big Data), and the Metadata and Linked Data Store along containing the RDF triples of metadata and data. Moreover, it is essential that a process for uploading large datasets in a timely and efficient manner is supported. The internal architecture of the Data Store will be decided in the next steps, as several requirements will be taken into account before finalizing the implementation. The storage solution should be capable to maintain large amount of unrelated complex data sources, offering the scalability, high availability and flexibility along with powerful querying and indexing capability needed in Big Data solutions.

The Data Indexing Engine in turn utilizes all the essential information which has been extracted from the primitive source of data and has been semantically annotated, as provided by the Data Harvester and Data Annotator during the previous steps, in order to achieve highly efficient near real-time indexing and advanced searching. In particular, the datasets are stored to a persistent location as a specific structural format of data, including all the indexes that have been defined during this process. After this process, the data is now in a form upon which subject indexing can be achieved.

The Query Builder is a graphic tool that can be used to create simple or complex queries in a user-friendly way. It will facilitate the query building procedure with a simple and easy-to-use user interface even for the complicated queries allowing the user to choose from multiple data sources and apply filters with less effort. The results of the query execution will be presented with pagination and additional filters could be applied on the results as selected by the user. The results of the execution will also be the input of the Visualizer.

Visualizer is the process that provides visualization capabilities on top of the content provided by the Algorithm Execution Container or the results of the query composed and executed by the Query Builder. It provides a variety of bar, line and scatter plots, charts, tables, and maps. Also, the visualizer will provide the ability to the user to quickly create and share flexible, dynamic dashboards.

The Algorithm Execution Container is the process where selected or requested algorithms are executed. It consists of two processes, the Algorithm Parameteriser and the Algorithm Executioner. The Algorithm Parameteriser is a small process only responsible for providing the parameter values of the algorithm to be executed, when applicable, as selected by the user to the main process, which is the Algorithm Executioner. The Algorithm Executioner is responsible for the initialization and monitoring of the execution of the selected algorithm, which includes communicating the Big Data Processing Cluster to initiate the execution and waiting for the results of the execution that will be later be provided to Visualizer.

For the data processing part of the AEGIS platform, a cluster-computing framework is needed. Big Data Processing Cluster should be powered by a processing engine built around speed, efficiency, ease of use and should be suitable for analytics execution purposes. Ideally, the framework should support clustering with data parallelism and fault-tolerance. For the AEGIS purposes, the cluster-computing framework should support data analysis in an interactive and

exploratory way supplemented with an easy-to-use interface for programming, facilitating the scheduling and execution of the AEGIS processing jobs. Along with cluster-computing framework, a cluster management technology is required. It is essential that it will provide efficient resource management and a centralized platform that will deliver consistent operations, security, and data governance tools across the clusters of the Big Data Processing Cluster. This operation should take advantage of cost effective, linear-scale storage and processing offered by the cluster-computing framework already described.

**Table 4-1: Mapping Technical Requirements to Components**

| ID | Need | Priority | Component |
|---|---|---|---|
| TR1 | Process measurements from "sensor nodes" | High | Data Harvester, Data Annotator, Algorithm Execution Container, Data Store |
| TR2 | Process measurements from social media sources | High | Data Harvester, Data Annotator, Algorithm Execution Container, Data Store |
| TR3 | Process data from mobile devices | High | Data Harvester, Data Annotator, Algorithm Execution Container, Data Store |
| TR4 | Process data from other open sources | High | Data Harvester, Data Annotator, Algorithm Execution Container, Data Store |
| TR5 | Process data from relational databases | High | Data Harvester, Data Annotator, Algorithm Execution Container, Data Store |
| TR6 | Process data from NoSQL databases | High | Data Harvester, Data Annotator, Algorithm Execution Container, Data Store |
| TR7 | Process data from structured sources | High | Data Harvester, Data Annotator, Algorithm Execution Container, Data Store |
| TR8 | Process data from semi-structured sources | High | Data Harvester, Data Annotator, Algorithm Execution Container, Data Store |
| TR9 | Connect to data sources through well-defined APIs. | High | Data Harvester, Data Annotator, Data Store |

| | | | |
|---|---|---|---|
| TR10 | Configure data sampling / update period | High | Data Harvester |
| TR11 | Store big data | High | Data Harvester, Data Annotator, Data Store |
| TR12 | Produce RDF Triples | High | Data Harvester, Data Annotator |
| TR13 | Store RDF Triples | High | Data Store |
| TR14 | Handle big data scalability | High | Data Store, Query Builder, Algorithm Execution Container, Big Data Processing Cluster |
| TR15 | Ensure data availability | High | Data Store |
| TR16 | Ensure data persistence | High | Data Store |
| TR17 | Ensure data consistency | High | Data Store, Data Harvester, Data Annotator |
| TR18 | Offer distributed storage | High | Data Store |
| TR19 | Offer public storage | High | Data Store |
| TR20 | Offer private storage | High | Data Store |
| TR21 | Provide ability to users to securely upload datasets | High | Data Store, Data Harvester, Data Annotator |
| TR22 | Provide ability to declare IPR on uploaded datasets and set access restrictions | High | Brokerage Engine |
| TR23 | Access data | High | Data Store, Query Builder |
| TR24 | Select data | High | Data Store |
| TR25 | Implement a data management policy | High | Data Store, Brokerage Engine |
| TR26 | Support authentication mechanisms for access to data | High | Data Store |
| TR27 | Secure datasets | High | Data Store |
| TR28 | Clean datasets | Medium | Data Harvester, Data Annotator |
| TR29 | Anonymise datasets | Medium | Anonymization Tool |
| TR30 | Support interlinking of AEGIS triplestore with SLOD | Medium | Data Store |
| TR31 | Provide SPARQL Endpoints | High | Data Store |
| TR32 | Support conversion of multiple source formats to RDF | High | Data Harvester, Data Annotator |

| | | | |
|---|---|---|---|
| TR33 | Support conversion of multiple source formats to multiple formats | Medium | Data Harvester, Data Annotator |
| TR34 | Support interlinking of datasets from different sources. | High | Data Harvester, Data Annotator |
| TR35 | Support the reuse of available canonical standard vocabularies and ontologies. | High | Data Harvester, Data Annotator |
| TR36 | Deal with corrupted, inconsistent data or misconfiguration of mapping rules | Medium | Data Harvester, Data Annotator |
| TR37 | Support management of data sources | High | Data Harvester, Data Annotator, Data Store |
| TR38 | Add metadata, enriching the existing data sources | Medium | Data Harvester, Data Annotator |
| TR39 | Select / Search for / Define vocabularies | Medium | Data Annotator |
| TR40 | Inspect data sources | Medium | Data Harvester, Data Annotator |
| TR41 | Ability to handle queries spanning multiple datasets | High | Algorithm Execution Container, Query Builder |
| TR42 | Support format conversion | High | Data Harvester |
| TR43 | Manually balance transformation automation and semantic soundness | Low | Data Harvester, Data Annotator |
| TR44 | Support data analytics on distributed big data | High | Algorithm Execution Container, Query Builder, Big Data Processing Cluster |
| TR45 | Support data analytics on distributed big data | High | Algorithm Execution Container, Big Data Processing Cluster |
| TR46 | Support initialization of algorithms | High | Algorithm Execution Container |
| TR47 | Support the processing of data in multiple formats. | High | Algorithm Execution Container |
| TR48 | Support a set of different types of output formats. | High | Algorithm Execution Container |

| | | | |
|---|---|---|---|
| TR49 | Customise the analytics process | High | Algorithm Execution Container |
| TR50 | Support real-time analysis | | Algorithm Execution Container |
| TR51 | Manage the results of the big data enabled analytics | High | Algorithm Execution Container, Visualizer |
| TR52 | Support overview of resources | High | Data Management, Visualizer |
| TR53 | Preview a small selection of the results of the generated query | Medium | Visualizer |
| TR54 | Support a set of different types of visualizations | High | Visualizer |
| TR55 | Create advanced graphs based upon queries spanning multiple datasets | High | Visualizer |
| TR56 | Support secure communication among the AEGIS components | High | Data Harvester, Data Annotator, Data Indexing Engine, Algorithm Execution Container, Query Builder, Visualizer, Big Data Processing Cluster, Brokerage Engine, Data Store |
| TR57 | Save analytics results | High | Visualizer, Data Store |
| TR58 | Support authorized sharing of saved results | Medium | Visualizer, Brokerage Engine |
| TR59 | Support local deployment of the anonymization tool | Low | Anonymization Tool |
| TR60 | Support local deployment of the cleansing tool | Low | Cleansing Tool |
| TR61 | Support local deployment of the harvesting tool | Low | Data Harvester |
| TR62 | Support local deployment of the transformation tool | Low | Data Harvester |
| TR63 | Provide push notifications | High | Data Harvester |
| TR64 | Provide secure user-friendly interface and flexible navigation. | High | Visualizer, Query Builder, Data Harvester, Data Annotator |
| TR65 | Provide system management dashboard | High | AEGIS Web GUI |

| | | | |
|---|---|---|---|
| TR66 | Support geospatial data analysis | High | Algorithm Execution Container |
| TR67 | Provide a solid deployment workflow of the system | Medium | |
| TR68 | Support uploading user's custom algorithm implementation | Low | Algorithm Execution Container |

## 4.3. AEGIS COMPONENTS SPECIFICATIONS

### *4.3.1. Data Harvester*

Data Harvester is the component enabling the import of data and metadata from heterogeneous sources and their transformation to the required data format/structure.

The ability to work with different data sources, types and formats is one of the key requirements to AEGIS directly affecting its exploitation potential. Data Harvester has to be able to import data as well as metadata. The main types of data sources are:

- APIs providing access to static data or to dynamic data streams
- Individual files of heterogeneous type and structure
- Databases

For accessing each of the data sources, a dedicated connector has to be implemented and/or configured. In some cases, the connectors can be re-used if several data sources have the same API.

The (meta-)data harvested from different sources or even from the same source may have different structure. In AEGIS, all metadata has to be transformed to linked data format and then refined with the help of the Annotator component explained in the next section. The data may be transformed to Linked Data or to another format or kept in the original structure. The transformation of the (meta-)data is a part of the harvesting process. It is done by the Transformer component, which applies predefined transformation rules to the (meta-)data collected by the Harvester. The transformed data then are stored in the Data Store. The transformation rules are the source and target data format specific and need to be defined and properly managed.

Data Harvester needs to provide a possibility to plan harvesting tasks for their regular execution.

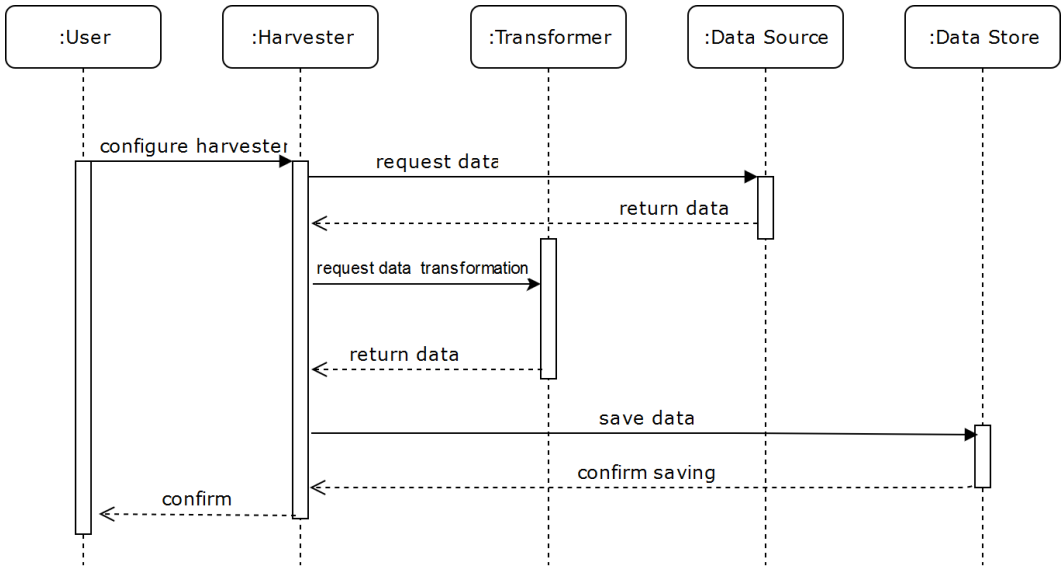



**Figure 4-2: Data Harvester**

*4.3.2. Data Annotator and Vocabularies*

The annotator is the component in the AEGIS platform that is responsible for interactively equipping input data with suitable metadata. What follows is the properties the annotator and the AEGIS ontologies should have, and the basic concepts for them.

4.3.2.1. Annotator concept

The annotator is an interactive metadata editor with a graphical user interface. It is capable of guessing various data formats, and presents its guesses to the user for further refinement. The annotator has a built-in knowledge of the AEGIS ontologies. In particular, the annotator "knows" what metadata need to exist for what kind of input data, and guides the user through the metadata generation process accordingly. Additionally, it knows to which installation of the AEGIS platform it belongs, and hence knows without user assistance where to store its output.

4.3.2.2. Requirements to Annotator

For the input data, the following aspects hold:

- The input data may come in a variety of structures and formats, e.g., very frequently as tabular data.
- The input data may or may not already be equipped with some metadata, and these may potentially be incomplete or incorrect and may thus require complementation or correction.
- Furthermore, the input data may come in various sizes; this means in particular that it can be Big Data or non-Big Data.

For the output data, there are two options to select from:

- The output consists of the original input Big Data, or even only a reference to it (URL), and the annotations (metadata). The latter are produced in RDF. These two parts of the output will be stored in different dedicated stores, e.g., in a Data Store and in a triple store, respectively. – For Big Data input, this is the only option available.
- The data and its metadata are both converted to Linked Data and are delivered as joint, semantically self-describing data. Linked Data is a data format where data plus its metadata are both represented in RDF, using appropriate ontologies.

#### 4.3.2.3. Requirements to Annotations and Vocabularies

The purpose of the metadata is to enable the AEGIS platforms and its users to do their work. This comprises typical analytics and Big Data operations (such as correlations, visualisations, searches, text analyses, graph operations and others). It also comprises exploratory analyses to detect unforeseen relationships, in an automated fashion whenever possible.

This has not self-evident implications for the ontologies devised for AEGIS. These implications originate from the design objective that the collected data will be subject to Big Data evaluations. Hence, the AEGIS ontologies need to provide means to express those meta information that are relevant for subsequent typical Big Data operations.

A relevant example of such operations is correlations. To enable these, the domain and the range of a dataset should be expressible. Other Big Data operations will require their own specific information.

Moreover, not all data are converted to the (self-describing) Linked Data format; instead, Big Data input is retained as it is. In order to enable automated data accesses in that case, the metadata must be capable of stating technical format aspects, like column numbers, byte offsets, floating point formats and the like.

