EUROPEAN LANGUAGE GRID

D2.1

User Requirements and Functional Specifications

Authors:	Julija Melnika (TILDE), Dr. Andrejs Vasiljevs (TILDE), Dr. Raivis Skadins (TILDE), Andis Lagzdins (TILDE)
Dissemination Level:	Public
Date:	30-04-2019

About this document

Project	ELG – European Language Grid
Grant agreement no.	825627 – Horizon 2020, ICT 2018-2020 – Innovation Action
Coordinator	Dr. Georg Rehm (DFKI)
Start date, duration	01-01-2019, 36 months
Deliverable number	D2.1
Deliverable title	User Requirements and Functional Specifications
Type	Report
Number of pages	30
Status and version	Final – Version 1.0
Dissemination level	Public
Date of delivery	Contractual: 30-04-2019 – Actual: 30-04-2019
WP number and title	WP2: Grid Platform – Language Grid
Task number and title	Task 2.1: Requirements collection and specification of the higher-level ELG architecture
Authors	Julija Melnika (TILDE), Dr. Andrejs Vasiljevs (TILDE), Dr. Raivis Skadins (TILDE), Andis Lagzdins (TILDE)
Reviewers	Khalid Choukri (ELDA), Georg Rehm (DFKI)
Consortium	Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Germany
	Institute for Language and Speech Processing (ILSP), Greece
	University of Sheffield (USFD), United Kingdom
	Charles University (CUNI), Czech Republic
	Evaluations and Language Resources Distribution Agency (ELDA), France
	Tilde SIA (TILDE), Latvia
	Sail Labs Technology GmbH (SAIL), Austria
	Expert System Iberia SL (EXPSYS), Spain
	University of Edinburgh (UEDIN), United Kingdom
EC project officers	Philippe Gelin, Alexandru Ceausu
For copies of reports and other ELG-related information, please contact:	DFKI GmbH European Language Grid (ELG) Alt-Moabit 91c D-10559 Berlin Germany Dr. Georg Rehm, DFKI GmbH georg.rehm@dfki.de Phone: +49 (0)30 23895-1833 Fax: +49 (0)30 23895-1810 http://european-language-grid.eu © 2019 ELG Consortium

Table of Contents

List of Acronyms

Abbreviation	Full form or explanation
ASR	Automatic Speech Recognition
CEF AT	Connecting Europe Facility Automated Translation building block
DoA	ELG project Description of Action document
DSI	Digital Service Infrastructure
ELRC	European Language Resource Coordination action of CEF AT DSI
LT	Language Technologies
MVP	Minimum Viable Product
META	Multilingual Europe Technology Alliance
MT	Machine Translation
RMS	Requirements Management System
SME	Small and Medium Size Enterprises

List of Terms

Term	Explanation
Functional content	Language processing services and tools of different types, i.e., running services and technologies
Non-functional content	Language resources of different types (e.g., corpora, lexicons, terminologies, models), different media (e.g., text, audio, video), including language-related datasets (e.g., collections of blogposts, recordings of parliamentary debates, etc.)

Summary

This document reflects the current results of analysing user requirements and developing functional specification for the European Language Grid platform. The document starts with a review of the methodology used, tools applied and the roadmap planned for the continuous requirement elaboration process (Section 2). Section 3 lists the ELG platform objectives and its stakeholders. Section 4 contains the specification and characteristics of the main user groups who will use the platform. For each user group, we have identified and described the main user scenarios, encapsulating specific user needs to be addressed by the platform (Section 5). Section 6 is dedicated to example application scenarios that are prioritised for the Minimum Viable Product release of the platform. Finally, from these user scenarios, we elicit the initial functional and non-functional system requirements (Section 7).

1   Introduction

This deliverable is a result of Task 2.1 'Requirements collection and specification of the higher-level ELG architecture'. Although it is primarily intended for the project partners to guide the ELG platform development, it may also be useful for other stakeholders who are interested in providing their feedback and insights.

A comprehensive set of clearly defined requirements is an essential prerequisite for developing an ELG platform so that it efficiently serves the needs of its users, meets the expectations of all stakeholders and reaches the broader objectives of the project.

As ELG platform development is a long term project for developing a novel platform with a broad scope and variety of stakeholders, it is neither feasible nor possible to detail all requirements in the first few months of the project. This is why we have decided to use an agile approach in working on user requirements and functional specification. In the agile approach, requirement specification is an ongoing iterative process that actively involves intended users of the platform.

This document details the methodology that will be applied in the ongoing user requirement elaboration process through the platform development lifecycle. It also specifies the Requirement Management System that will serve as the project backlog and day-to-day tool for granular detailing of requirements into user stories and functions and for organising and tracking their implementation.

The document defines the main elements that set the basis for further elaboration and detailing of requirements. We describe the objectives of the platform and its stakeholders, define the key user groups and define initial user scenarios and the functional and non-functional requirements. This work is based on the analysis performed by the consortium partners during the project preparation as reflected in the ELG DoA document and discussions at the project meetings. The agile approach was first presented to the Consortium at the project kick-off meeting. It was further detailed at the project face-to-face technical meeting in Berlin on February 18-19th, 2019. At this meeting, a user definition session was also held with participation of the partners from this project area (DFKI, Tilde, ILSP, USFD). It was elaborated later through the feedback from Consortium members and an online meeting on April 2.

In line with the agile principles, the initial requirements included in this document will be continuously updated, elaborated and detailed during the course of the project by actively involving target users of the ELG

platform. All up-to-date requirement information at the most detailed level will always be available in the project online Requirement Management System. This document will also be periodically reviewed as defined in the Requirements Specification Roadmap (Section 2.4). If user feedback will lead to major updates or changes in the basic project elements described in this document, it will be revised accordingly .

2 Methodology

2.1 Agile Software Requirements Approach

The ELG work on user requirements and functional specification is based on the Agile software development principles. In the last two decades, the Agile development methodology has been established as the industry best practice in development of software systems. Due to its iterative nature, Agile is especially suitable for long-term new product development projects, when it is neither possible nor feasible to define and describe the whole system from the very beginning. Agile allows iterative development (the requirements become more and more detailed with each iteration), and additional requirements can be provided when necessary. It allows for natural development of the platform, without imposing unnecessary restrictions resulting from the need to make cursory decisions with lack of information and feedback from the intended users.

We will follow the well-established guidelines as defined in the *Agile Software Requirements*[1] – an elaborated and industry proven methodology that integrates the basis of several variations of agile approach.

Accordingly, ELG requirement specification will be an ongoing activity that is aligned with the deployment cycles of interim platform releases as shown in Figure 1.

Figure 1: Iterative process of requirement specification through the ELG project lifecycle

---

[1] Leffingwell, Dean. Agile Software Requirements (Agile Software Development Series). Pearson Education. First printing by Addison-Wesley Dec. 2010

Figure 2: The main steps in the ELG requirement specification process

The following are the main steps in our agile requirement specification process:

A. **Defining objectives.** We start with listing the objectives of the platform based on the ELG Description of Action (DoA) document. This establishes the broader framework for all requirement work and guides the next steps and setting of priorities.