#### 4.3.2.4. Potentials of possible automation

The process of incorporating metadata that were already found in the harvesting process can certainly be automated. On the other hand, the process of automatically making semantics of data explicit (making semantic annotations) is very difficult. Examples of this problem are:

- There is a time series of numeric sensor data. Appropriate ontologies (and descriptive properties therein) cannot be derived from the data, let alone automatically. It needs to be known what kind of sensor is involved, what machine it is a part of, etc.
- There is tabular (or relational) data with a table name and with column headers. The procedure of just looking up words (proper names; column headers) in a repository of descriptions can certainly be automated to a large extent. However, the relevance and semantic accuracy of the results is more than questionable.
- There is a data column containing proper names. Annotating them with appropriate URIs is hindered by the potential ambiguity of the proper names and by difficulty of finding canonical knowledge bases.

In summary, the potential of automation of generating valid metadata is very limited. In those cases where it seems to be feasible, one must be prepared for considerable imprecision of the result.

4.3.2.5. Possible Extensions

There are some concepts for later extensions:

- **Data curation**. The annotation process may be scripted, so that it can be repeated many times with new versions of the data in question. The scripts may automatically be checked for inconsistencies every time, so as to keep it up-to-date with changing formats and contents.
- **NLP**. Applying NLP stacks may also be subsumed under automated annotation processes. No surprise, that would employ vocabularies specific for NLP, and, due to the intrinsic properties of natural language, the results would not be as accurate and reliable as computer scientist are accustomed to. (This does not make them useless; one just needs to be careful.)

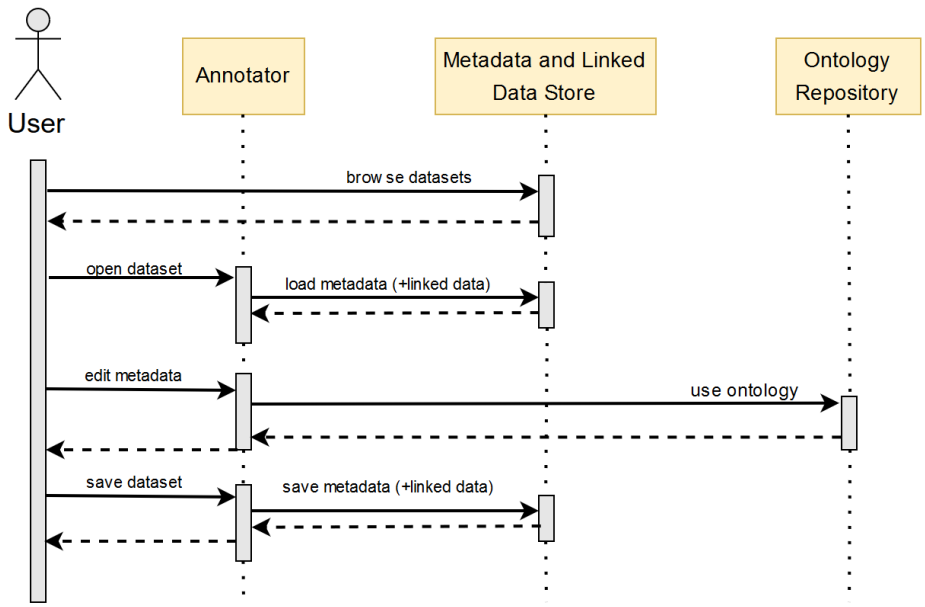



**Figure 4-3: Data Annotator**

*4.3.3. Algorithm Execution Container*

The Algorithm Execution Container is the component where the data analysis algorithm selected by the user is executed. This component can be viewed as the facilitator of the execution of analysis over multiple data, originating from different sources and already stored in the Data Store, in order to provide the results to satisfy the stakeholders' needs. The Algorithm Execution Container contains all necessary sub-components to facilitate the execution of all steps needed for the complete execution flow of a data analysis algorithm over selected datasets in order to provide the results in the Visualizer component.

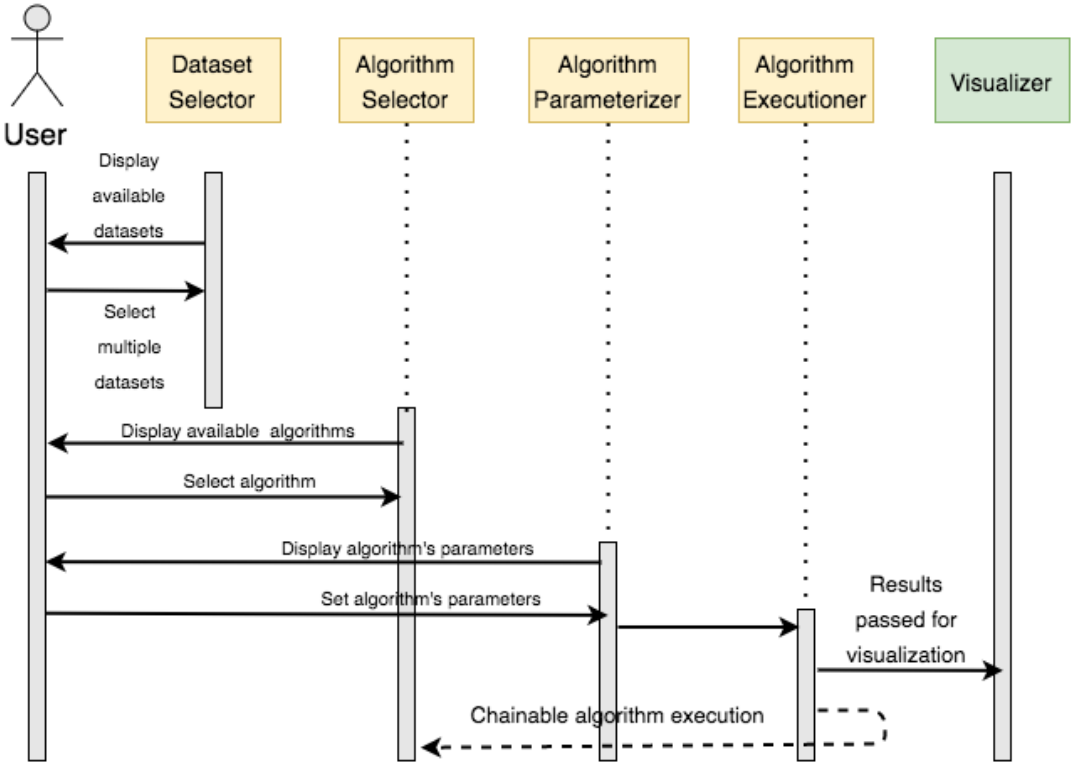
Within the context of AEGIS, the Algorithm Execution Container is the cornerstone of the AEGIS platform, as it is highly crucial that the algorithm's execution will deliver the business

intelligence that addresses the needs of the stakeholders across the Public Safety and Personal Security value chain. It is critical that a plethora of robust, big-data enabled analysis algorithms is already implemented and offered to the users. It is also crucial that those algorithms can be initialised / parameterized by the requestor in order to personalise the analysis to the widest possible extent. Depending on the algorithm the corresponding parameters of the algorithms will be offered in predefined default values and the user will be able to configure the parameter values upon needs before initiating the execution.

Another critical aspect of the Algorithm Execution Container is the ability to support various kinds of analysis spanning across multiple datasets that are already available through the AEGIS platform. Prior to the execution phase, the user will be able to select one or more of the available datasets. The Algorithm Execution Container will offer the ability to save the user's selection (datasets, algorithm with the preferable parameter values) for easy re-execution in the future. Also within the Algorithm Execution Container the user will be able to request for the execution of an algorithm implemented by the user itself. This user will be provided with an easy way to upload his code which implements the custom algorithm and request for a review and approval from the AEGIS moderators. The user will be informed of the rules and restrictions that will be applied by the AEGIS moderators so that his implementation is fully compatible. Once the AEGIS approve his custom algorithm, the user will be able to select the uploaded custom algorithm and execute it just like the rest predefined algorithms.

Once the preferable datasets have been identified and the algorithm has been selected along with the preferable parameter values, the execution can be initiated immediately upon user's request or can be scheduled for future time. The execution will result in a request in the Big Data Processing Cluster framework which will handle the actual execution of the selected algorithm over the selected data and will provide the execution results back to the Algorithm Execution Container. The results will be provided then to the Visualizer component which will provide visualization capabilities over the results. Also, it shall be possible that the results can be used to initiate a new algorithm execution on top of them, implementing a chainable execution of algorithms.

The following figure displays the sequence diagram for the Algorithm Execution Container.




Metadat

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

*4.3.4. Query Builder*

The Query Builder is the component that provides the capability to interactively define and execute queries on data available in the AEGIS system. The purpose of the Query Builder is to provide AEGIS users, especially those with limited technical background, with the potential to explore AEGIS data without a requirement to learn a query language or understand the underlying structure of data stored in the system.

Queries defined using the Query Builder are the main "view" of AEGIS data for the average user. The output of the Query Builder, effectively a view on AEGIS data, will be the input for most high level AEGIS tools, such as the Visualizer and the Algorithm Execution Container. As a result, the following objectives arise for this tool:

- Allow users to create complex queries from a simple, easy-to-use GUI, without imposing significant limitations compared to using a query language directly.
- Use state-of-the-art techniques in order to cope with the volume of the underlying big, potentially real-time data.
- Depend on the PSPS AEGIS ontology in order to present "smart" recommendations during query building.
- Become schema and database agnostic, thus minimizing developer intervention when changes occur in the structure of AEGIS data.
- Build upon common user interface techniques used by other query building tools such as Oracle Query Builder, MySQL Query Builder and LinDA Query Designer, in order to hide the underlying complexity from the end user.

- Exploit common patterns in order to employ generally accepted optimisations, in order to cope with the volume of AEGIS data.

The queries defined can be saved, executed, retrieved and updated, while users can also share queries with others. The visual representation of queries in AEGIS means that new users have a very low entry barrier, and can almost immediately understand the structure of the underlying data, easily reusing existing and defining new queries.

The high level workflow followed in the Query Builder is:

a) Datasets available to the user are presented with some high level information about the available data
b) Each dataset will break down into several "Entities" (models, classes or variables, depending on the context), each of them containing homogeneous data.
c) Various relations between entities will be presented and proposed in order to link data.
d) User will pick the entities they want to obtain and link them using the proposed relations.
e) Filters, aggregates and transformations will be imposed on the attributes of the selected entities
f) Once the query is defined it can be executed, after which its results will be previewed and available for further use in the system.

The following sequence diagram shows the aforementioned steps:

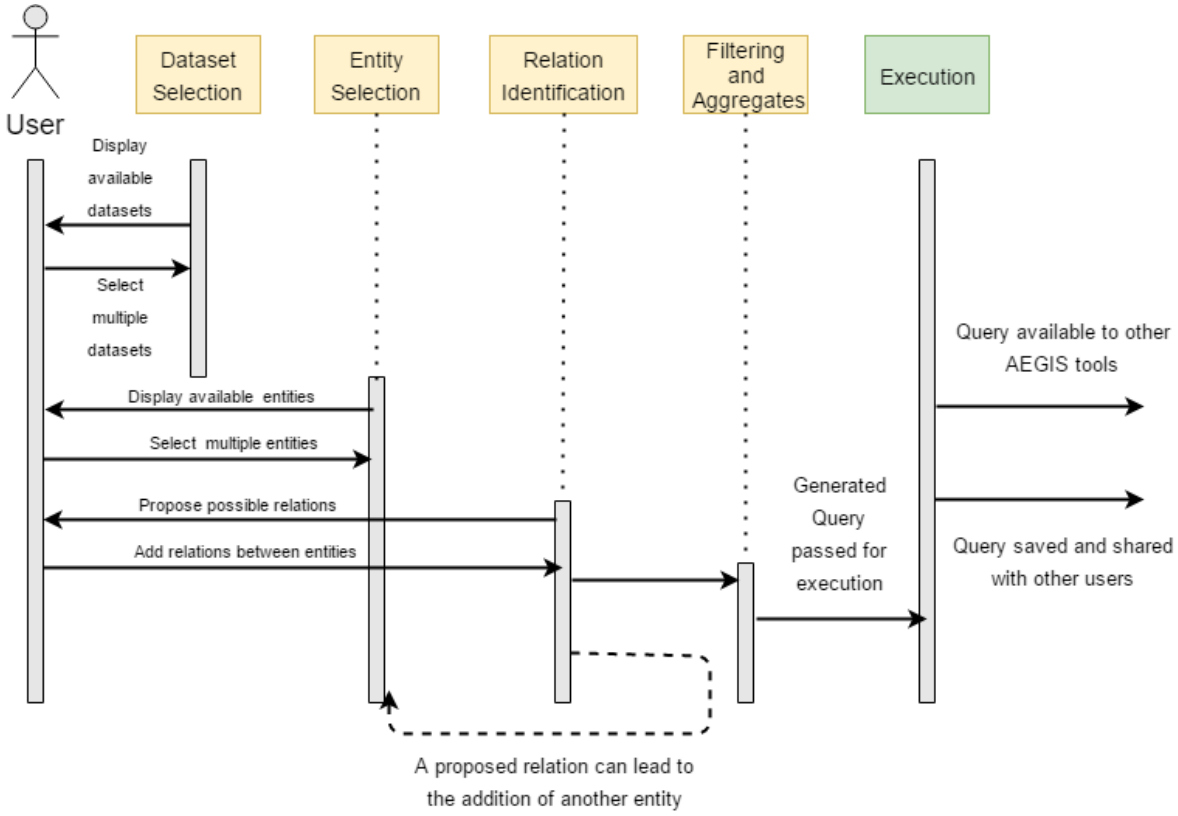



**Figure 4-5: Query Builder**

### *4.3.5. Visualizer*

The visualizer is the component that provides visualisation capabilities for the output of (1) the analysis results coming from the Algorithm Execution Container component and (2) the results of the queries composed and executed by the Query Builder. In order to support the heterogeneous data in the AEGIS system, the visualizer provides a variety of stock visualisation formats, including bar, line and scatter plots, charts, tables, and maps, easily selectable and applicable on the data through the user interface.

In order to facilitate the selection of the most appropriate visualisation, the component offers automatically generated suggestions that leverage the underlying common AEGIS schemas to infer, when possible, the semantics of the data to be visualised. Furthermore, in the case of geo-referred information, which is very common in PSPS-related datasets, the visualizer offers various pre-configured maps that allow the user to add new layers of information coming from both the algorithm and query execution results, thus creating rich visualisations. The provided gallery of visualisations spans from simple static charts to browsable interactive charts that seamlessly interact with the back-end components to update and refine the visualised information according to the user's actions (e.g. progressively drilling down to a map and performing new computations to show the required fine-grained data). For the data-savvy users, visualizer also offers certain configurable, to an extent, visualisation options, enabling the users to build custom solutions.

All visualisations can be saved independently, but can also be embedded in flexible dashboards that allow the user to select, position and organise in the same interface multiple visualisations to be accessed together. The visualizer may, upon user request, generate a link for each dashboard so that it can be easily accessed independently. Finally, the visualizer allows the user to publish dashboards as AEGIS services to be consumed by other stakeholders.

The following figure displays the sequence diagram for the Visualizer component.

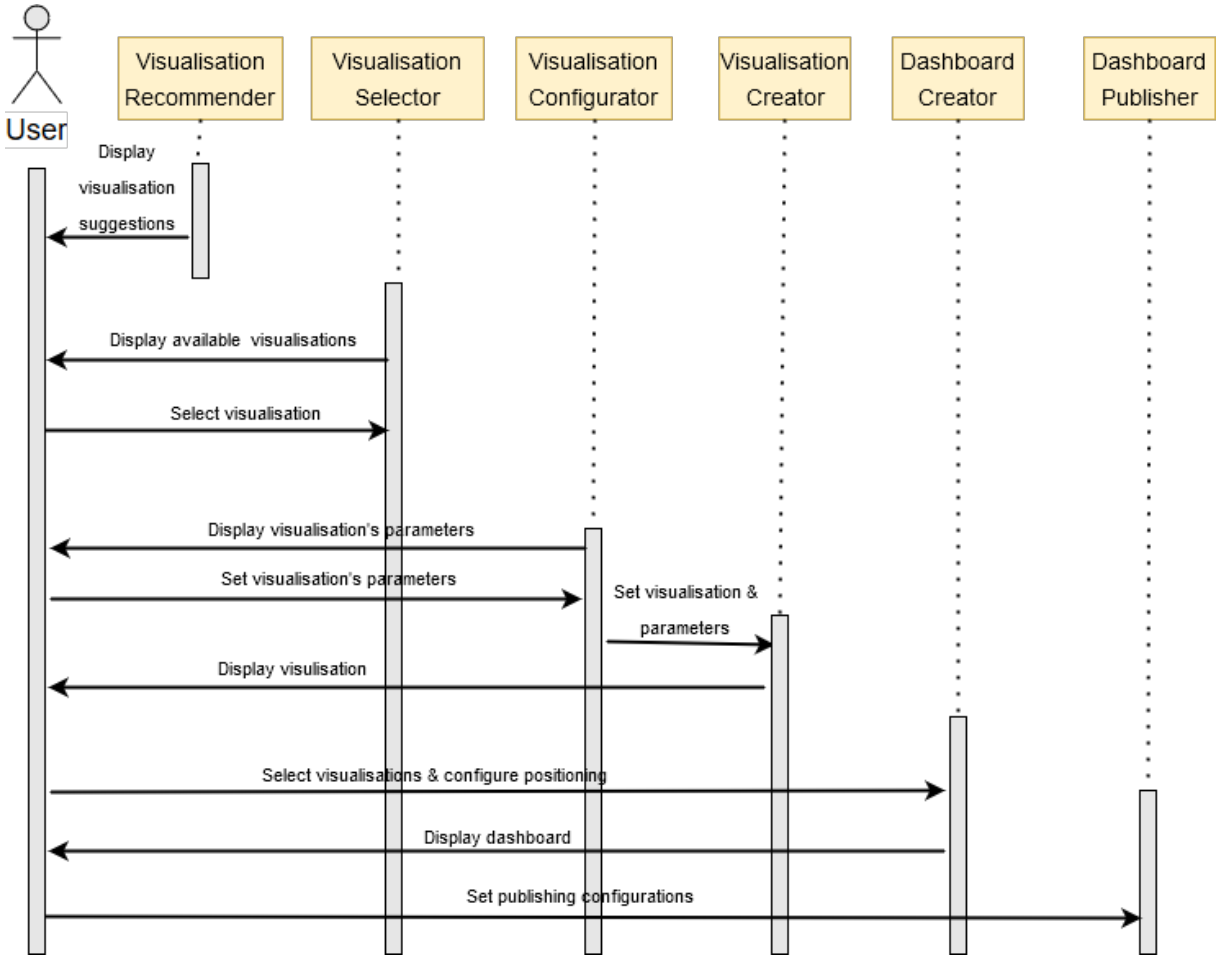



**Figure 4-6: Visualizer**

### *4.3.6. Big Data Processing Cluster*

The Big Data Processing Cluster component is responsible for storing and processing of the Big Data files. The core component of the cluster is the distributed file system. The distributed file system stores the data in a distributed fault-tolerant fashion to ensure the availability of the files. It should be highly performant to cope with thousands of concurrent requests from users, and should be scalable to support a large number of files, potentially billions. The distributed file system serve as the data storage.

The Big Data Processing Cluster manages execution of distributed applications on the cluster. It has a central authority which collects usage information from all nodes in the cluster, that is then used to correctly assign nodes to run a specific job/application. When a user wants to run a job on the cluster, the cluster allocates some resources in the cluster to run the application master for the submitted job. Then, the application master is responsible for running the job on the cluster, requesting further resources (CPU, memory) as required by the job, and monitor the status of the job.

On Top of cluster management, the Big Data Processing Cluster provides several frameworks for Big Data processing. These frameworks, should provide high-level APIs in variety of

languages such as Java, Scala, Python, and R. Also, a special-purpose tools for machine learning, graph processing, and streaming are highly beneficial.

In AEGIS, once the preferable algorithm and datasets have been selected the Algorithm Executioner schedules a job for the algorithm on the Big Data Processing Cluster as shown in Figure 4-7. Then, the Big Data Processing Cluster starts an application master for the job, which in turns requests resources from the Big Data Processing Cluster as it needs to execute the requested algorithm. Once everything is set up, the algorithm starts to read and/or write data to the Data Store according to the nature of the algorithm. When the algorithm is finished, the results are sent back to the Algorithm Executioner.

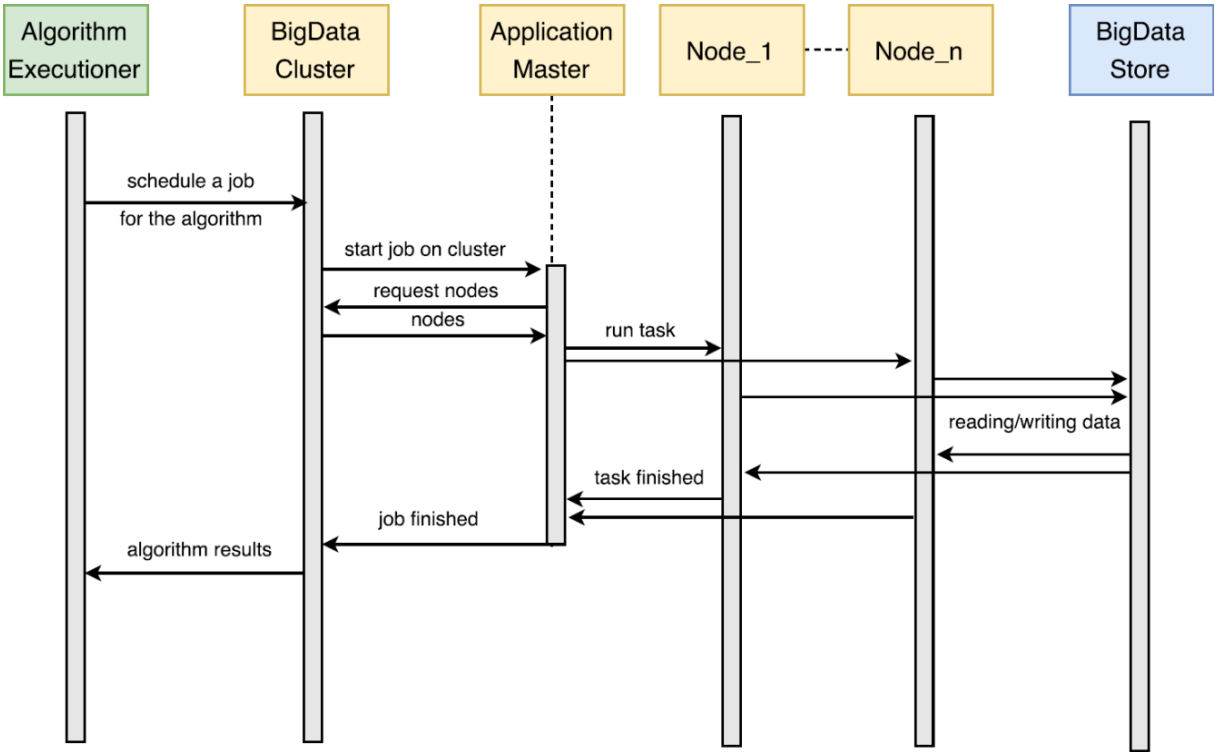



**Figure 4-7: Big Data Processing Cluster**

*4.3.7. Brokerage Engine*

The AEGIS Brokerage Engine will be the component that will instantiate part of the methods that will be included in the AEGIS Data Policy and Business Brokerage Frameworks, which are delivered under WP2. Although the title of these Framework points to a data-focused brokerage and business value generation methodology, the overall frameworks, as well as the Brokerage Engine, will support also the brokerage of other artefacts that can be designed, developed, shared and used over the AEGIS infrastructure, such as (micro-)services, algorithms, visualisations, etc. Thus, in the Brokerage Engine we refer to "Artefacts" with those being Datasets, Services or Algorithms.

The final aim of this component is to serve the AEGIS infrastructure with an endpoint that is able to create micro-contracts for artefact sharing, managing IPRs, quality and privacy issues as well.

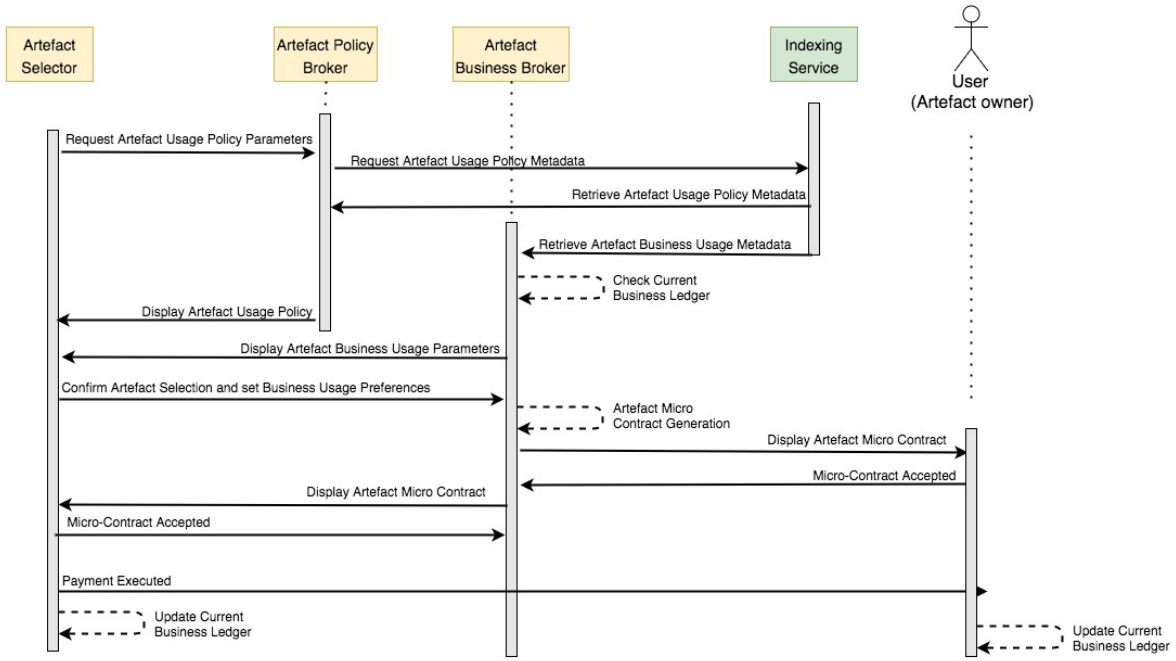



**Figure 4-8: Brokerage Engine**

As shown in the figure above, upon the pre-selection of an Artefact by another component (such as the "Algorithm Execution Container") which we call here "Artefact Selector", the Brokerage Engine engages initially with the Artefact Broker that is responsible of handing the incoming request in terms of retrieving both the policy usage (e.g. IPRs and other artefact related metadata), as well as the Business Usage parameters of the artefact under request. As part of this process, the artefact's metadata are requested from the Indexing Service and are packaged together ready to be served to the requested component. As a last step before serving those to the requestor, an internal check of the AEGIS ledger is performed, to identify whether previous transactions regarding this artefact impose a change in the artefact's general business usage (for example if this has been purchased for sole usage by another entity).

Upon serving all the information back to the requestor, the latter is able to confirm his interest and set his Business Usage Preferences. The Business Broker then generates a micro contract, which is sent to both the requestor and the artefacts owner. Upon acceptance by both sides and execution of the payment, the AEGIS ledger is updated with this information.

It needs to be noted that the steps of micro-contract creation, acceptance/rejection of it and of the payment are subject to the usage and business policies of each artefact, meaning that these steps might be automatically executed (and in certain cases for free if these refer to open source/open access artefacts), or that they would need the intervention of users).

### *4.3.8. Holistic Security Approach*

In addition to the aforementioned components, the AEGIS conceptual architecture incorporates a holistic security approach. This security approach which will be more thoroughly analysed in the future versions of the current document integrates security aspects throughout the whole

lifecycle of the data exploitation, from its check in to the AEGIS platform and its storage, to its usage and exploitation during query time.

Towards this end, the AEGIS platform will safeguard security of data in storage, in transit and in use.

- Security of data in storage is the first of the three parts of the data lifecycle security and includes all raw data and metadata stored on the platform infrastructure. Data in storage security refers to the preservation of the security, privacy and integrity of data that is stored physically in any digital form. It deals with any type of security around the storage architecture and the data stored on it. This is of particular value to the private repositories envisaged in the context of AEGIS, but applies also in the case of public storage since these data may become corrupted and be of limited value to the AEGIS stakeholders. The AEGIS technical partners have identified a number of candidate technologies for securing data at storage including Symmetric Encryption Algorithms, Asymmetric Encryption algorithms as well as Attribute-Based Encryption.
- Security of data in transit, or data in motion, is the second part of the data lifecycle security and includes data actively moving from one location to another such as across the internet or through a private network. Data protection in transit is the protection of this data while it's traveling from network to network or being transferred from a local storage device to a cloud storage device – wherever data is moving, effective data protection measures for in transit data are critical as data is often considered less secure while in motion. The AEGIS technical partners have identified a number of candidate technologies for securing data at storage including IPsec and TLS.
- Security of "Data in Use" is the third and last part of the data lifecycle security and includes all data not in an at-rest state, which is kept only one particular node in a network (for example, in resident memory, or swap, or processor cache or disk cache, etc. memory). This data can be regarded as "secure" if and only if (a) access to the memory is rigorously controlled, and (b) regardless of how the process terminates (either by successful completion, or killing of the process, or shutdown of the computer), the data cannot be retrieved from any location other than the original at rest state, requiring re-authorization. The AEGIS technical partners have identified a number of candidate technologies for securing data at storage including Homomorphic Encryption which allows the chaining together of different services without exposing the data to each of those services, and Verifiable Computation which enables a computer to offload the computation of some function, to other perhaps untrusted clients, while maintaining verifiable results, functionalities very important to AEGIS when criticality of privacy preservation is of increased importance and data needs to be transferred across and used by a plethora of different components, guaranteeing that all of this components are trusted and non-tampering with sensitive information is safeguarded.