B. **Identifying stakeholders.** We then continue with identification of the key stakeholders, including all parties who may have an interest in the project results – the ELG platform.

C. **Defining user groups.** We continue with elicitation of the main user groups based on the assessment of stakeholders who represent the target users of the system. We perform the initial analysis of the characteristics of every user group.

D. **Describing user scenarios**. We describe user scenarios pertinent to each user group. The user scenarios encapsulate the essential task that users want to perform with the platform. We start with very generic scenarios and provide only a very basic initial description. This helps to avoid preconceptions and leaves a space for involving users in detailing initial assumptions.

E. **Defining user stories**. User stories are decompositions of user scenarios detailed by granular elements. A user story can be defined as a description of a granular functionality from the user's perspective. Although the definition of user stories may be limited to just one sentence, they are often supplemented with additional elements needed for actual development (workflows, diagrams, design sketches, etc.).

F. **Defining functional and non-functional requirements**. User stories may be directly assigned for development or lead to functional requirements for the system. Besides the user perspective that is reflected in the user scenarios and user stories, the architectural and technical perspective of the system requirements are defined through specific functional and non-functional requirements.

G. **Collecting user feedback**. Every development cycle should result with a number of implemented user stories or system functions. This enables us to show this functionality to the users for validation and feedback. Collected feedback is used to initiate necessary changes in the system.

Steps D, E, F and G are repeated in a circular process through all project iteration cycles.

For each development iteration (also called Sprint), user scenarios will be elaborated through prioritised user stories. After developing and testing by the development team, implemented functionality will be provided to the potential users for validation and feedback. Thus, in effect we will be using a dual-track agile, which contains both discovery (new requirements development) and delivery (development of new features according to ready requirements) tracks running simultaneously. In parallel, we will be working with potential users to detail requirements, and specified requirements will be implemented in the system.

In addition to the User scenarios, we have defined four example application scenarios. These scenarios that encompass real-life situations for different user groups were discussed and selected by the consortium as priorities for the first version of the platform. Initial functional and non-functional requirements were reviewed against the example application scenarios, and the ones corresponding to the scenarios were marked as the first priority requirements.

2.2 User Involvement

Active involvement of users is essential in the agile process for detailing and verification of requirements, validation of scenarios, assessing the user interface and user testing of implemented functionality.

To cover the full spectrum of user needs, representatives from all key user groups will be involved in the process. This will include but will not be limited to language technology related associations and networks such as META-NET, LT-Innovate, ICT-29b) projects and CEF AT generic services projects. Communities from target user groups will be approached to invite representatives who are interested and willing to commit their time and provide useful feedback.

The appropriate tools and methods will be used to elicit user needs, verify requirements and elaborate system UX. A smaller group of selected users will be involved in the following:

- User interviews to get new insights and make an in-depth analysis of the user needs in the specific area.
- UX design sessions using prototyping tools such as wireframes and design mock-up.
- For complicated features, users may be involved in creating and/or validating User story maps or Workflow diagrams.

As these methods can involve only a small group of people, important findings will be validated by surveying larger user communities. User surveys will be conducted to verify assumptions, determine priorities and collect feedback on implemented functionality.

Principles of Design Thinking[2] will be applied to maximise the efficiency of user involvement. The idea of Design Thinking is to start with a human-centric approach and to gain an emphatic understanding of the problem from the user's point of view. It involves a constant shift between abstract and concrete extremes and requires engagement in concrete experience and abstract conceptualisation, reflective observation and active experimentation. The first part will be exemplified by using open style interviews encouraging users to share both their practical needs and emotional experiences. For the latter, users will be involved in assessing mock-ups and evaluating implemented functionality at every release cycle of the ELG platform.

---

[2] Beckman, Sara & Barry, Michael. (2007). Innovation as a Learning Process: Embedding Design Thinking. California Management Review. 50. 10.2307/41166415.

In line with the agile principles, a lightweight and very pragmatic approach will be used in documenting user involvement. All the information collected from users will be documented using the most efficient tools for supporting user-centric system development while not creating information and administrative overload.

2.3   Requirements Management System

ELG project requirements management system (RMS) is used for storing, managing, documenting, analysing, tracing, prioritising and agreeing upon requirements. It is also used to store artefacts related to requirements such as wireframes, workflows and design sketches. RMS serves as a backlog for the platform development.

The selection of an RMS for the ELG project is based on the following criteria:

- Hierarchical organisation – ability to split requirements into smaller sub-items
- Traceability – ability to connect requirements with later developed feature
- Collaboration features – with many project consortium members
- Import/export options to exchange requirement information with other tools/systems
- Security and permission control
- Define/set relationships/reuse – the RMS should ensure the ability to establish relationships be-tween requirements, show the hereditary features or possibly reuse the already done work for other requirements
- Integration with different software development environments

Based on these criteria, our current choice for RMS is Azure DevOps. Besides meeting the set criteria, Azure DevOps provides a range of extended functionality. It offers rich configuration options and is integrated with various development tools enabling dynamic tracking of requirement implementation. Requirements can be exported for use in other systems, which will make it easy to switch to another RMS if Azure DevOps will not fully meet the needs of the project.
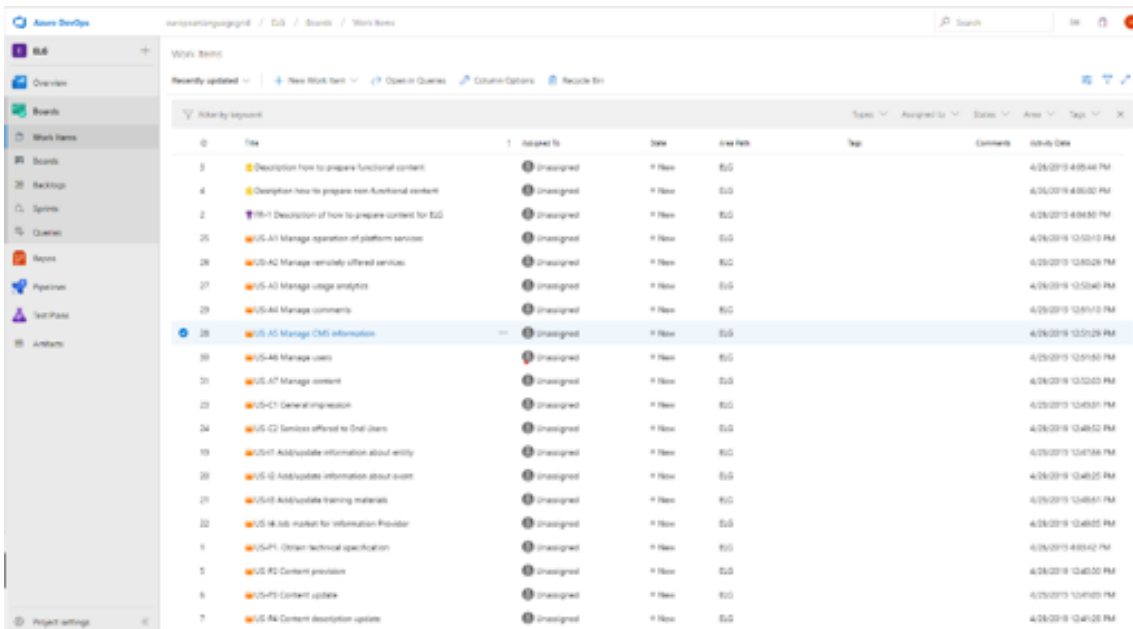

Figure 3: Screenshot of Azure DevOps online system selected for managing the ELG requirements

2.4   Requirements Specification Roadmap

The initial work on all the steps described in Section 2.1 (except E and G) was performed inside the project consortium and is reflected in the current document.