Apart from the three parts of the data lifecycle security aforementioned, the AEGIS consortium has identified two additional security aspects which will be handled appropriately throughout the platform implementation.

- Security of technical interfaces (e.g. REST) amongst the various AEGIS components. This security aspect refers to the minimisation of the risk associated with the exploitation of the operation of various AEGIS components from external malicious

components. Several ways to secure APIs can be encountered in the literature, including for example basic authentication w/ TLS, utilisation of OAuth protocol, etc.

- Data Access Control, which in general includes authorization, authentication, access approval, and audit. Authentication and access control are often combined into a single operation, so that access is approved based on successful authentication, or based on an access token. Authentication methods and tokens include passwords, biometric scans, physical keys, electronic keys and devices, and other means. The operation of the data access control will be tightly inter-related in the context of the project with the operation of the Brokerage Engine, which, as aforementioned, will instantiate part of the methods that will be included in the AEGIS Data Policy and Business Brokerage Frameworks.

# 5. CONCLUSION

The objective of this deliverable was to deliver the user requirements along with the technical requirements of the AEGIS and also to deliver the conceptual architecture of the platform. More specifically, in order to collect and analyse the user requirements the agile software development extraction techniques were adopted. At first the actors of the AEGIS platform were identified and the principle of user stories was used. From these user stories the functional and non-functional requirements were extracted and collected by the technical partners and the pilots. In the next step these collected requirements were analysed and validated in a specific workshop in order to obtain the concrete list of AEGIS user requirements.

This list of user requirements has supported the compilation of the list of technical requirements of the AEGIS platform. The collection and construction of the technical requirements is of particularly importance for the AEGIS, since those requirements will drive the architectural decisions that were (for the preliminary version) and will be made for the AEGIS platform design. So, in this deliverable a complete requirement backlog has been provided for AEGIS. After performing a comprehensive analysis of the technical requirements, the conceptual architectural design was designed.

The scope of the architecture is to reach the goals and expectations of the users and the stakeholders. In this context, the key components of the AEGIS platform and their specific functionalities and interactions have been described. Each component was carefully described having in mind that it should address a specific set of technical requirements from the list of technical requirements extracted in the previous step. The complete architecture was compiled by the list of components making sure that every requirement was addressed towards the required AEGIS complete required functionality.

In the next steps the platform architecture compiled in this deliverable along with the list of components specified will be utilized in order to design the components, the different micro-services that will drive the AEGIS execution flow along with the appropriate API interfaces that will be implemented. It should be stressed at this point that the current deliverable comprises the preliminary version of the technical and user requirements and also delivers the conceptual architecture of the AEGIS platform. The forthcoming versions of this deliverable comprise updated iterations and will include updates on the platform and components' architecture, as well as the APIs interfaces, based on the feedback received by the project's demonstrators. Thus, the future versions of this deliverable will include the more detailed technical architecture of the platform, updating the requirements backlogs provided in the annexes, as well as additional information including for example where all data sources are located, where the processing power resides, what analytics are supported and what services are delivered to the users etc. in full detail.

## APPENDIX A: STAKEHOLDERS & LITERATURE REQUIREMENTS BACKLOG

| Id | Non-functional requirement type | Description |
|---|---|---|
| NFR_TL1 | Customizable stakeholder tool | A customized and optimized tool for a particular field of use. |
| NFR_TL2 | Manage account | Create/delete/modify accounts (stakeholder account, customer account, operator account…). |
| NFR_TL3 | User profiling | Manage groups and roles for the definition of user's functions and permissions. |
| NFR_TL4 | Authentication and authorization | Managing access to the AEGIS tool/platform (local and remote). |
| NFR_TL5 | Customizable workspace | The user can organize his work environment as desired. |
| NFR_TL6 | Multilingual workspace | The user can change the language of their work environment. |
| NFR_TL7 | Local client tool | A client tool which can also work offline (with reduced functionality). |
| NFR_TL8 | Online browser tool | User's workspace can be run on a simple web browser. |
| NFR_TL9 | Mobile app | User's workspace can be run on a mobile device. |

| Id | Functional Requirement | Description |
|---|---|---|
| FR_QR1 | Manage query | Create/delete/modify/duplicate a query, an object that represents a search, a view. |
| FR_QR2 | Online store query | A query can be stored on AEGIS' cloud. |
| FR_QR3 | Local Import/export query | A query can be stored locally on the user's terminal. A compressed and password-protected file is generated. |
| FR_TK1 | Manage task | Create/delete/modify/duplicate task, a programmable object based on an expected result (an environment, overcoming a threshold…). |
| FR_TK2 | Online store task | A task can be stored on AEGIS' cloud. |

| | | |
|---|---|---|
| FR_TK3 | Local import/export task | A task can be stored locally on the user's terminal. A compressed and password-protected file is generated. |
| FR_TK4 | Schedule task | Execution of a task can be timed. |
| FR_TK5 | Enable/disable task | Timed or triggered task can be activated or deactivated. |
| FR_RL1 | Manage semantic rule | Create/delete/modify/duplicate semantic rule, an object that represents a given meaning to one or more data sets. |
| FR_RL2 | Online store semantic rule | A semantic rule can be stored on AEGIS cloud. |
| FR_RL3 | Local Import/ Export semantic rule | A semantic rule can be stored locally on the user's terminal. A compressed and password-protected file is generated. |
| FR_RT1 | Displays result | A result can be displayed in different ways. |
| FR_RT2 | Local Import/export result | A result can be stored locally on the user's terminal. A compressed and password-protected file is generated. |
| FR_RT3 | Publish result | A result can be published on a social platform (open or close). |
| FR_RT4 | Buy/Sell result | A result can be sold or purchased among members of AEGIS community. |
| FR_RT5 | Mask result | It is possible to define different levels of visibility and details of a result (published or sold). |
| FR_RT6 | Search result | Searches can be made between results published or sold by the AEGIS community. |
| FR_RT7 | Vote result | It is possible to express a quantitative (vote) and/or qualitative (comment) judgment on a published result. |
| FR_RT8 | Translate result | A result can be translated in another language. |
| FR_RT9 | Manipulate result | It is possible to change a result by arbitrarily (not from a data source) changing the variables that produced it. This operation can be done offline. |
| FR_RT10 | Validate result | It is possible to validate a result using a different semantic rule. |
| FR_RT11 | Compare results | It is possible to compare two results that share the same query or the same semantic rule. |
| FR_RT12 | Historicise result | Inside the result can also be recorded the previous versions of the same. |

| | | |
|---|---|---|
| FR_RT12 | Static update result | It is possible to update a result by re-running the query contained in it. It is necessary to access the original data sources. |
| FR_RT14 | Dynamic update result | A result is automatically updated in real time. |
| FR_RT15 | Periodic update result | A result is automatically updated and periodically. |
| FR_DS1 | Manage data source | Create/delete/modify/duplicate a data source. |
| FR_DS2 | Search data source | Searches can be made between data sources shared by the AEGIS community. |
| FR_DS3 | Expose read-only data source | It is possible to share a data source within AEGIS community, cannot be changed. |
| FR_DS4 | Share data source | It is possible to share a data source within AEGIS community, can be changed. |
| FR_DS5 | Rent data source | It is possible to provide access to a data source for a limited time. Exclusively or not. |
| FR_DS6 | Mask data source | It is possible to define different levels of visibility and details of a data source (shared or rented). |
| FR_DS7 | Vote data source | It is possible to express a quantitative (vote) and/or qualitative (comment) judgment on a shared data source. |
| FR_DS8 | Book data source | It is possible book access to a data source. |
| FR_DS9 | Link data source | To run a query, some task or service, is necessary to link a data source to your environment. |
| FR_DS10 | Translate data source | A data source can be translated in another language. |
| FR_DS12 | Analyse data source | It is possible to extract generic information about a data source. |
| FR_SC1 | Manage service | Create/delete/modify/duplicate service, an object that providers derivate information from a result to generic users. |
| FR_SC2 | Expose service | It is possible to expose a service to the AEGIS community or to users outside it (generic citizens). |
| FR_SC3 | Rent service | It is possible to provide access to a service for a limited time. Exclusively or not. |

## APPENDIX B: FUNCTIONAL REQUIREMENTS BACKLOG

| Id | Description (Detailed description of the requirement) | References |
|---|---|---|
| Analytic | | |
| CFR17 | AEGIS has to implement an advanced clustering algorithm | GFT26 (HDI26) |
| CFR18 | AEGIS provides advanced analytics to compare different results | GFT28(HDI28, HDI29, HDI43, HDI55, HDI56) |
| CFR22 | AEGIS should be able to process daily routines as self-reported from users or automatically extracted by wearables | UBI6 (HYP9) |
| CFR23 | AEGIS should be able to process external weather data | UBI7 (HYP10) |
| CFR24 | AEGIS should be able to process Public Health Information data and statistics | UBI8 (HYP11) |
| CFR25 | AEGIS should be able to process energy (electricity) retailer prices data | UBI9 (HYP12) |
| CFR26 | AEGIS should be able to process CO2 emissions Footprint data | UBI10 (HYP13) |
| CFR27 | AEGIS should be able to process crime and accident raw data | UBI11 (HYP14) |
| CFR39 | AEGIS should allow uploading of user's data to be processed for example trips data | KTH2 (VIF2) |
| CFR4 | AEGIS has to support many analysis types (e.g. estimation of correlations between variables, linear regression, predictive analysis, clustering algorithms, simulations) | GFT4 (HDI3, HDI6) |
| CFR41 | AEGIS should be able to process sensor data (including environmental (temperature/humidity/luminance), occupancy sensing and Air Quality monitoring) from installed physical devices | UBI1 (HYP1, HYP2, HYP3) |
| CFR42 | AEGIS should be able to process data (energy consumption and set points) from in-house physically installed devices through interfaces | UBI2 (HYP4, HYP5) |
| CFR43 | AEGIS should be able to process positioning information and location data from mobile phones and wearables | UBI3 (HYP6) |
| CFR44 | AEGIS should be able to process health information and activity data from wearable devices | UBI5 (HYP8) |
| CFR45 | AEGIS should be able to process real-time data on PSPS related events from social media and RSS channels | UBI12 (HYP15, HYP16, HYP17) |
| CFR46 | AEGIS should be able to get weather data and correlate it with users' imported data | KTH10 (VIF15, VIF16) |
| CFR5 | The user should be able to perform the same kind of analysis in different ways (e.g. in critical situation analysis through a simplified algorithm) | GFT5 (HDI4, HDI5) |
| Correlation | | |
| CFR10 | AEGIS has to correlate datasets to geospatial data with their description | GFT12 (HDI14, HDI23) |
| CFR11 | AEGIS should work simultaneously with public and private (customers) data | GFT17 (HDI18) |
| CFR12 | AEGIS has to correlate customer events, habits and policy data | GFT19 (HDI19) |

| | | |
|---|---|---|
| CFR15 | AEGIS has to support interlinking of datasets from different sources, also correlating unstructured data sources with internal/IoT data | GFT24 (HDI22, HDI34) |
| CFR28 | AEGIS should be able to correlate positional information (and additional information from wearables) with Public Health Information data and announcements, taking into account also time | UBI16 (HYP22, HYP23) |
| CFR29 | AEGIS should be able to correlate selected historical data from specific sources (indoor/outdoor environmental weather data) with HVAC and lighting device data towards the extraction of user behavioural patterns that interpret the behaviour of building occupants at specific devices under different contextual conditions | UBI17 (HYP24) |
| CFR3 | AEGIS should be able to cross-check data (structured data) | GFT3 (HDI2, HDI19, HDI25) |
| CFR30 | AEGIS should be able to correlate streams of HVAC, lighting devices data, indoor environmental data and operational patterns towards the implementation of control strategies on Smart Home Environment Devices under specific building contextual conditions | UBI19 (HYP25) |
| CFR31 | AEGIS should be able to correlate Indoor Air quality monitoring data with data for the usage of HVAC devices towards facilitating the implementation of control strategies on HVAC devices by taking into account information about IAQ conditions | UBI21 (HYP26) |
| CFR32 | AEGIS should be able to correlate energy consumption data for HVAC and lighting devices with Energy (electricity) Retailer Prices towards the implementation of control strategies on specific devices by taking into account retailer prices | UBI22 (HYP27) |
| CFR33 | AEGIS should be able to correlate crime related data with HVAC devices usage data towards the implementation of control strategies on devices under specific critical conditions | UBI23 (HYP28) |
| CFR34 | AEGIS should be able to correlate occupancy data and with usage of HVAC and lighting devices data towards the implementation of an occupancy based automation framework in premises | UBI24 (HYP29) |
| CFR35 | AEGIS should be able to correlate public data (weather, public safety, etc..) with personal health vulnerability data, Smartphone/Wearables data, taking into account locality and time | UBI26 (HYP31) |
| CFR36 | AEGIS should be able to correlate real-time data that come either from third party sources or from individuals, based on their location or on the time of their occurrence, with risk patterns | UBI28 (HYP32) |
| CFR37 | AEGIS should be able to correlate personal safety profiles with smartphone data, based on location, activity type and time of execution | UBI29 (HYP33) |
| CFR38 | AEGIS should be able to correlate personal safety profiles with real time data from various sources (smartphones/wearables/sensors), based on their location and time of occurrence | UBI30 (HYP34) |
| CFR47 | AEGIS should be able to correlate environmental data with Public Health Information data, based on their location and time | UBI15 (HYP21) |

| | | |
|---|---|---|
| CFR49 | AEGIS should be able to correlate indoor environmental with outdoor environmental conditions towards the extraction of patterns that affect the operation over specific devices (how the external environmental conditions affect internal environmental conditions) | UBI37 |
| CFR50 | AEGIS should be able to correlate indoors IAQ with outdoors IAQ towards the extraction of patterns that will further facilitate the implementation of control strategies | UBI38 |
| CFR51 | AEGIS should be able to proceed with correlation analysis over indoor and outdoor environmental conditions towards the extraction of seasonal patterns - time based correlation | UBI39 |
| CFR52 | AEGIS should be able to proceed with correlation analysis over indoor and outdoor environmental conditions towards the extraction of seasonal patterns - time based correlation | UBI40 |
| CFR53 | AEGIS should be able to proceed with correlation analysis over retailer prices and energy consumption towards the extraction of patterns of energy usage affected by external factors | UBI41 |
| CFR54 | AEGIS should be able to proceed with correlation analysis over retailer prices towards the extraction of seasonal patterns - time based correlation | UBI42 |
| Input / Output | | |
| CFR16 | AEGIS shall provide capability to manage data sources (import, select, upload and download) from/to different formats (e.g. Excel, download in local storage, USB pen) | GFT25 (HDI24) |
| CFR40 | Users should be able to share and export their analysed data into different systems | KTH7 (VIF11) |
| Privacy | | |
| CFR21 | AEGIS should ask for user's approval to use their data | KTH13 (VIF25) |
| CFR48 | AEGIS should be able to anonymise personal data | UBI32 (HYP38) |
| CFR6 | AEGIS has to implement security mechanisms as well as proper handling of privacy issues (e.g. in case of using private datasets) | GFT6 (HDI7, HDI11, HDI16, HDI52, HDI66, HDI71) |
| Security | | |
| CFR20 | AEGIS should allow moderators to review and approve uploaded user's code | KTH17 |
| CFR7 | AEGIS needs to display different levels of information depending on who is accessing to the data | GFT7 (HDI8) |
| CFR8 | AEGIS should allow the creation of the different users/groups and access rights for authorized system user | GFT8 (HDI9, HDI61, HDI68) |
| Other | | |
| CFR1 | AEGIS has to provide an organized (online-offline/in private cloud) storage for the data analysis | GFT1 (HDI1) |
| CFR13 | AEGIS provides a function to upload in-house datasets to AEGIS cloud | GFT20 (HDI20) |
| CFR14 | AEGIS provides a private cloud area for dataset storage | GFT21 (HDI21) |
| CFR19 | AEGIS should allow uploading of compatible user's code to process their data | KTH4 (VIF4) |
| CFR2 | AEGIS should allow to load previous results (if available), select data sources, type of analytics and input parameters | GFT2 (HDI1) |

| | | |
|---|---|---|
| CFR9 | AEGIS should provide the ability to handle queries spanning multiple datasets/data sources and the possibility to search in a 'query storage | GFT9 (HDI12, HDI34, HDI40, HDI41, HDI42, HDI45, HDI49, HDI50, HDI62, HDI63, HDI64) |