In the next phase of the project (starting in April 2019), we validate, elaborate and detail initial requirements with active involvement of representatives from target user groups and major stakeholders as described in Section 2.

The roadmap for the requirement specification is aligned with the Technical development roadmap v.1.0 that is agreed on by the Consortium (Figure 4). Besides a basic architecture, it contains 3 levels of prioritisation, thus allowing to define the Minimum Viable Product, priority features which are marked with 'A' .
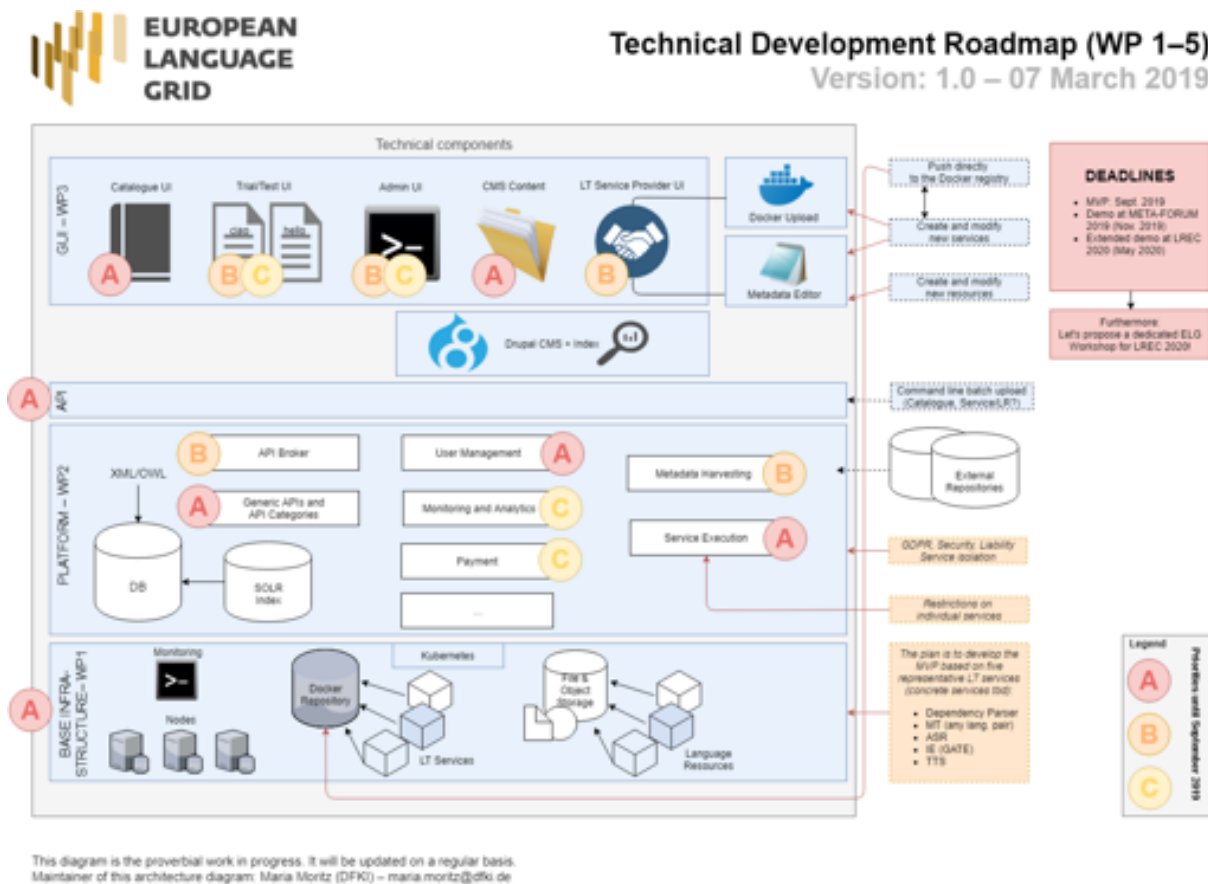
Figure 4: Technical Development Roadmap

June 2019 (M6)

In June 2019, *D3.1 Requirements and Design guidelines deliverable* will be submitted, which is tightly connected with the current document. It will contain a short description of GUI features, design sketches and design guidelines. Simultaneously, we are working with user stories that were defined as necessary for the MVP release. We will use the Requirements Management System for specifying and detailing respective user stories. For this purpose, we will employ in-depth user interviews with various platform users and conduct a user survey.

September 2019 (M9)

According to the timeline agreed upon by the consortium partners, the initial MVP should be ready by M9. It should cover A-priority features in such a way that they can be demonstrated to the potential users. It is planned that initial implementation of all A-priority features will be finished and tested in M9, and we will start gathering from the users.

If necessary, after reaching this milestone, the current document (D2.1) will be revised to reflect the major changes in ELG requirement specification. The revised version will be submitted to the European Commission and other stakeholders interested in it.

October 2019 (M11)

In October 2019, it is planned to present the demo of the system at META-FORUM 2019. The demo version of the ELG platform will be the MVP developed by M9 and improved with the feedback received from its users.

March 2020 (M15)

According to the ELG DoA, the first release of the platform GUI is planned. The platform backend will be re-leased one month later in M16. The first release will contain A-priority features in their full version. As an extended demo, this version will also be ready for demonstration at the LREC conference in May 2020.

If decided by the project partners or requested by the key stakeholders, the D2.1 may be revised again to reflect the changes in the ELG requirement specification.

February 2021 (M26)

The second official release will add B-priority features as well as improvements and bugfixes from the previous release.

With this last milestone, revision of D2.1 is still feasible if major changes require fixes in the form of the formal document. After this point, further requirement and functional specification will relate only to granular user stories and specific functions. In the Agile approach, this level of details will be managed only in the RMS system (ELG backlog).

October 2021 (M34)

The final official release will add C-priority features as well as improvements and bugfixes for the previous release.

3   Platform Objectives and Stakeholders

According to the chosen Agile methodology, we start requirement specification by stating the objectives of the system and identifying the key stakeholders.