## APPENDIX C: NON-FUNCTIONAL REQUIREMENTS BACKLOG

| Requirement Sub-category | Id | Description (Detailed description of the requirement) |
|---|---|---|
| Functional Suitability | NFR1 | AEGIS should be able to perform a great variety of analysis on selected multiple datasets (private or public) in order to provide state of the art analytics |
| Performance efficiency | NFR2 | AEGIS should be able to perform analytics in a timely and efficient manner |
| | NFR3 | AEGIS should be able to guarantee the full optimization of the response time to ensure a functional and flexible navigation through the AEGIS solution |
| Compatibility | NFR4 | AEGIS should be able to interact and exchange information with other products / IoT devices in a secure way (for example secure REST API) |
| | NFR5 | AEGIS should have the information distributed in a cloud system and be able to merge the heterogeneous information of the cloud |
| | NFR6 | AEGIS has to provide communication capabilities to allow other applications to interact with the AEGIS platform |
| Usability | NFR7 | AEGIS shall feature a user-friendly interface, provide an overview of supported kind of analytics and visualization through both online and offline user guide (i.e. demo guide, manuals and documentation database) |
| | NFR8 | AEGIS should have a Multi-language user interface |
| | NFR9 | AEGIS should provide an attractive application icon that makes the application easily distinguishable. |
| | NFR10 | AEGIS should enable to track logs about evolution and faults history and periodically send debug reports |
| Reliability | NFR11 | AEGIS should be able to securely store uploaded private datasets |
| | NFR12 | AEGIS should provide a strong algorithm to recognize aliases whenever there is a duplicate resolution |
| | NFR13 | AEGIS should ensure high availability of the system and the stored datasets |

| | | |
|---|---|---|
| Security | NFR14 | AEGIS should offer login with user credentials |
| | NFR15 | AEGIS should handle the geo-referencing data only at the application level |
| | NFR16 | AEGIS should take into account privacy and security rules according to national legislation |
| | NFR17 | AEGIS should verify the authenticity of its data sources |
| Maintainability | NFR18 | AEGIS should provide a periodically data maintenance and/or replacement during the period 8p.m to 8a.m. |
| | NFR19 | AEGIS should be easily maintainable (e.g. upgrades, bug fixing) and the user has to be informed on time |
| | NFR20 | AEGIS should be able to raise alarms about hardware/software failures of the solution |
| Portability | NFR21 | AEGIS should be able to be deployed on various Linux distributions |
| | NFR22 | AEGIS should be able to be deployed in a timely and efficient manner. |

## APPENDIX D: TECHNICAL REQUIREMENTS BACKLOG

| ID | Correlates to functional/ non-functional requirement: | Need | Description | Priority | Component |
|---|---|---|---|---|---|
| TR1 | CFR15, CFR22, CFR29, CFR30, CFR31, CFR32, CFR33, CFR34, CFR37, CFR41, CFR42, CFR43, CFR44, CFR49, CFR50, NFR4 | Process measurements from "sensor nodes" | AEGIS should be able to connect to, receive and process data from sensors and IoT devices, such as from smart home sensors and/or wearables. | High | Data Harvester, Data Annotator, Algorithm Execution Container |
| TR2 | CFR45 | Process measurements from social media sources | AEGIS should be able to connect to, receive and process data from social media sources such as Facebook and Twitter. | High | Data Harvester, Data Annotator, Algorithm Execution Container |
| TR3 | CFR10, CFR35 CFR37, | Process information from mobile devices | AEGIS should be able to connect to, receive and process data from mobile devices, | High | Data Harvester, Data Annotator, |

| | | | | | |
|---|---|---|---|---|---|
| | CFR38, CFR43, NFR4 | | including GPS signal, acceleration / gyroscope information etc. | | Algorithm Execution Container |
| TR4 | CFR23, CFR24, CFR25, CFR26, CFR27, CFR28, CFR35, CFR36, CFR46, CFR47 | Process information from open sources | AEGIS should be able to connect to, receive and process data from open sources including for example weather data etc. | High | Data Harvester, Data Annotator, Algorithm Execution Container |
| TR5 | CFR9, CFR11, CFR13, CFR15, CFR16 | Process information from relational databases | AEGIS should be able to connect to, receive and process data from relational databases. | High | Data Harvester, Data Annotator, Algorithm Execution Container |
| TR6 | CFR9, CFR11, CFR13, CFR15, CFR16 | Process information from NoSQL databases | AEGIS should be able to connect to, receive and process data from NoSQL databases. | High | Data Harvester, Data Annotator, Algorithm Execution Container |
| TR7 | CFR9, CFR11, CFR13, CFR15, CFR16 | Process information from structured sources | AEGIS should be able to connect to, receive and process documents in structured format (e.g. JSON, XML). | High | Data Harvester, Data Annotator, Algorithm Execution Container |

| | | | | | |
|---|---|---|---|---|---|
| TR8 | CFR9, CFR11, CFR13, CFR15, CFR16 | Process information from semi-structured sources | AEGIS should be able to connect to, receive and process documents in semi-structured format (e.g. csv files). | High | Data Harvester, Data Annotator, Algorithm Execution Container |
| TR9 | CFR15, CFR16, NFR4 | Connect to information sources through well-defined APIs. | AEGIS should be able to consume and store data through well-defined APIs. | High | Data Harvester, Data Annotator |
| TR10 | CFR39, CFR51, CFR52, CFR53, CFR54 | Configure data sampling / update period | AEGIS should be able to allow for the configuration of the data sampling / update period. | High | Data Harvester, Data Annotator |
| TR11 | CFR1, CFR9, CFR13, CFR14, CFR16 | Store big data | AEGIS should be able to store large datasets (Big Data Infrastructure). | High | Data Harvester, Data Annotator, Data Store |
| TR12 | CFR9, CFR18 | Produce RDF Triples | AEGIS should be able to produce RDF Triples out of the big datasets or out of the metadata of these big datasets. | High | Data Harvester, Data Annotator |
| TR13 | CFR1 | Store RDF Triples | AEGIS should be able to allow for the storage of the produced RDF triples. | High | Data Store |
| TR14 | NFR5 | Handle big data scalability | AEGIS should be able to handle (query and process) large datasets (Big Data) and be able to scale horizontally. | High | Data Store, Query Builder, Algorithm Execution Container, Big Data Processing Cluster |

| | | | | | |
|---|---|---|---|---|---|
| TR15 | NFR6 | Ensure data availability | AEGIS should be able to ensure data availability, so that data is available for data consumers at a required level of performance during normal or extraordinary operation. | High | Data Store |
| TR16 | NFR13 | Ensure data persistence | AEGIS should be able to ensure data persistence so that users can be confident that their own data, and the data used by their processes/services, is always available. | High | Data Store |
| TR17 | CFR1, CFR3 | Ensure data consistency | AEGIS should be able to ensure consistency, so that data storage is always conducted according to specific rules, and it is therefore easier to consume the required data. | High | Data Store,<br>Data Harvester,<br>Data Annotator |
| TR18 | CFR1, CFR9, CFR13, CFR14, CFR16, NFR5 | Offer distributed storage | AEGIS should be able to offer distributed storage. | High | Data Store |
| TR19 | CFR1, CFR8 | Offer public storage | AEGIS should be to able to provide public storage in order to allow access to all (non-malicious) third parties. | High | Data Store |
| TR20 | CFR1, CFR6, CFR7, CFR8, CFR11, CFR14, | Offer private storage | AEGIS should be to able to provide private storage in order to prevent access from unauthorised third parties. | High | Data Store |
| TR21 | CFR1, CFR14, NFR11, NFR17 | Provide ability to users to securely upload datasets | AEGIS should provide the ability to its users to upload datasets. | High | Data Store, Data Harvester, Data Annotator |

| | | | | | |
|---|---|---|---|---|---|
| TR22 | CFR6, CFR21, CFR48, NFR16 | Provide ability to declare IPR on uploaded datasets and set access restrictions | AEGIS should provide the ability to its users to set access restrictions to the datasets they upload to serve as the pre-conditions for the formulation of micro-contracts. | High | Brokerage Engine |
| TR23 | CFR6, CFR7, CFR8, CFR21, CFR48 | Access data | AEGIS should be able to provide to its (authorised) users data access capabilities from original and/from originating datasets, allowing them to export the required data from the available knowledge bases (e.g. based on SPARQL queries). | High | Data Store, Query Builder |
| TR24 | CFR39 | Select data | AEGIS should be able to provide to its users data selection capabilities from original and/from originating datasets. | High | Data Store |
| TR25 | CFR7, CFR8, | Implement a data management policy | AEGIS should define and implement a data management policy to restrict unauthorised access to restricted datasets. | High | Data Store, Brokerage Engine |
| TR26 | CFR6, CFR7, CFR8, NFR14 | Support authentication mechanisms for access to data | AEGIS should be able to support authentication mechanisms for restriction of malicious-driven access to data. | High | Data Store |
| TR27 | CFR6, CFR7, CFR8 | Secure datasets | AEGIS should be to offer security and encryption of the datasets so as to prevent unauthorised access to and tampering with stored datasets. | High | Data Store |
| TR28 | CFR3, NFR12 | Clean datasets | AEGIS should be able to offer (offline) tools for data cleansing. | Medium | Data Harvester, Data Annotator |
| TR29 | CFR48 | Anonymise datasets | AEGIS should be able to offer (offline) tool for (sensitive / personal) data anonymization. | Medium | Anonymization Tool |

| | | | | | |
|---|---|---|---|---|---|
| TR30 | CFR9, CFR11, CFR15 | Support interlinking of AEGIS triplestore with SLOD | AEGIS should be able to support the interlinking of the AEGIS triplestore with the SLOD to provide open access to data. | Medium | Data Store |
| TR31 | CFR9 | Provide SPARQL Endpoints | AEGIS should be able to provide SPARQL Endpoints for supporting federated queries over distributed datasets. | High | Data Store |
| TR32 | CFR3, CFR16, | Support conversion of multiple source formats to RDF | AEGIS should be able to provide the means for transforming data to RDF and support the conversion of multiple source formats – including structured and unstructured data. | High | Data Harvester, Data Annotator |
| TR33 | CFR3, CFR16, | Support conversion of multiple source formats to multiple formats | AEGIS should be able to provide the means for transforming data in one format to another format (selection amongst a close list of formats) and support the conversion of multiple source formats –including structured and unstructured data. | Medium | Data Harvester, Data Annotator |
| TR34 | CFR3, CFR9, CFR11, CFR15, CFR16, | Support interlinking of datasets from different sources. | AEGIS should be able to support automatic or semi-automatic interlinking of datasets from different sources. AEGIS should be able to configure the interlinking task by selecting meta level items (properties, classes) or instance level items (individuals) to be interlinked. | High | Data Harvester, Data Annotator |
| TR35 | | Support the reuse of available canonical standard vocabularies and ontologies. | AEGIS should be able to support the automatic or semi-automatic construction and derivation of vocabularies and ontologies from the data sources. | High | Data Harvester, Data Annotator |
| TR36 | CFR3 | Deal with corrupted, inconsistent data or | AEGIS should be able to provide measures to deal with corrupted, inconsistent data or misconfiguration of mapping rules and means | Medium | Data Harvester, Data Annotator |

| | | | | | |
|---|---|---|---|---|---|
| | | misconfiguration of mapping rules | for maintenance of data, in order to ensure the quality of the checked in and/or the transformed RDF data. | | |
| TR37 | CFR1, CFR13, CFR14 | Support management of data sources | AEGIS should be able to support the management of data sources through various functionalities: e.g. add/import, select and upload functionalities. | High | Data Harvester, Data Annotator, Data Store |
| TR38 | | Add metadata, enriching the existing data sources | AEGIS should provide the means for adding metadata/enriching the existing data sources. | Medium | Data Harvester, Data Annotator |
| TR39 | | Select / Search for / Define vocabularies | AEGIS should provide the means for selecting/ searching/ defining (Linked Data) vocabularies. | Medium | Data Annotator |
| TR40 | CFR1, NFR17 | Inspect data sources | AEGIS should allow for the inspection of data sources (and for the tentative mapping of the data sources to the available vocabularies). | Medium | Data Harvester, Data Annotator |
| TR41 | CFR9, NFR1 | Ability to handle queries spanning multiple datasets | AEGIS should be able to handle queries spanning multiple datasets. | High | Algorithm Execution Container, Query Builder |
| TR42 | CFR16 | Support format conversion | AEGIS should be able to support conversion of RDF to any format as well as various export formats including CSV, JSON, XML. | High | Data Harvester |
| TR43 | | Manually balance transformation automation and semantic soundness | AEGIS should provide the means to the user to fine-tune the balance of transformation automation and semantic soundness, whenever the automatization of some transformation is potentially speculative (e.g., guessing link types from column headers). | Low | Data Harvester, Data Annotator |

| | | | | | |
|---|---|---|---|---|---|
| TR44 | CFR4, CFR5, CFR17, CFR18, | Support data analytics on distributed big data | AEGIS should be able to support a set of different types of robust algorithms for data analytics on top of big data based on the needs of end users. Ideally, AEGIS should be able to reuse open source frameworks for machine learning algorithms. | High | Algorithm Execution Container, Query Builder, Big Data Processing Cluster |
| TR45 | CFR4, CFR5, CFR17, CFR18, | Support data analytics on distributed big data | AEGIS should be able to support a set of different types of robust algorithms for data analytics on top of distributed big data based on the needs of end users. Ideally, AEGIS should be able to reuse open source frameworks for machine learning algorithms. | High | Algorithm Execution Container, Big Data Processing Cluster |
| TR46 | CFR2 | Support initialization of algorithms | AEGIS should be able to support the initialisation of available algorithms supporting big data analytics, by defining the available set of algorithm-specific parameters prior to the execution of the algorithms. | High | Algorithm Execution Container |
| TR47 | CFR4, CFR17, CFR18, NFR2 | Support the processing of data in multiple formats. | AEGIS should be able to support the execution of robust algorithms for data analytics on top of (distributed) big data made available in various formats, whether structured or semi-structured. | High | Algorithm Execution Container |
| TR48 | CFR2, CFR40, NFR2 | Support a set of different types of output formats. | AEGIS should be able to support the execution of robust algorithms for data analytics on top of (distributed) big data making the results (data output) available in various formats. | High | Algorithm Execution Container |
| TR49 | CFR2, NFR2 | Customise the analytics process | AEGIS should be able to support functionalities for customization of the analytics process. AEGIS should be able to load previous results (optional and if | High | Algorithm Execution Container |

| | | | | | |
|---|---|---|---|---|---|
| | | | available), select data sources select type of analytics and input parameters. | | |
| TR50 | CFR45 | Support real-time analysis | AEGIS should be able to support analysis based on real time data feeds and predefined processes. | | Algorithm Execution Container |
| TR51 | CFR2, CFR40 | Manage the results of the big data enabled analytics | AEGIS should be able to support the management of the results of the big data enabled analytics, including ability to export the analytics results / output, to save and/or to visualize the produced results. | High | Algorithm Execution Container<br>Visualizer |
| TR52 | CFR2, NFR2 | Support overview of resources | AEGIS should be able to provide an interface to the users to overview the resources to be used during analytics, including the resources description, their preliminary visualization, the view of sample of the whole dataset etc. | High | Data Management<br>Visualizer |
| TR53 | CFR18, NFR3 | Preview a small selection of the results of the generated query | AEGIS should provide the means to the user to preview a small selection of the results of the generated query so as to extract some initial insights out of the foreseen analytics | Medium | Visualizer |
| TR54 | CFR5, CFR9, CFR18 | Support a set of different types of visualizations | AEGIS should be able to support a set of different types of visualizations based on different types of input datasets formats. Provide means for visualizing different data modalities (e. g. special, temporal, statistical) and provide an overview of the supported kinds of visualization | High | Visualizer |
| TR55 | CFR5, CFR9, CFR18, | Create advanced graphs based upon queries | AEGIS should be able to provide to its users the ability to create advanced graphs based upon queries spanning multiple datasets | High | Visualizer |

| | | | | | |
|---|---|---|---|---|---|
| | | spanning multiple datasets | | | |
| TR56 | CFR6, CFR7 | Support secure communication among the AEGIS components | AEGIS should be able to guarantee security and trustworthiness in the communication among the various platform components, providing secure technical interfaces between them | High | Data Harvester,<br>Data Annotator,<br>Data Indexing Engine,<br>Algorithm Execution Container,<br>Query Builder,<br>Visualizer,<br>Big Data Processing Cluster, Brokerage Engine,<br>Data Store, |
| TR57 | CFR2, CFR40 | Save analytics results | AEGIS should be able to provide to the users the ability to save their projects / results of the executed analysis for further reuse or analysis | High | Visualizer, Data Store |
| TR58 | CFR40 | Support authorized sharing of saved results | AEGIS should be able to support the "distribution" of projects / results saved by the users, with configurable sharing visibilities, namely making the results available to a subset of users authorised to access them. | Medium | Visualizer, Brokerage Engine |
| TR59 | CFR13, CFR48 | Support local deployment of the anonymization tool | AEGIS should be able to support and provide deployment options for the anonymization tool on a local platform, and provide the interfaces for the management of the tools and its execution options | Low | Anonymization Tool |

| | | | | | |
|---|---|---|---|---|---|
| TR60 | CFR3 | Support local deployment of the cleansing tool | AEGIS should be able to support and provide deployment options for the cleansing tool on a local platform, and provide the interfaces for the management of the tools and its execution options | Low | Cleansing Tool |
| TR61 | CFR3 | Support local deployment of the harvesting tool | AEGIS should be able to support and provide deployment options for the data harvesting tool on a local platform, and provide the interfaces for the management of the tools and its execution options | Low | Data Harvester |
| TR62 | CFR3 | Support local deployment of the transformation tool | AEGIS should be able to support and provide deployment options for the transformation engine on a local platform, and provide the interfaces for the management of the tools and its execution options | Low | Data Harvester |
| TR63 | CFR8 | Provide push notifications | AEGIS should be able to support the ability to provide notifications to end users (e.g. in the case an analytics task has been finalised, a dataset has been updated etc.) | High | |
| TR64 | CFR2, CFR16, CFR21, CFR40, NFR3, NFR7, NFR8 | Provide secure user-friendly interface and flexible navigation. | AEGIS should be able to provide a user-friendly interface so that the users can easily navigate between the functionalities offered by the platform in a secure way. | High | Visualizer, Query Builder, Data Harvester, Data Annotator |
| TR65 | NFR10, NFR20 | Provide system management dashboard | AEGIS should be able to offer to the Administrator a dashboard with system overview like log reports, transaction and activity report, system alarms, storage and general health report of the system. | High | |

| | | | | | |
|---|---|---|---|---|---|
| TR66 | CFR10, NFR15 | Support geospatial data analysis | AEGIS should be able to support a set of different types of algorithms for data analytics on top of geospatial data analysis | High | Algorithm Execution Container |
| TR67 | NFR18, NFR19 | Provide sustainability of system | AEGIS should be easily maintainable (upgrade and update) with a consistent and robust procedure. | High | |
| TR68 | NFR21, NFR22 | Provide a solid deployment workflow of the system | AEGIS should be able to be deployed in a timely and efficient manner. | Medium | |
| TR69 | CFR19, CFR20 | Support uploading user's custom algorithm implementation | AEGIS should provide a secure way for the user upload code implementing algorithms for data analysis. The uploaded code should be reviewed and approved by the moderators before it can be used. | Low | Algorithm Execution Container |