3.1   ELG Objectives

The main objective of the project is to address the fragmentation in the European LT business and research community by establishing the European Language Grid as the primary platform for Language Technology in Europe and, thus, to strengthen the European LT business community. To achieve that and to be able to better

focus the efforts, a list of specific objectives has been identified. It is vital to consider them when working on user requirements, as all requirements should be compared against objectives to be certain that the former always contribute to the latter. The objectives directly related to ELG platform are the following:

- Business category:
    - Contribute to the establishment of the Multilingual Digital Single Market through ensuring the availability of a large amount of LT tools for European languages provided by various stakeholders, mainly **LT companies**, **research groups**, **academic institutions**.
    - Provide LT companies with a **platform & marketplace** for LT tools, services and resources. ELG should help European companies to connect to each other and serve as a smart marketplace, facilitating trade and ensuring information exchange.
- Societal category:
    - Provide a vast set of LTs to **European citizens**, including stakeholders such as **public administrations** and **NGOs**.
    - Help address the threat of digital language extinction by being a united place for LT for **all European languages**.
- Technology and Innovation category:
    - Provide European industry and research organisations with a **reliable** and **scalable** platform and marketplace.
    - Design, develop and deploy a unique LT grid architecture using the principle of **encapsulating services into containers**, thus avoiding the issues of interoperability.

These high level objectives further lead to specification of ELG stakeholders, defining user groups and user scenarios (Section 3, 4 and 5). They should also be respected in prioritisation of the system functions through the development of the platform.

3.2   ELG Stakeholders

According to the chosen Agile methodology, the project objectives and their setup are determined by the following key project stakeholders:

- Project partner organisations
- Language technology industry:
    - LT companies, including SMEs and start-ups
    - LT industry association LT-Innovate
    - Individual developers of LT resources
- Language technology researchers:
    - Research and academic institutions
    - LT associations and networks – META-NET, CLARIN, ELRA, EAMT, EACL and others
    - Individual researchers
- Developers and integrators:
    - ICT companies
    - Individual developers
- Projects in European research, innovation and deployment programmes ICT-29-2018 b) and CEF AT
- European research and industry networks that are interested in contributing or benefiting from the ELG platform such as the Big Data Value Association (BDVA), euRobotics, CLAIRE and others

- The Platform for AI related services and resources AI4EU
- European Commission DG CONNECT unit G.3 Accessibility, Multilingualism & Safer Internet
- CEF Automated Translation building block team, including respective participants from DGT eTranslation, DG CONNECT G.3 and Innovation and Network Executive Agency INEA
- EU member state institutions and networks in the LT field
- European citizens representing European multilingual diversity, with the primary focus on EU member states and Economic Cooperation Countries Norway and Iceland

The project objectives determine the primary focus of the project to be on the European language technology community although it does not exclude involvement of industry and research participants from other regions if they can contribute to the achievement of the project objectives.

The respective stakeholders will be involved in the project development process in various capacities with respect to their vested interests – requirements elicitation, design validation, testing, feedback collection, resource collection, system integration, user involvement, platform promotion and others.

From the list of stakeholders, we further select those stakeholders that represent system users - LT industry, LT researchers, Developers and integrators, and European citizens. We further analyse these stakeholders to elicit user groups (see Section 4).

4 User Groups and Characteristics

In order to fully identify all user scenarios and elicit user requirements from them, we have identified the main ELG system users and divided them into the following groups in relation to their needs from ELG:

- Content Providers
- Developers and Integrators
- Information Seekers
- Information Providers
- Citizens
- ELG Platform Administrators

It is important to understand that during one session, one user can play several roles. The same user can act as an Integrator of content, then as an Information Seeker, and then as Citizen. However, we consider them to be different User Groups, as their distinctive actions will be different in every case.

4.1 Content Providers

This category encompasses various entities that might offer their services or resources through the platform. They can be LT companies and start-ups, research organisations, universities conducting research in Language Technologies and public institutions or language communities with data sets that can be incorporated into ELG. Individuals who have created some service, tool or resource and wish to provide it on the ELG platform will also belong to this category. Finally, as part of the ELG project, it is planned to implement two open calls to demonstrate the use of ELG as a basis for more advanced LT-based modules or components useful to the industry. It is planned that the pilot projects' results will also be offered on the ELG platform.

In the ELG, the Content Providers will be interested in the following actions:

- to upload/provide a link to their content
- to provide a description and metadata that will allow potential users to find necessary types of content
- to provide content for testing
- to define a specific type of pricing or offer the content for free
- to see the analytics of content demand on the ELG

4.2   Developers and Integrators of Content

This category includes grouped users who might want to use content provided through the ELG platform by Content Providers. For example, these might be SMEs or start-ups searching for a paid product to use for their own business expansion or commercial technology players searching for an additional module in order to enrich their own product development. This category also includes various researchers and institutes, both from LT and non-LT industries, looking for services or resources necessary for moving their own research forward. Finally, public institutions can also belong to this group of users.

This user group will want to search for specific content on the ELG platform, read its full description and possibly contact its provider in order to clarify issues about the content's characteristics. They might want to preview and run testing of the content (especially necessary for paid content) before deciding to use it. Finally, the content will be downloaded and integrated with the user's own systems or executed on a cloud. In the case of paid content, the payment will have to be made according to the suggested schema. Users can also be interested in analytics about content usage, which is also especially important for paid content.

Developers and Integrators can be split into two main sub-categories:

- Commercial Developers, who use/integrate content and provide content to its customers
- Academic Researchers and Developers , who test/use/integrate content for their research needs

However, from the point of view of the ELG platform, these groups have strongly overlapping needs, and they can act as representatives of both categories in certain situations (e.g. a university that needs services or resources both for running research and for the machine translation of its website). Further on, we are treating them as one.

4.3   Information Seekers

Users in this user group are those who come to the ELG platform searching for specific information but without the aim of using the ELG marketplace, i.e. they are not interested in services or resources. An Information Seeker can be a research or a commercial organisation, a PhD student or public institution.

Information Seekers will have many various needs. They will be interested in browsing for specific material. They will try to find information about other organisations. Individuals and organisations related to the LT industry will be interested in searching for specific information about trainings or events. PhD students might come looking for job or internship possibilities.

4.4   Information Providers

This group represents users who wish to provide information about events, training, job opportunities, or information about things other than the services or tools themselves. Such information is typically provided by industry associations, academic networks, companies or research institutions, but it may also be provided by an individual expert.

4.5 Citizens

This category of users usually does not have any kind of professional connection to language technologies. In some cases, they will wander onto the ELG website with curiosity in mind. Their area of interest will be understanding what the website is about, learning about language technology in layman's terms, finding out the importance of the platform for the future and trying to understand how ELG influences their lives.

It is important to understand that while Citizens might not be the primary users for whom the ELG platform is developed, they are always the end users of the services provided by previous user groups. Thus, their importance should not be underestimated.

4.6 ELG Platform and Content Administrators

The users who will ensure the platform functioning, keep it alive, manage its content and do day-to-day maintenance belong to this group. They will be responsible for tasks concerned with the platform's technical state and for managing and organising its content, improving metadata if necessary, adding articles to the blog, cleaning up spam comments and managing ELG platform users, their groups and roles. At the same time, they will need instruments to monitor the operation of platform services, as well as remotely offered services. Finally, they will need to manage other users' accounts.