## APPENDIX E: USER STORIES COLLECTED

| Id | Epic | User story | | | Priority | Value | Acceptance |
|---|---|---|---|---|---|---|---|
| | | As a <type> | I want to <user requirement> | So that <reason> | | | |
| VIF1 | Browse services | Data Provider | select broken road indicator service | use broken road indicator service | 2 Med | 2 Med | Users should be able to select and use the broken road indicator service. |
| VIF2 | Data Upload | Data Provider | import collected trips data from individual vehicle drivers into AEGIS platform | analyse and visualize the data | 1 High | 1 High | Data should be correctly uploaded to the AEGIS platform |
| VIF3 | Analytic Tools | Data Consumer (Data Analyst) | use predefined algorithms for broken road detection | analyse and detect broken roads | 2 Med | 1 High | The algorithms provided should be able to detect the broken roads if any. |
| VIF4 | Analytic Tools | Data Consumer (Data Analyst) | upload their own algorithm code to be run on AEGIS platform | analyse and detect broken roads | 2 Med | 1 High | The code should be compatible with the AEGIS platform. |
| VIF5 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | specify departure and destination points on map | choose a route from departure to destination | 2 Med | 1 High | Users can enter their departure and destination then choose a route. |
| VIF6 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | choose different routes according to the road conditions | use the route that the vehicle driver chose | 2 Med | 1 High | Users can choose between different routes according to the road conditions. |
| VIF7 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | view the list of roads with their reported damage level | visualize and prioritize the damaged roads | 2 Med | 1 High | Users can view the list of damaged roads with their damage level if any. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| VIF8 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | prioritize damaged roads according to traffic level, duration, and type of damage | visualize and prioritize the damaged roads | 2 Med | 1 High | Users can prioritize damaged roads according to different criteria. |
| VIF9 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | view the list of roads with their reported damage level | help road maintenance authorities in planning of road maintenance work | 2 Med | 2 Med | Road maintenance authorities can use list of damaged roads as input for their planning of road maintenance work. |
| VIF10 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | group damaged roads list based on location and/or date sequence | visualize the grouped damaged roads | 2 Med | 1 High | Users can group damaged roads according to different criteria. |
| VIF11 | Sharing/Visibility | Data Consumer (Developer that want to use API) | allow users to share/export processed data after analysis | export processed data to be used by other systems/users | 2 Med | 1 High | Users are allowed to export the resulted data after analysis. |
| VIF12 | Connection/Integration | Data Consumer (Developer that want to use API) | get historic traffic data from governmental organization | conduct analysis on the roads and correlate with users imported data | 2 Med | 1 High | AEGIS should process historic traffic data and correlate with user's imported data. |
| VIF13 | Browse services | Data Provider | select safe driving indicator service | use safe driving indicator service | 2 Med | 1 High | Users should be able to select and use the safe driving indicator service. |
| VIF14 | Connection/Integration | Data Consumer (Developer that want to use API) | get accident hotspots | correlate accident hotspots with users imported data | 2 Med | 1 High | AEGIS should process accident hotspots data and correlate with users imported data. |
| VIF15 | Connection/Integration | Data Consumer (Developer that want to use API) | get weather data | correlate weather data with users imported data | 2 Med | 1 High | AEGIS should process weather data and correlate with user's imported data. |
| VIF16 | Analytic Tools | Data Consumer (Data Analyst) | correlate public data (weather, accidents, traffic) with users imported data | infer the driver's safety style | 2 Med | 1 High | AEGIS can correlate public data (weather, accidents hotspots, and traffic) with the driver's imported data to infer their driving safety style. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| VIF17 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | generate driver's safety style report | infer the driver's safety style | 2 Med | 1 High | User can generate a report to infer his driving safety style. |
| VIF18 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate driver's safety style report | incorporate driver safety style as selection criterion for ride sharing services | 2 Med | 2 Med | Ride sharing services can incorporate driver safety style as a selection criterion. |
| VIF19 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate driver's safety style report | evaluate the driving style of students in a driving school | 2 Med | 2 Med | Driving schools can use the driver safety report to evaluate their students. |
| VIF20 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate driver's safety style report | estimate the insurance holders accident risk in an insurance company | 2 Med | 2 Med | Insurance companies can estimate the insurance holders accident risk based on their driving style report. |
| VIF21 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate driver's safety style report | improve accuracy of driver models and testing for Advanced Driver Assistance Systems (ADAS) | 2 Med | 2 Med | Automotive developers and suppliers can use the driver's safety style report to improve accuracy of driver models and testing for ADAS. |
| VIF22 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate driver's safety style report | optimize powertrain calibration for special usage behaviour for automotive manufacturers with small production lines | 2 Med | 2 Med | Automotive manufacturers with small production lines can use driver's safety style report to optimize powertrain calibration for special usage behaviour. |
| VIF23 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | compare driver's safety style to other drivers | quantify driver's safety style in comparison to other drivers | 2 Med | 2 Med | User can generate a report to infer his driving safety style in comparison to other drivers. |
| VIF24 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate road safety status | publish road safety status and dangerous spots on the www | 2 Med | 2 Med | AEGIS can generate road safety status. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| VIF25 | Sharing/Visibility | Data Consumer (Non-expert users just to see visualizations) | ask the users for their approval to use their imported data | authorize AEGIS to use my imported data | 2 Med | 1 High | AEGIS can access user's imported data. |
| VIF26 | Browse services | Data Consumer (Developer that want to use API) | select regional driving safety risk estimator service | use the regional driving safety risk estimator service | 2 Med | 2 Med | Users should be able to select and use the regional driving safety risk estimator service. |
| VIF27 | Connection/Integration | Data Consumer (Developer that want to use API) | crawl social networks for traffic conditions | conduct analysis on the traffic conditions and correlate with users imported data | 2 Med | 2 Med | AEGIS should process social networks data on traffic and correlate with user's imported data. |
| VIF28 | Connection/Integration | Data Consumer (Developer that want to use API) | crawl newspapers for traffic messages | conduct analysis on the traffic conditions and correlate with users imported data | 2 Med | 2 Med | AEGIS should process traffic messages in newspapers and correlate with user's imported data. |
| VIF29 | Analytic Tools | Data Consumer (Data Analyst) | correlate crawled data from social networks and newspapers with public data and users imported data | calculate a regional driving safety risk metric for certain regions | 2 Med | 1 High | AEGIS can correlate crawled data from social networks and newspapers with public data and the drivers imported data to calculate regional driving safety risk metric. |
| VIF30 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | generate a regional driving safety risk metric report for specific region | identify regions with high safety risk and take actions accordingly | 2 Med | 1 High | Users can generate a regional driving safety risk metric report to identify regions with high risk and act accordingly. |
| VIF31 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate a regional driving safety risk metric report for specific region | help smart city planners, local and federal governments to prioritize and justify infrastructure investments | 2 Med | 2 Med | Smart city planners can use the generated regional driving safety risk metric report to prioritize and justify infrastructure investments. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| VIF32 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate a regional driving safety risk metric report for specific region | help Insurance companies to assess regional risk and estimate appropriate insurance premium baselines | 2 Med | 2 Med | Insurance companies can use the generated regional driving safety risk metric report to assess regional risk and estimate appropriate insurance premium baselines. |
| VIF33 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate a regional driving safety risk metric report for specific region | help automotive engineers to customize vehicles according to the requirements of a specific region | 2 Med | 2 Med | Automotive engineers can use the generated regional driving safety risk metric report to customize vehicles according to the requirements of a specific region. |
| VIF34 | Dashboards/Report/Notifications | Data Consumer (Developer that want to use API) | generate a regional driving safety risk metric report for specific region | offer services for customers such as portals and newspapers | 2 Med | 2 Med | Web content suppliers can publish the generated regional driving safety risk metric report in portals and newspapers. |
| | | | | | | | |
| HYP1 | Connection/Integration | Data Consumer (Developer that want to use API) | get indoors environmental data from installed physical devices (sensors) | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can receive and process indoor environmental data (Humidity, Temperature and Luminance) from sensors and conduct analysis or correlation with other data |
| HYP2 | Connection/Integration | Data Consumer (Developer that want to use API) | get occupancy sensing (PIR) data from installed physical devices (sensors) | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can receive and process occupancy sensing (PIR) data from sensors and conduct analysis or correlation with other data |
| HYP3 | Connection/Integration | Data Consumer (Developer that want to use API) | get Indoor Air Quality monitoring data from installed physical devices (sensors) | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can receive and process Indoor Air Quality data from sensors and conduct analysis or correlation with other data |
| HYP4 | Connection/Integration | Data Consumer (Developer that want to use API) | get data through interfaces with HVAC and lighting devices | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can receive and process data (status, mode and operational state) from HVAC and lighting |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | devices and conduct analysis or correlation with other data |
| HYP5 | Connection/Integration | Data Consumer (Developer that want to use API) | get energy consumption data for HVAC and lighting devices (from metering devices) | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can receive and process energy consumption data (from metering devices) from HVAC and lighting devices and conduct analysis or correlation with other data |
| HYP6 | Connection/Integration | Data Consumer (Developer that want to use API) | get positioning information data from mobile phones and wearable devices | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can receive and process positioning data from mobile phones and wearables and conduct analysis or correlation with other data |
| HYP7 | Connection/Integration | Data Consumer (Developer that want to use API) | get the location of a user from his device | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can receive and process location data from user's device and conduct analysis or correlation with other data |
| HYP8 | Connection/Integration | Data Consumer (Developer that want to use API) | get health information data from wearable devices | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can receive and process health information data from user's wearable device and conduct analysis or correlation with other data |
| HYP9 | Connection/Integration | Data Consumer (Developer that want to use API) | get daily behavioural routines as self-reported by users or automatically tracked by smart devices | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can get behavioural routines submitted by the user or automatically extracted by smart devices |
| HYP10 | Connection/Integration | Data Consumer (Developer that want to use API) | get external weather conditions data (luminance, temperature, humidity, IAQ data) | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can get external weather conditions data from a provided source or a consumable API |
| HYP11 | Connection/Integration | Data Consumer (Developer that want to use API) | get Public Health Information data and statistics | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can get PHI data and statistics from a provided source or a consumable API |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HYP12 | Connection/Integration | Data Consumer (Developer that want to use API) | get energy (electricity) retailer prices data | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can get and process energy (electricity) retailer prices data from provided sources and correlate with other data |
| HYP13 | Connection/Integration | Data Consumer (Developer that want to use API) | get CO2 Emissions Footprint data | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can get and process CO2 Emissions Footprint data from provided sources and correlate with other data |
| HYP14 | Connection/Integration | Data Consumer (Developer that want to use API) | get crime & accident raw data | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can get and process crime and accident raw data from provided sources and correlate with other data |
| HYP15 | Connection/Integration | Data Consumer (Developer that want to use API) | get real-time data on PSPS related events from social media channels | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can get real-time events from provided sources of PSPS related events |
| HYP16 | Connection/Integration | Data Consumer (Developer that want to use API) | get real-time data on PSPS related media streams from online RSS channels | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can get real-time media streams from provided RSS sources of PSPS related events |
| HYP17 | Connection/Integration | Data Consumer (Developer that want to use API) | get real-time data on the location of PSPS related events from social media channels | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can get real-time data on the location of PSPS related events from provided social media channels |
| HYP18 | Connection/Integration | Data Consumer (Developer that want to use API) | collect social activity data of end-users | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS can collect and process social activity data from users |
| HYP19 | Sharing/Visibility | Data Consumer (Non-expert users just to see visualizations) | consent to the retrieval of my data by AEGIS Apps | choose when to authorize AEGIS to use my data | 1 High | 1 High | User is asked to authorize AEGIS to get his data. If no authorization is granted no data must be retrieved |
| HYP20 | Sharing/Visibility | Data Consumer (Non-expert users | follow a mixed approach towards sharing data to the AEGIS Apps | choose which of my data I will authorize AEGIS to use | 1 High | 1 High | User is given the ability to choose between the types of data that will be sharing to AEGIS |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | just to see visualizations) | | | | | |
| HYP21 | Analytic Tools | Data Consumer (Data Analyst) | process environmental data from sensors and correlate with Public Health information sources | conduct analysis for the health and well-being of individuals | 2 Med | 1 High | AEGIS should process indoors and/or outdoors environmental data (Humidity, Temperature, VOC and CO2 emissions) from installed physical devices (sensors) and correlate with Public Health information sources |
| HYP22 | Analytic Tools | Data Consumer (Data Analyst) | correlate positioning data from mobile phones/wearables with Public Health Information Sources | identify pattern irregularity which could signify cognitive deterioration | 2 Med | 1 High | AEGIS should be able to process and correlate positioning information data from mobile phones and wearable devices with Public Health Information Sources |
| HYP23 | Analytic Tools | Data Consumer (Data Analyst) | correlate positioning data, additional information from wearables and PHI sources & Literature | identify physical wellbeing deterioration and frailty status | 2 Med | 1 High | AEGIS should be able to correlate positioning information data from mobile phones and wearable devices with additional information from wearable devices and Public Health Information Sources & Literature |
| HYP24 | Analytic Tools | Data Consumer (Data Analyst) | extract thermal and visual comfort profiles by taking into account selected historical data | further enable the implementation of a home automation framework towards optimal comfort levels and compliance with special, ambience-related, health requirements. | 2 Med | 1 High | AEGIS should be able to correlate selected historical data from selected sources (indoor/outdoor environmental data, weather data, HVAC and lighting devices data) |
| HYP25 | Analytic Tools | Data Consumer (Data Analyst) | correlate streams of HVAC, lighting devices data, indoor environmental data and operational patterns | provide control actions on HVAC and lighting devices towards a home automation environment fully preserving user preferences and needs. | 2 Med | 1 High | AEGIS should be able to process the streams of HVAC and lighting devices data, correlated with indoor environmental conditions (temperature, humidity and luminance) further correlated with operational/usage patterns |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HYP26 | Analytic Tools | Data Consumer (Data Analyst) | monitor Indoor Air quality conditions and further correlation with usage of HVAC devices | provide control actions on HVAC devices towards the establishment of an ambient living environment. | 2 Med | 1 High | AEGIS can process monitor data Indoor Air quality conditions and further correlation with data for the usage of HVAC devices |
| HYP27 | Analytic Tools | Data Consumer (Data Analyst) | monitor energy consumption data for HVAC and lighting devices along with Energy (electricity) Retailer Prices | provide control actions on HVAC and lighting devices as an Energy efficiency mode toward avoiding extraordinary energy costs and optimize balancing (“sweet spot”) with personal comfort/ health | 2 Med | 1 High | AEGIS can process energy consumption data for HVAC and lighting devices along with Energy (electricity) Retailer Prices |
| HYP28 | Analytic Tools | Data Consumer (Data Analyst) | monitor crime data (post-processed data from PSPS events) and correlate with HVAC devices usage | provide control actions on home controllable devices towards the enhancement of security in cases of high crime rates | 2 Med | 1 High | AEGIS can process crime data (post-processed data retrieved from an analytics process PSPS events) and correlate with usage of HVAC devices |
| HYP29 | Analytic Tools | Data Consumer (Data Analyst) | monitor occupancy data and further correlation with usage of HVAC and lighting devices | provide control actions towards an Active and Assisted Living mode | 2 Med | 1 High | AEGIS can correlate occupancy data from sensors with usage of HVAC and lighting devices data |
| HYP30 | Analytic Tools | Data Consumer (Data Analyst) | correlate public data (weather, public safety, etc..) with personal health vulnerability data, Smartphone/Wearables data | extract accurate personal safety profiles / personas to classify elderly people to different vulnerability levels | 2 Med | 1 High | AEGIS can correlate data from weather, Public safety /Crime, social media / Broadcast and news, Health provider records and announcements, along with personal health vulnerability data, Smartphone data / Wearable sensors data to extract personas with different vulnerability levels |
| HYP31 | Analytic Tools | Data Consumer (Data Analyst) | correlate public data (weather, public safety, etc..) with personal health vulnerability data, Smartphone/Wearables data | extract accurate risk patterns | 2 Med | 1 High | AEGIS can correlate data from weather, Public safety /Crime, social media / Broadcast and news, Health provider records and announcements, along with personal health vulnerability data, |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | Smartphone data / Wearable sensors data |
| HYP32 | Analytic Tools | Data Consumer (Data Analyst) | correlate real-time data with risk patterns | identify potentially dangerous incidents | 2 Med | 1 High | AEGIS can correlate selected real-time data with provided risk patterns |
| HYP33 | Analytic Tools | Data Consumer (Data Analyst) | correlate personal safety profiles with smartphone data | provide recommendations on avoidance of accidents | 2 Med | 1 High | AEGIS can correlate personal safety profiles with data from smartphone devices |
| HYP34 | Analytic Tools | Data Consumer (Data Analyst) | correlate personal safety profiles with real time data from various sources(smartphones/wearables/ sensors) | issue (real-time) notifications to app consumers | 2 Med | 1 High | AEGIS can correlate personal safety profiles with data from selected sources |
| HYP35 | Analytic Tools | Data Consumer (Data Analyst) | correlate positioning information data from mobile phones and wearable devices with identified PSPS events | issue notifications to end-users at risk | 2 Med | 1 High | AEGIS can correlate positioning information data with identified PSPS events (post-processed data retrieved from an analytics process) |
| HYP36 | Sharing/Visibilit y | Data Consumer (Developer that want to use API) | ask the app consumers for their approval/authorisation of accessing their data | use it to conduct analyses | 1 High | 1 High | User is asked to authorize AEGIS to get his data. If no authorization is granted no data must be retrieved |
| HYP37 | Analytic Tools | Data Consumer (Developer that want to use API) | perform analyses on social activity data of end users | identify patterns of behaviour | 2 Med | 1 High | AEGIS can correlate social activity data of end users |
| HYP38 | Analytic Tools | Data Consumer (Data Analyst) | anonymise and create personas out of end-users' data | identify patterns of behaviour for larger groups and classify individuals to those | 1 High | 1 High | AEGIS can anonymise data and create personas correlating user data |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HYP39 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | create rules for notifications | (semi) automatically execute once a critical situation is detected | 2 Med | 1 High | AEGIS provides the ability to create rules for notifications and execute the rules for every event detected |
| HYP40 | Analytic Tools | Data Consumer (Data Analyst) | identify clusters of users similar characteristics considering indoor environment | correlate them with other data to conduct analysis. | 2 Med | 2 Med | AEGIS should be able to identify clusters of users with similar characteristics |
| HYP41 | Analytic Tools | Data Consumer (Data Analyst) | identify trends and outliners on end users performance considering indoor environment | correlate them with other data to conduct analysis. | 2 Med | 1 High | AEGIS should be able to identify trends and outliners on end-users' performance considering indoor environment |
| HYP42 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | use a dashboard with analytics and visualisations regarding risks for the health and well-being | identify risks for the health and well-being of individuals | 2 Med | 1 High | AEGIS should provide a dashboard with advanced analytics and intuitive visualisations regarding risks for the health and well-being of individuals |