5 User Scenarios

A User scenario describes the type of user, what they want and why they want it. According to the chosen Agile Software Requirements methodology, the standard statement form is used: *"As a <role>, I want <feature> so that <reason>"[3].* Although such a short statement may suffice in smaller projects where User Stories can be directly defined, for larger projects such as ELG, we provide an extended description in the form of User Scenarios. At the same time the description is deliberately generic and limited to one or a few paragraphs to not restrict its deliberation by involving user representatives.

For the user groups described above, the following initial generic user scenarios have been identified. In the next project stage these scenarios will be revised, elaborated and detailed through prioritized user stories.

5.1 User Scenarios for Content Providers

5.1.1 US-P1. Obtain Technical Specification

Statement: As a Content Provider, I want to get information on how to prepare my functional or non-functional content for the ELG in order to provide them through the ELG.

Content Providers get technical information about supported data formats and tools and step-by-step descriptions of actions needed in order to prepare functional or non-functional content for the ELG. The description provides various variants depending on content type or uses a wizard-process where the user needs to make choices in order to obtain suitable answers. A special option is provided to define the license type, which the Provider might apply to the content.

5.1.2 US-P2. Content Provision

Statement: As a Content Provider, I want to provide my content so that it will be available on the ELG.

---

[3] See also "What is Role-Feature-Reason?". Retrieved 10 April 2019 from https://www.agilealliance.org/glossary/role-feature/

This User Story is specific for each type of content listed:

- Containerised tool
- Downloadable content
- API
- Link to GitHub, container image repository or other place where the content is stored

Containerised tool

The Provider prepares the container with the functional content as well as the description of it. The container is built and pushed to the container registry (hosted by ELG or by a third party). The link to the container image is provided to ELG, and the description of the container (metadata) is manually entered or uploaded to ELG. The Provider chooses if the container image will be available for download or should be deployed only on the ELG platform. Next, the provider chooses if the tool should be provided for free or not and selects the pricing model if necessary. The Provider can also switch off user-generated rankings, reviews, etc. for paid content.

Downloadable content

The Provider prepares downloadable content and its description. Content is uploaded to the ELG, and the description of the content is manually entered or uploaded to ELG according to the metadata schema. The Provider chooses if the content is provided for free or not, and applies the pricing model if necessary. The Provider can also switch off user-generated rankings, reviews, etc. for paid content.

API

The Provider prepares and hosts the functional content and its API on infrastructure controlled by him. The API description is submitted to the ELG, and the description of the content is manually entered or uploaded to the ELG according to metadata schema. The Provider chooses if the API is available for free or not, and describes the pricing model if necessary. The Provider can also switch off user-generated rankings, reviews, etc. for paid content.

Link to GitHub or other repositories where content is stored

The Provider prepares the information about the content (link to GitHub, container registry or other repository where content is stored) and the content description. A link is entered in a specific field in ELG, and a description of the content is manually entered or uploaded to ELG according to metadata schema.

5.1.3   US-P3. Content Update

Statement: As a Content Provider, I want to provide a new version of the content so that the latest version is available on the ELG.

A Provider prepares a new version of the content. In ELG, the Provider chooses the content for which the version needs to be updated and provides the new version. The Provider needs to update the content description (according to the metadata schema), the new version can also have a payment model and a license different from the older version. If there are external implementations done with the old version of the content and the new version introduces breaking changes, it needs to be ensured that the procedures are not disrupted for Integrators.

5.1.4 US-P4. Content Description Update

Statement: As a Content Provider, I want to update/change any of the provided content meta-information/payment schema so that the content has the correct description.

The Provider prepares new meta-information and either enters it manually or uploads it to ELG. In case the payment schema is changed, it will apply only to the new content purchases.

5.1.5 US-P5. Content Removal

Statement: As a Content Provider, I want to remove the content so that it is not available on ELG anymore.

In case a Content Provider does not want to provide the content anymore, he is able to remove or temporarily disable the content from ELG. It needs to be ensured that the content removal is appropriately handled with the Integrators of the content and complies with the Terms of Use and the licensing conditions.

5.1.6 US-P6. Content Usage Analytics for Providers

Statement: As a Content Provider, I want to get information about content usage (time or amount of usage, number of downloads, etc.) so that I know how useful, popular or successful the content is and how much I have earned with it (for paid content).

A Content Provider can view the analytics about the content. The analytical dimensions shown depend on the type of the content. The Content Provider is able to see analytics for an item for a chosen period of time.

5.1.7 US-P7. Content Promotion

Statement: As a Content Provider, I want to increase the content exposure on ELG so that it becomes more visible.

The Content Provider can order additional exposure of his content in various ways offered by the platform (banner, textual information, primary position in search results, etc.).

5.2 User Scenarios for Developers and Integrators of Content

5.2.1 US-U1. Content Search

Statement: As a Developer/Integrator, I want to search/browse the content in order to find the content I need.

A Developer/Integrator wants to find content. The Developer/Integrator is able to browse the catalogue drilling down into various categories. He searches for content using various options, keyword search as well as faceted search based on selected elements of the metadata. Search results show the list of content items and basic elements of their description. There is also a preview option available from search results. For non-functional resources, a part of the content is shown as a preview. For functional services or tools, a preview might be created as a description page or as a demo version provided to ELG.

5.2.2 US-U2. Technical Information for Developer/Integrator

Statement: As a Developer/Integrator, I want to find out how to test or use the functional service or tool so that I can use it.

A Developer/Integrator gets a step-by-step description of actions needed in order to test or to use the functional or non-functional content. The description provides various variants depending on content type or uses a wizard-process where the user needs to make choices to obtain a suitable result.

5.2.3   US-U3. Using Content

Statement: As a Developer/Integrator, I want to use the content provided on ELG to achieve my goals.

This User Story is specific for each type of content listed:

- use the functional service or tool on ELG (container running on ELG)
- use the remotely invoked content (e.g. API)
- download and deploy the functional content on the user's infrastructure
- use/download the non-functional resource

Use the functional service or tool on ELG (container running on ELG)

A Developer/Integrator chooses the content. If it is necessary, the Developer/Integrator contacts the Content Provider for additional information. If the content is paid, the Developer/Integrator can order it and pay via online payment. The billing process is handled by ELG. The Developer/Integrator can test the functional service or tool if the Content Provider has provided such an option. If the Developer/Integrator wants to use the running service, the runtime instance of a container is created on ELG, and the service becomes available to the Developer/Integrator for a definite period of time.

Use remotely invoked content (API)

A Developer/Integrator chooses the content. If it is necessary, the Developer/Integrator contacts the Content Provider for additional information. If the content is paid, the Developer/Integrator can order it and pay via online payment. The billing process is handled by ELG. The Developer/Integrator can test the functional service or tool if the Content Provider has foreseen such a possibility, i.e., provided a specific content version for testing purposes. The Developer/Integrator can also remotely invoke such service using RESTful API.

Download and deploy the functional content on the user's infrastructure (e.g., container deployed on private cloud, script, link to GitHub, executable file saved on ELG or any other repository)

The Developer/Integrator chooses the content. In case it is necessary, the Developer/Integrator contacts the Content Provider for additional information. If the content is to be paid for, the Developer/Integrator can order it and pay via online payment. The billing process is handled by ELG. Afterwards, the content is offered for download. For paid content, an option for download should remain available according to the payment schema. The Developer/Integrator can deploy the containerised content on his own infrastructure.

Use/download the non-functional resource (e.g., dataset, model)

The Developer/Integrator chooses the resource. In case it is necessary, the Developer/Integrator contacts the Content Provider for additional information. If the resource is to be paid for, the Developer/Integrator can order it and pay via online payment. The billing process is handled by ELG. Afterwards, the Developer/Integrator is offered an option to download the resource. For paid resources, an option to download should remain available according to the payment schema.

5.2.4   US-U4. Content Usage Analytics for Developer/Integrator

Statement: As a Developer/Integrator, I want to get information about my content usage (time or amount of usage, number of downloads, etc.) so that I know how much time or money I have spent for it.

A Developer/Integrator can review the usage information about the content (which is especially important for paid content). The dimensions shown depend on the type of the content. The Developer/Integrator is able to see the analytics for a chosen period of time.

5.3   User Scenarios for Information Seekers

5.3.1   US-S1. Information Search

Statement: As an Information Seeker, I want to find information about an entity (company, organisation, etc.) so that I can use information about it.

The Information Seeker searches for information about entities (e.g., companies or organisations) in the catalogue. The user is looking for a unified search experience or browsing options.

Possible search domains can be name, location, specialisation, etc.

5.3.2   US-S2. Search for Information about Event

Statement: As an Information Seeker, I want to find information about educational or other events so that I can visit them.

An Information Seeker searches for information about events (educational or other). The User is looking for a unified search experience or browsing options.

Possible search domains can be event name, organiser, type, location, etc.

5.3.3   US-S3. Use Training Materials

Statement: As an Information Seeker, I want to find training materials so that I can fill in my knowledge gap.

An Information Seeker searches for various training materials. The User is looking for a unified search experience or browsing options.

Possible search domains can be: training material name, producer, topic, etc.

5.3.4   US-S4. Job Market for Information Seeker

Statement: As an Information Seeker, I want to find information about available job opportunities/internships/freelance opportunities so that I can change my job/find a more interesting job.

An Information Seeker searches for employment information. The User is looking for a unified search experience or browsing options.

Possible search domains can be position title, location, organisation type, remuneration type, etc.

5.4   User Scenarios for Information Providers

The following scenarios all have the same step-by-step sequence, only the object of the scenario changes, thus each scenario has a separate name and statement, but the step-by-step description is united at the end of this section.

5.4.1   US-I1. Add/Update Information about Entity

Statement: As an Information Provider, I want to place or update information about an entity (company, organisation, etc.) so that this information can be found by Information Seekers.

5.4.2 US-I2. Add/Update Information about Event

Statement: As an Information Provider, I want to place or update information about an educational or other event so that potential visitors can visit it.

5.4.3 US-I3. Add/Update Training Materials

Statement: As an Information Provider, I want to place or update training materials, so that users can learn something new.

5.4.4 US-I4. Job Market for Information Provider

Statement: As an Information Provider, I want to place/update information about available job opportunities/internship/freelance opportunities so that I can find a suitable candidate to fill a given position.

Step-by-step description

The Information Provider wants to record information about an entity in the catalogue or CMS. This entity might be an organisation working in the LT area, using LT content but not providing it yet, and the information might be about education, training materials, job opportunities, internships, freelance opportunities or other events. In order to add new information, the Information Provider fills in a special form, which is then reviewed by the moderator, who decides if this information should be published on the website. Another option is immediate publishing of the entity's information on the website with post-factum check by the administrator.
To update the information, the Information Seeker fills in the same form.

5.5 User Scenarios for Citizens

5.5.1 US-C1. General Impression

Statement: As a Citizen, I want to know what ELG is and how it affects me so that I can understand whether it is useful for me.

The Citizen, during internet surfing or through search, comes to the ELG page and becomes curious about ELG. He will look for information, using the same approach to the website structure as used on a website he normally visits. The next step will be natural curiosity: 'How does it affect me? Why is this website useful to me?' However, the focus time is very short, so this information should be very easily available, otherwise the Citizen will leave the page.

5.5.2 US-C2. Services Offered to End Users

Statement: As a Citizen, I want to use ELG services offered to end users so that I can fulfil my own tasks easier with the help of these services.

The Citizen has found the page with services offered to the end user. He is able to choose easily between various categories, as well as between competing services used for the same goal (e.g. between various translation engines for a specific language pair).

5.6 User Scenarios for ELG Platform and Content Administrators

5.6.1 US-A1. Manage Operation of Platform Services

Statement: As an Administrator, I need to manage platform service operations so that I am able to keep the platform going.

The Administrator of the platform wants to be able to monitor and manage all the platform functions, to be notified about possible issues (especially about problems that have already occurred) and to be able to review the reports about the system performance in a chosen period of time. The Administrator needs to be able to start and stop specific platform services and fulfil other operations necessary to ensure the platform performance.

5.6.2   US-A2. Manage Remotely Offered Services

Statement: As an Administrator, I need to monitor and manage remotely offered services to ensure that they are available.

The Administrator needs to be able to follow up on the availability of the remotely offered services. The system checks their availability automatically after some time and notifies the Administrator in case any of the services becomes unavailable. The system also provides reports about the availability of services in a selected time span.

5.6.3   US-A3. Manage Usage Analytics

Statement: As an Administrator, I need to monitor usage analytics so that I can react to malfunctions or security threats in the system.

The Platform administrator needs to be able to follow up on the usage of functional and non-functional services provided on ELG. If one service usage raises significantly, the Administrator needs to be notified.

5.6.4   US-A4. Manage Comments

Statement: As a Content Administrator, I need to moderate comments in order to provide a safe and friendly environment for all users.

The Content Administrator needs a comfortable interface to be able to review all comments left by other users, independent from place/entities for which these comments were left. The Content Administrator needs the possibility to edit or remove user comments, e.g. in case they are offensive or contain spam.

5.6.5   US-A5. Manage CMS information

Statement: As a Content Administrator, I need to approve or update the information of the CMS (blog, other articles, editing the meta-information or user-generated content) so that fresh information is available for users.

Requests for information approval or update come to the corresponding moderator from various sources. Some updates can come through outside communication means, such as emails, while others will come as internal requests for information approval from Information Providers. The Content Administrator updates the information with the ability to roll-back at any point in time.

5.6.6   US-A6. Manage Users

Statement: As an Administrator, I need to manage active users on the platform to ensure user authentication and authorisation.

The Administrator needs to be able to temporarily freeze or block accounts of users in case of suspicious activities from user accounts or when a user creates an unsafe environment for everyone else.