| HYP43 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | use a dashboard with analytics and visualisations regarding risks for the health and well-being | identify potential cognitive deterioration | 2 Med | 1 High | AEGIS should provide a dashboard with advanced analytics and intuitive visualisations regarding potential cognitive deterioration |
| HYP44 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | use a dashboard with analytics and visualisations regarding risks for the health and well-being | identify potential physical wellbeing deterioration and frailty status | 2 Med | 1 High | AEGIS should provide a dashboard with advanced analytics and intuitive visualisations regarding potential physical wellbeing deterioration and frailty status |
| HYP45 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | get notifications when risks for the health and well-being of individuals are identified | proactively provide on-time assistance | 2 Med | 1 High | AEGIS should identify and provide notifications when risks for health and well-being of individuals is identified |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HYP46 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | get notifications when potential cognitive deterioration of individuals is identified | proactively provide on-time assistance | 2 Med | 1 High | AEGIS should identify and provide notifications when potential cognitive deterioration of individuals is identified |
| HYP47 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | get notifications when potential physical wellbeing deterioration and frailty status of individuals is identified | proactively provide on-time assistance | 2 Med | 1 High | AEGIS should identify and provide notifications when potential physical wellbeing deterioration and frailty status of individuals is identified |
| HYP48 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | get simplified alerts when risks for the health and well-being of individuals are identified | proactively protect myself | 2 Med | 1 High | AEGIS should identify and provide simplified alerts when risks for health and well-being of individuals is identified |
| HYP49 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | get simplified alerts when potential cognitive deterioration of individuals is identified | proactively protect myself | 2 Med | 1 High | AEGIS should identify and provide simplified alerts when potential cognitive deterioration of individuals is identified |
| HYP50 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | get simplified alerts when potential physical wellbeing deterioration and frailty status | proactively protect myself | 2 Med | 1 High | AEGIS should identify and provide simplified alerts when potential physical wellbeing deterioration and frailty status of individuals is identified |
| HYP51 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | get personal safety recommendations | reduce my exposure to dangerous conditions (weather-wise, health-related, safety-related). | 2 Med | 1 High | AEGIS should provide personal safety recommendation based on the analysis of weather, health and safety data |
| HYP52 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | get insights about the streams of building environment data | establish of a home automation environment | 2 Med | 1 High | AEGIS should provide insights through visualizations of analytics to the user concerning the environmental data based on the data from the sensors |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HYP53 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | get insights about energy, comfort and IAQ KPIs | evaluate the performance of establishing a home automation environment | 2 Med | 1 High | AEGIS should provide insights through visualizations of analytics to the user concerning the energy efficiency, comfort levels and Indoor Air Quality of the building based on the data from the sensors |
| | | | | | | | |
| HDI1 | Services/Micro-services | Data Consumer (Data Analyst) | mark each analysis result in a structured folder (organized tree) | facilitate further traceability | 2 Med | 2 Med | AEGIS needs to implement a system for information traceability<br>- The system should have a data/knowledge base to handle traceability and submitted issues status<br>- The system should provide services to allow traceability queries |
| HDI2 | Services/Micro-services | Service Provider | automate the analysis quality control process | done the analysis faster | 1 High | 1 High | AEGIS needs to automate the quality control process, that can be useful also for the maintenance |
| HDI3 | Services/Micro-services | Administrator | have an interactive demo guide | reduce the training costs or the loss time due to inexpert use of the system | 1 High | 2 Med | AEGIS shall feature a guided user interface |
| HDI4 | Dashboards/Report/Notifications | Service Provider | store information about AEGIS failures | perform maintenance and troubleshooting activities with the support of logs about evolution and faults history | 1 High | 1 High | The system should enable to track logs about evolution and faults history |
| HDI5 | Dashboards/Report/Notifications | Service Provider | send debug reports | facilitate bug fixing | 1 High | 1 High | The system should periodically send debug reports |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HDI6 | Dashboards/Report/Notifications | Service Provider | provide overview of supported kind of analytics and visualization (e. g. temporal, statistical) | provide an easy to use interface | 1 High | 1 High | provide overview of supported kind of analytics and visualization through both online and offline user guide.<br><br>The AEGIS platform should have manuals and documentation database<br>User-friendly interface for the selection and the execution of the analytics task |
| HDI7 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | export reports on different media and format, such as USB pen or paper | share them with other users | 2 Med | 1 High | AEGIS shall provide capability to export reports in different format.<br>Manage data sources: add/import, select and upload |
| HDI8 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | have a Multi-language user interface | use AEGIS in my language | 2 Med | 1 High | AEGIS should have a Multi-language user interface |
| HDI9 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | provide overview means for visualizing different data modalities | chose easily the ones suitable for my needs | 2 Med | 1 High | Configure an analytics process: load previous results (optional and if available), select data sources select type of analytics and input parameters |
| HDI10 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | visualize the remaining time of processing | better organize my work | 2 Med | 2 Med | AEGIS needs to indicate the remaining time of processing in a processing mask<br>AEGIS should provide a tool to plan the time to begin an analysis (an Analysis queue function) |
| HDI11 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | transfer data to or from Excel sheets | enable the integration with most common platform for corporate data management | 1 High | 1 High | AEGIS needs to import/export and manage data from and to Excel (interoperable with Excel) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HDI12 | Data Upload | Data Curator | process data gathered from recording sensors and dedicated devices | access information in real time | 1 High | 1 High | AEGIS needs to have ability to process data gathered from recording sensors and dedicated devices |
| HDI13 | Data Upload | Data Consumer (Data Analyst) | edit specific features on the data to analyse | enrich data with my experience | 2 Med | 1 High | provide means for adding additional metadata to data sources<br>- Provide means for selecting/searching/defining vocabularies |
| HDI14 | Data Storage | Data Provider | distribute the information in a cloud system | guarantee an easy sharing and the calculations will be facilitated | 1 High | 1 High | AEGIS needs to have the information distributed in a cloud system<br>The system data/knowledge base should be able to merge heterogeneous information of the cloud |
| HDI15 | Analytic Tools | Data Consumer (Data Analyst) | reproduce a sequence of hypothetical events (what-if scenarios) | cross-check the expected behaviour with real data for evaluation purposes | 1 High | 1 High | AEGIS should be able to reproduce a sequence of events (scenarios) to cross-check data (structured data) |
| HDI16 | Analytic Tools | Data Consumer (Data Analyst) | merge different types of input datasets formats | perform more accurate analysis | 1 High | 1 High | Must provide capabilities for creation of set of graphs/analytics based on different types of input datasets formats.<br>Must provide (semi-) automatized interlinking of datasets from different sources.<br><br>Allow for inspection of data sources (and tentative mapping to vocabulary).<br>May have to work simultaneously with public and private data |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HDI17 | Analytic Tools | Data Consumer (Data Analyst) | have integrated algorithms to perform the estimation of indicators, estimation of correlations between variables, linear regression, predictive analysis and aggregation | choose the best algorithm | 2 Med | 1 High | AEGIS has to support the following analysis types: Estimation of indicators; Estimation of correlations between variables; Linear regression; Predictive analysis; Aggregation. |
| HDI18 | Analytic Tools | Data Consumer (Data Analyst) | chose a simplify algorithm | perform an analysis faster for critical situations | 2 Med | 1 High | AEGIS should implement an emergency event model. |
| HDI19 | Analytic Tools | Data Curator | recognize aliases | delete duplicate resolution and make the whole process more strength | 1 High | 1 High | AEGIS should provide a strong algorithm to recognize aliases whenever there is a duplicate resolution<br>Must let the user reuse existing canonical standard vocabularies and ontologies<br>Must let the user employ his/her own vocabularies and relate them to existing vocabularies. |
| HDI20 | Analytic Tools | Administrator | offer a personalized service | to satisfy each customer needs | 2 Med | 1 High | Available at different service levels based on the end users' needs |
| HDI21 | Connection/Integration | Data Consumer (Developer that want to use API) | AEGIS to provide communication capabilities using standards | other applications can interact with the database | 1 High | 1 High | AEGIS has to provide communication capabilities to allow others applications to interact with the database software interface<br>Must be able to set up server and mixed client/server infrastructures as specified by the user<br>Must provide interfaces for management of the tools and execution options based on the hosting platform environment |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HDI22 | Software/System Security | Administrator | the AEGIS solution to implement security mechanisms | the correct access and exchange of information could be assured | 1 High | 1 High | AEGIS has to implement security mechanisms<br>Include security protection as well as proper handling of privacy issues (e.g. in case of using private datasets) |
| HDI23 | Data Management | Data Consumer (Data Analyst) | the AEGIS solution to handle heterogeneous data | gather and cross-reference data | 1 High | 1 High | The AEGIS platform should be able to handle heterogeneous kind of data |
| HDI24 | Data Management | Service Provider | periodically data maintenance and/or replacement | AEGIS to optimize the analysis process | 2 Med | 1 High | Periodically data maintenance and/or replacement. |
| HDI25 | Data Management | Data Curator | manage corrupted or inconsistent data | achieve a strong data base | 2 Med | 1 High | provide measures to deal with corrupted, inconsistent data |
| HDI26 | Sharing/Visibility | Service Provider | AEGIS have a role based access depending on who is accessing the data and a customized visualization | data analysts, and other users participation can be improved | 2 Med | 2 Med | have a role based access to accomplish customized visualization.<br><br>AEGIS needs to display different levels of information depending who is accessing to the data |
| HDI27 | Sharing/Visibility | Administrator | have a role-based access | different users with different access capabilities could access only the pages/data that are required for his/her skills/function or role | 2 Med | 2 Med | The system should allow the creation of the different users/groups and access rights for authorized system user<br>AEGIS business models should be accessible to different users (managers, developers, operators, etc.), presenting an adequate interface to allow all of them to participate in the specification of the new processes and services |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HDI28 | Query Tools | Data Consumer (Data Analyst) | built the query | work with specific and less data | 2 Med | 1 High | The rules for the identification of the data to analyse should be editable by the end-users<br>Ability to handle queries spanning multiple datasets<br>Configure an analytics process: load previous results (optional and if available), select data sources select type of analytics and input parameters |
| HDI29 | Query Tools | Data Consumer (Developer that want to use API) | store queries | reuse the same query time after | 2 Med | 1 High | AEGIS should provide a 'query storage' |
| HDI30 | Software/System Management | Administrator | easily maintainable and configurable | satisfy the business needs fast | 2 Med | 1 High | must be easily maintainable (i.e. upgrades, bug fixing, ….) and installable |
| HDI31 | Software/System Management | Service Provider | inform the customer about product constraints/enhancements/updates | keep he informed | 2 Med | 2 Med | The platform should be able to inform the customer about constraints/enhancements/updates |
| HDI32 | Software/System Management | Service Provider | have alarms about hardware/software failures of AEGIS | notifications are received in real time and activities optimised without the need to pool for information | 2 Med | 1 High | It is needed to have alarms about hardware/software failures of the AEGIS solution |
| HDI33 | Linked Data | Data Curator | cross-check information automatically | decrease my workload and the whole process speeded up | 1 High | 1 High | The AEGIS solution should have a reasoning algorithm able to cross-check information automatically |
| HDI34 | Data Collection | Data Consumer (Data Analyst) | localize the user on time | offer a personalized service | 1 High | 1 High | the Costumer has to approve his data diffusion |
| HDI35 | Dashboards/Report/Notifications | Service Provider | upload only events of interest from different web sources | view events that can be in some way related to the guarantees | 2 Med | 1 High | events must be filtered as soon as they appear in the web page; sematic rules must be created |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HDI36 | Dashboards/Report/Notifications | Service Provider | automate the upload of events from different web sources | be updated in real time | 2 Med | 2 Med | events must be uploaded as soon as they appear in the web page |
| HDI37 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | see all the phenomenon in real time both on map and on a table | have an overview of the situation in real time | 2 Med | 1 High | create rules that could connect events shown on map with their description, plus the possibility to choose between different sources |
| HDI38 | Dashboards/Report/Notifications | Data Consumer (Non-expert users just to see visualizations) | see all the phaenomenon in real time both on map that on a table | monitor events, view descriptions and locations | 2 Med | 1 High | geo-localized data (customers + events) |
| HDI39 | Dashboards/Report/Notifications | Service Provider | automate the upload of events, customer (with their policy) from different sources | have an overview of the situation in real time and the possibility to evaluate the impact of an event | 2 Med | 1 High | automate the upload and create rules that could evaluate possible losses |
| HDI40 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | easily access customers' and events' data | understand situation of risk and send notification in real time | 2 Med | 1 High | heat map of events with plus customer |
| HDI41 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | easily access customers' and events' data | be able to do cross-selling | 2 Med | 1 High | |
| HDI42 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | easily access customers' and events' data | detect possible frauds | 2 Med | 2 Med | |
| HDI43 | Query Tools | Data Consumer (Data Analyst) | built the query | find a specific policyholder or event | 1 High | 1 High | The rules for the identification of the data to analyse should be editable by the end-users |
| HDI44 | Linked Data | Data Curator | cross-check information automatically | avoid mistakes | 1 High | 1 High | customer at risk might be identified automatically because of the event perimeter (or by the location of the event) |
| HDI45 | Dashboards/Report/Notifications | Data Consumer (Non-expert users | see all customers and correlated guarantees | monitor product distribution | 2 Med | 1 High | geo-localized data (customers + events) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | just to see visualizations) | | | | | |
| HDI46 | Dashboards/Report/Notifications | Service Provider | automate the upload of customer data (with their policy) from internal sources | have an overview of the situation in real time and the possibility to evaluate the coverage | 1 High | 1 High | automate the upload and create rules that could resume the commercial situation |
| HDI47 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | send notifications in real time | establish a connection with the customer | 2 Med | 1 High | be able to send email, call or get in contact with the customer in near real time |
| HDI48 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | predict the risk of a selected location | be able to establish the price of a policy based on the potential risk | 1 High | 1 High | |
| HDI49 | Dashboards/Report/Notifications | Data Consumer (Data Analyst) | cluster customers' needs and habits | segment customers data | 2 Med | 1 High | Application of non-intrusive predictive models to profile customers and identify a personalized offering for individuals to best fit their habits and needs |
| HDI50 | Query Tools | Data Consumer (Data Analyst) | personalized commercial offering | Customized offering to the customer to best fit behaviour and needs and improve customer satisfaction | 2 Med | 1 High | suggest new policy based on customer data |
| HDI51 | Linked Data | Data Curator | establish a connection with customer, events, habits and policy data | dynamically navigate customers' data | 2 Med | 1 High | all data must be connected automatically as soon as they are upload in the system |
| HDI52 | Data Upload | Data Consumer (Data Analyst) | Upload internal datasets | Use internal dataset to perform calculations | 1 High | 1 High | AEGIS provides a function to upload datasets from enterprise system to AEGIS cloud to use them for analysis |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HDI53 | Data Protection | Data Consumer (Data Analyst) | Select if an uploaded dataset can be shared or not with external users | Avoid data privacy issues | 1 High | 1 High | AEGIS provides a private cloud area for dataset storage |
| HDI54 | Data Protection | Data Consumer (Data Analyst) | Mask internal dataset content before uploading it | Avoid data privacy issues | 1 High | 1 High | AEGIS provides a user with tools to mask dataset before uploading |
| HDI55 | Browse services | Data Consumer (Data Analyst) | Select event sources to monitor | Monitor a multilanguage unstructured data source | 2 Med | 1 High | User can select event sources from a given list of preconfigured sources |
| HDI56 | Analytic Tools | Data Consumer (Data Analyst) | Select which kind of events to monitor | Customize AEGIS service by specifying which kind of event analysis is interested in | 2 Med | 2 Med | Depending on the event source, user can configure which type of phenomenon to monitor (e.g. Natural disasters) |
| HDI57 | Analytic Tools | Data Consumer (Data Analyst) | Set a warning threshold on the event to monitor. | Customize AEGIS service by specifying which kind of event analysis is interested in | 2 Med | 2 Med | Depending on the event source, user can configure "how dangerous" risk must be to be notified |
| HDI58 | Connection/Integration | Data Consumer (Data Analyst) | Enrich event detection with geo referenced coordinates | Correlate unstructured data source with internal / IoT georeferenced data | 1 High | 1 High | User can enrich service output with a geo referenced coordinates |
| HDI59 | Connection/Integration | Data Consumer (Data Analyst) | Enrich event detection with dataset from a given palette (e.g. Open Data) | Enrich analysis capabilities | 2 Med | 1 High | User can enrich service output with extra information from a list of preconfigured datasets (e.g. Institutional Open Data) |
| HDI60 | Data Collection | Data Consumer (Developer that want to use API) | Join an AEGIS event detection service | Include a new information source to analysis | 2 Med | 1 High | A developer can select an AEGIS event detection service to integrate it |
| HDI61 | Data Collection | Data Consumer (Developer that want to use API) | Dismiss an AEGIS event detection service | Exclude an information source from analysis | 2 Med | 2 Med | A developer can dismiss an AEGIS event detection service |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| HDI62 | Connection/Integration | Data Consumer (Developer that want to use API) | Connect to AEGIS a real time georeferenced stream data from IoT devices | Use IoT stream in the analysis | 1 High | 1 High | AEGIS can be used to store temporary IoT streams in order to process them during analysis |
| HDI63 | Query Tools | Data Consumer (Data Analyst) | Correlate internal georeferenced dataset (e.g. portfolio) with AEGIS event detection service | Use internal dataset in the analysis | 1 High | 1 High | AEGIS provides advanced algorithms to correlate event detection with internal georeferenced datasets |
| HDI64 | Query Tools | Data Consumer (Data Analyst) | Correlate real time IoT georeferenced streams (e.g. car black-box) with AEGIS event detection service | Link a potential risk to a IoT device | 2 Med | 1 High | AEGIS provides advanced algorithms to correlate event detection with IoT georeferenced streams |
| HDI65 | Query Tools | Data Consumer (Data Analyst) | Correlate real time IoT georeferenced streams (e.g. car black-box) with internal georeferenced dataset | Link a potential risk to an internal asset / customer | 2 Med | 1 High | AEGIS provides advanced algorithms to correlate internal georeferenced datasets with IoT georeferenced streams |
| HDI66 | Data Upload | Data Consumer (Data Analyst) | Download result of georeferenced correlated data | Use AEGIS result for other business applications | 2 Med | 1 High | User can download in a local storage computation results |
| HDI67 | Browse services | Data Consumer (Data Analyst) | Select a clustering algorithm from a palette to identify a clusterization in customer base | Choose the best type of customer segmentation for the current problem | 2 Med | 1 High | User can select an advanced clustering algorithm to use it for analysing his datasets |
| HDI68 | Analytic Tools | Data Consumer (Data Analyst) | Customize selected clustering algorithm | Customize clustering analysis | 2 Med | 1 High | User can configure algorithms parameters, input dimensions to analyse his datasets |
| HDI69 | Analytic Tools | Data Consumer (Data Analyst) | Simulate an algorithm on a given input dataset | Test algorithm appliance | 2 Med | 1 High | user can simulate the results of the selected algorithm applied to his dataset |
| HDI70 | Analytic Tools | Data Consumer (Data Analyst) | Compare different algorithms results on a given input dataset | Compare different type of analysis | 1 High | 1 High | AEGIS provides advanced analytics to compare different results |
| HDI71 | Data Upload | Data Consumer (Data Analyst) | Download result of the application of a selected | Use AEGIS result for other business applications | 1 High | 1 High | User can download in a local storage computation results |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | algorithm to a given input dataset | | | | |