5.6.7   US-A7. Manage Content

Statement: As an Administrator, I need to manage platform content to ensure its completeness and exhaustiveness.

The Administrator needs to be able to ensure and assist content provision from similar platforms.

6   Example Application Scenarios

The following scenarios encompass real-life situations for different user groups were discussed and selected by the consortium as priorities for the first version of the platform. In the next project phase these scenarios will be revised and elaborated involving user representatives.

6.1   University Provides Container-type Tool

A User (University) is already a user of the ELG platform and has an account. However, this is the first time the User wishes to provide a containerised tool, so there is a special training that helps to understand how to do that correctly. There is also a licensing wizard available to make sure that the chosen license type is correct. The User prepares the image according to instructions; however the user is experiencing problems with the image preparation and asks for help from the Helpdesk. After the image has been created and published to the container registry, the user logs into the ELG site, submits the access information for the image to the site and then provides a full description of the containerised tool according to the requested metadata. The User chooses if the tool will be available for use on the client's premise or should be deployed and exploited through ELG. Next, the user chooses that the tool is available for a fee and applies the pricing model. The User also decides not to switch off user-generated content (rankings, reviews, etc.)

6.2   Organisation Provides Non-functional Resource

A User (Organisation) has not used the ELG platform before. The resource has been uploaded to different databases before, so metadata have already been prepared and can be imported or even harvested. Additionally, the resource consists of several files. The user creates a new account, logs into the system, and then is able to upload all files in one approach. The metadata is automatically imported as well. Next, the user is able to edit any of the resource description information and to choose that the resource will be provided for free.

6.3   Company Needs Solution

A User is an e-commerce company that wishes to extend its geographical range to another EU country whose language is not known to anyone in the company. The User wants to find a specific solution that could be used for translations, preferably one that could be run from the users' servers. The User does not know the specific product name, so a simple search by the term 'translation' and a faceted search by content type (application), language pair, availability conditions and licensing type should be used. The search results in several options, all of which seem suitable. Some of the options offer a testing possibility, while others don't. The user contacts the content owners to find out more about the testing options. After an evaluation, the client makes a Service Level Agreement (SLA) with the content owner or automatically within the ELG platform and integrates the content into the client's product (remote calls to the ELG platform's translation service). In this example Clients pay a fixed price monthly to the ELG platform/Content Provider plus an amount depending on use statistics.

6.4   Organisation Needs Solution

A non-profit organisation needs a specific text-to-speech solution for all European languages. The organisation cannot pay for the solution, so they run a search on ELG, looking for free solutions. They find the most suitable (though not ideal) solution for their situation, one that provides a limited amount of services for free but demands payment after a certain threshold is reached. The solution is provided as an API. The organisation uses the API to integrate the provided solution into their own system. After the implementation, the organisation constantly monitors the usage of the content to make sure that the free amount provided is enough.

7   Initial Requirements

These initial requirements will be validated, elaborated and detailed in the next project phases through the involvement of users from various target user groups.

7.1   Functional Requirements

For priority definition, we used the MoSCoW model, which is a standard approach for Requirements prioritisation in the Agile environment, where each requirement is placed into 1 out of the 4 categories: *Must be done (M), Should be done (S), Could be done (C), Won't be done (W)*.

Initial prioritisation has been done based on the Platform objectives and the definition of the Minimum Viable Product, produced by Consortium members in the Technical kick-off meeting. The initial prioritisation will be iteratively reviewed during later phases of the project.

Table 1. ELG Functional requirements

Req ID	Description	Ref. to Scenario	Priority
FR-1	Detailed step-by-step description of the process defining how to prepare any type of content, both functional or non-functional for ELG	US-P1	M
FR-2	Wizard-type functionality for users explaining how to prepare any type of content, both functional or non-functional for ELG	US-P1	C
FR-3	Wizard-type functionality for users allowing to understand the type of licences that should be used for specific content	US-P1	M
FR-4	Context specific help to users	US-P1, US-U2	S
FR-5	Helpdesk option for non-standard questions	US-P1, US-U2	C
FR-6	Support for containerised tools, stored on a cloud container registry hosted by a third party	US-P2, US-U3	M
FR-7	Support for containerised tools, stored on a cloud container registry hosted by ELG	US-P2, US-U3	M
FR-8	Support for non-containerised content, uploaded to ELG (downloadable content)	US-P2, US-U3	M

FR-9	Support for content, for which an API is provided on ELG	US-P2, US-U3	M
FR-10	Support for content, for which a link to GitHub or another repository is provided on ELG	US-P2, US-U3	M
FR-11	Workflow for registering (including or excluding upload) of new content on ELG	US-P2	M
FR-12	Catalogue for all content and description of content items. Description fields for each content must be defined according to the metadata schema. The catalogue should also contain information about entities, necessary for describing the content, e.g., about organisations.	US-P2, US-U3, US-S1	M
FR-13	Upload of file with content description according to metadata schema	US-P2	S
FR-14	Running instances of containerised tools	US-P2, US-U3	M
FR-15	Download of content that was uploaded to ELG	US-P2, US-U3	M
FR-16	Various payment options for content usage	US-P2, US-U3	C-W
FR-17	Possibility for Content Provider to apply one of pre-defined payment schemas for content usage	US-P2	C-W
FR-18	Possibility to add user-generated ratings, reviews, etc. to content	US-P2	C
FR-19	Possibility for Provider of paid content to switch off user-generated ratings, review, etc. for specific paid content	US-P2	C-W
FR-20	Upload of new version of any content. The process must be similar to uploading new content, with the only difference being the pre-filled description file. ELG must state connection of versions on the platform.	US-P3	M
FR-21	Upload of additional versions of content with changed (reduced) functionality as teaser for paid content. The process must be similar to uploading new content, with the only difference being the pre-filled description file. ELG must state connection of versions on the platform.	US-P2, US-U3	C-W
FR-22	Content description update through manual adjustment	US-P4	M
FR-23	Content description update through description file upload	US-P4	S
FR-24	Content withdrawal from ELG by Content Provider. Withdrawal by provider must not be possible in the case when future use of content has been paid for by at least one user.	US-P5	S
FR-25	Content withdrawal from ELG by System Administrator	US-P5	M

FR-26	Possibility to see the content usage statistics in the chosen period of time. The statistics shown should depend on the content type (number of downloads, time or amount of usage, etc.).	US-P6, US-U4	S
FR-27	Mechanisms for content promotion	US-P7	W
FR-28	Simple keyword search possibility	US-U1	M
FR-29	Tree-like structured catalogue of all content with possibility to drill down into various categories	US-U1	M
FR-30	Faceted search possibility on the basis of selected elements of the metadata model.	US-U1	M
FR-31	Preview option in search results list, which is defined by Content Provider. Preview might be textual information or a reduced version of the functional service or tool.	US-U1	S
FR-32	Detailed step-by-step description of the process defining how to test or execute any type of content, both functional and non-functional for ELG	US-U2	M
FR-33	Wizard-type functionality for user explaining how to test or execute any type of content, both functional and non-functional for ELG	US-U2	C
FR-34	Content usage for a defined amount of time/defined amount of usage counts	US-U3	S
FR-35	CMS for various information types • Events (date and description) • Training materials (structured text, pictures, videos, files) • Vacancies (Text)	US-S1	M
FR-36	Search options must work through all catalogue items and CMS. Search results must be prioritised based on from where the search was invoked. Search results can also be prioritised by the user with the help of search options.	US-S1, US-S2, US-S3, US-S4	M
FR-37	Possibility for Users to add information in the CMS according to previously defined formats. Added information is approved in two ways: 1) Submitted information must be approved by the Administrator before publishing or 2) Submitted information is published and shown to the Administrator as new and can be removed afterwards by the Administrator. It must be possible to use different approval modes for different CMS parts.	US-I1, US-I2, US-I3, US-I4	M
FR-38	Published information can have an expiration date, after which it is not being shown on the website (but can remain searchable).	US-I1, US-I2, US-I3, US-I4	C

FR-39	Descriptive information about platform for all audiences	US-C1	M
FR-40	Interface for execution of specific content and showing results to end users	US-C2	S
FR-41	Search interface for executable content	US-C2	S
FR-42	Administrator interface to monitor system activities	US-A1	S
FR-43	Notification should be sent to Administrator through various channels about malfunctions	US-A1	S
FR-44	Administrator should be able to stop and start specific platform services	US-A1	S
FR-45	Reports about system performance in a chosen period of time	US-A1	S
FR-46	Interface to follow up on availability of remotely offered services	US-A2	S
FR-47	Availability check of remotely offered services after defined time periods	US-A2	S
FR-48	Notification sent to Administrator through various channels about the unavailability of a remotely offered service	US-A2	S
FR-49	Interface for Administrators to monitor usage of content	US-A3	S
FR-50	Notifications should be sent to Administrator through various channels about the irregularities in content usage	US-A3	S
FR-51	Interface for Administrator to see newly created User-generated content and edit/remove it if necessary	US-A4	S
FR-52	Interface for Administrator to approve content generated by Information Providers for publishing	US-A5	S
FR-53	Interface for Administrator to review all content generated by Information Providers and published in the previous period of time	US-A5	S
FR-54	Saving previous versions of information published in CMS with possibility for Administrators to roll back to previous versions	US-A5	S
FR-55	Interface for Administrator to manage users, i.e., to assign and define roles, rights, groups, and assign users to them, etc.	US-A6	S
FR-56	Interface for Administrator to block or delete a specific user	US-A6	S
FR-57	Interfaces for information harvesting and metadata storing with existing similar initiatives	US-A7	S
FR-58	API for the integrators of tools and services	US-U3	M

FR-59	The availability of operational context-sensitive user assistance across all ELG Platform user and administration interfaces should be ensured. User Help should be provided as: • Information available to the user about each of the fields that should be filled in • Link to help files • Search Option for ELG Platform Interfaces	US-P1, US-U2	S
FR-60	Authentication required from users in all operations that: • are related to provision of content • involve financial transactions • involve content for which authentication is necessary	US-A6	M
FR-61	The following authentication options must be supported: • authentication using system built-in authentication mechanism with user name and password and/or any chosen external authentication provider • authentication must be provided in web services (API) using user credentials Authentication should be provided for shared service interfaces.	US-A6	M
FR-62	The ELG Platform must account for privacy, ethics and other legal requirements (e.g., GDPR).	all	M

7.2 Non-functional Requirements

Revision and prioritization of non-functional requirements will be performed in the next phase of the project.

Table 2. ELG Non-functional Requirements

Req ID	Description
NFR-1	The ELG Platform should scale to provide smooth and uninterrupted operations for a greatly fluctuating number of simultaneous end-users.
NFR-2	The ELG Platform must support multiple simultaneous run-time instances of containers deployed without any system malfunction.
NFR-3	System performance limitations, if exist, must be properly handled.
NFR-4	The system must be built on a scalable and robust Open Source framework for building and operating cloud platforms.

NFR-5	Base infrastructure should be able to receive, manage and run many images in parallel. The basic layer of the system should be set up using an IaaS solution that enables the use of large pools of resources for computing, storage, networking, etc.
NFR-6	The ELG platform should respond in an appropriately short number of seconds to every user request.
NFR-7	The platform should provide GUI with rich user experience and be simple and easy to use.
NFR-8	The ELG Platform should be scalable to ensure an increase of performance by adding additional hardware without system malfunctioning.
NFR-9	ELG architecture needs to provide solutions for automatic load-balancing.
NFR-10	The architecture and design should provide for notification of software errors and exceptions to the server. System monitoring should be carried out at user interface, application and database level.
NFR-11	The ELG Platform must have user interfaces based on the Design guide and ensure unified graphical and function experience across the platform.
NFR-12	The interface must be user-friendly and intuitive without the overhead of understanding how the services are working.
NFR-13	When developing the ELG Platform interface, the following guidelines should be applied: • ANSI/HFES 200 'Human Factors Engineering of Software User Interfaces' • Microsoft 'Inductive User Interface Guidelines' • 'WEB Content Accessibility Guidelines' (WCAG) 2.1
NFR-14	The graphical User interface must support different types of platforms – computers, tablets, mobile devices.
NFR-15	The web interface should be compatible with the latest versions of common web browsers, such as: • Microsoft Edge (chromium based) • Google Chrome • Mozilla Firefox • Safari • Opera
NFR-16	The ELG graphic user interface must be usable on mobile devices (tablets, smartphones, etc.) with different versions of operating systems: • Android • Apple iOS
NFR-17	ELG Platform user interfaces (menus, buttons, messages, etc.) must be considered to be provided in all official EU languages.
NFR-18	ELG Platform development must follow application security principles (see for example, OWASP principles and guidelines http://www.owasp.org/index.php/Category:Principle)

7.3   Functional View on User Requirements

The following table shows interaction between User Groups and Objects:

Table 3. Interaction between User Groups and Objects

User group/Object	Content Provider	Developer/ Integrator	Information Seeker	Information Provider	Citizen	Administra-tor
Functional Service or Tool	+	+				+
Non-functional resource	+	+				+
Trial content version	+	+				+
Executable content	+				+	+
Content description	+	+				+
Metadata schema	+	+				+
Catalogue entity not describing the content (e.g. organisation)	+	+	+	+		+
CMS entity (e.g. article, event, training, vacancy)			+	+		+
User-generated Content	+	+	+	+	+	+
Payment	+	+				+
Payment schema	+					+
Content usage statistics	+	+				+

User group/Object	Content Provider	Developer/ Integrator	Information Seeker	Information Provider	Citizen	Administrator
Search	+	+	+	+	+	+
Helpdesk	+	+	+	+	+	+
Newsletter	+	+	+	+		+
Blog	+	+	+	+	+	+
System Monitoring and Analysis						+
User Management/ User Account	+	+				+