Research and Innovation Action (RIA) H2020 – 957017

SELMA

Stream Learning for Multilingual Knowledge Transfer

https://selma-project.eu/

D1.1 Use Case Description and Requirements

Work Package	1
Responsible Partner	Deutsche Welle
Author(s)	Hala Attig, Andreas Giefer
Contributors	Sebastiao Miranda, Peggy van der Kreeft, Joscha Simon Rieber, Yannick Esteve, Guntis Barzdins
Version	1.0
Contractual Date	30 June 2021
Delivery Date	30 June 2021
Dissemination Level	Public

Version History

Version	Date	Description
0.1	10/05/2021	Initial Table of Contents (ToC)
0.2	26/05/2021	First draft with template for contributions
0.3	02/06/2021	Integrated all contributions from all partners
0.4	18/06/2021	First full version ready for review
0.5	22/06/2021	Internally reviewed version
0.6	24/06/2021	Finalized version after internal review
1.0	28/06/2021	Final version for submission

1. Executive Summary

The Use Case Description and Requirements report gives a detailed overview of the main use cases and the requirements associated with them. It includes the personae, the user scenarios and the workflows.

SELMA's central concept is to build a deep-learning NLP platform that trains unsupervised language models, using a continuous stream of textual and video data from media sources and make them available in a user/topic-oriented form in over 30 languages.

The knowledge learnt from deep contextual models is transferred to a set of NLP (Natural Language Processing) tasks and made available to users through a **Media Monitoring Platform** (Use Case 1) that will be able to handle up to ten million news items per day. The media monitoring platform will be able on demand to transcribe, translate, aggregate, write abstractive summaries, classify, and extract knowledge in the form of entities, relations and topics and present all this to the user using new visualizations and analytics over the data. The contextual models will also be applied to a **News Production Tool** (Use Case 2), where enriched models for transcription (ASR) and translation (MT) will be used. This will give journalists in an operational editorial environment a multilingual tool that will be able to learn continuously over time.

Table of Contents

Table of Figures

Table of Tables

2. Introduction

This document describes the use case definition and user requirements for the SELMA project, which covers two use cases: a media monitoring platform and a news production tool, and several use case applications within those areas.

It includes the definition of the use cases which will serve as a basis for the personae, workflows and scenarios. User and technical requirements alongside the Key Performance Indicators (KPI) are extracted from the previous sections.

This deliverable contains 8 sections: the definition of the use cases, the description of personae, the workflow diagrams, the scenarios, the user and technical requirements, the KPIs and finally, the conclusion.

3. Use Case Description

We look at two main use cases, i.e., Media Monitoring and News Production, as well as four use case applications covered within those two principal use cases. The use case applications, in which the technologies and components developed in the project will be implemented, focus on specific challenges and direct usage in the media industry.

3.1 Media monitoring

The *Media Monitoring* use case focuses on the observation of web feeds with media content. In most cases, these are RSS feeds or website site maps which give access to written articles, video and audio content in multiple languages.

The user selects from available feeds and assigns them to a group, saved for easy reference and subsequent use. This allows the system to aggregate the media items contained in these groups. For each group, an automatic analysis identifies clusters of media items that are related by topic and other common attributes. This provides a visualization of the contained topics and entities along various facets, for instance how a cluster evolves over time or how gender is distributed.

An important feature is that the clustering works independently of the language, which means that related media items can be linked with each other across multiple languages.

The results of the analysis can be visualized in dashboards and used to trigger alerts. The results are also available via an API for post-processing.

3.1.1 Advanced Content Analysis

This use case application covers advanced analysis of topics of journalistic content and summarization, which require more research. This may include a wide range of topics and themes.

3.1.1.1 *Broadcast Monitoring and Analysis*

Goal

The goal is to enhance internal and external monitoring and analysis activities based on real-life requests from within the broadcast (in particular DW) community, using advanced SELMA research results.

Concept

During the project, DW will apply the SELMA platform to monitor broadcast material, from DW internal sources as well as from other news providers. This covers ingestion and processing of massive data streams, with video, audio and text, including feeds in all 30+ languages from DW. Advanced content analysis supports a more efficient editorial workflow and decision-making process through better information on trending topics and breaking news. It is a major step forward towards breaking the language barrier, as awareness of and access to content from different languages brings the different language departments within DW (and by extension different communities) closer together. In addition, ad hoc, specific analysis requests will be done based on topics of interest reported by DW users, e.g. elections or high-profile events, statements or actions by specific politicians. NLP technology enabling high-quality, reliable filtering and clustering, resulting in cross-language summarized reports allows our editorial staff to work at a next level in a fully customized way.

3.1.1.2 *Diversity*

DW proposes to analyze journalistic content with respect to diversity, a major topic of interest, high on the agenda in the media as well as in the society. It is a complex analysis, thus well-suited for advanced analysis applications.

Goal

Media needs and wants to be representative in their journalistic output.

"An editorial team loses credibility if it doesn't represent diversity in its many forms: gender, ethnicity, sexual orientation, disability and different world views. Credibility is the most important currency in journalism." (Peter Limbourg, Director General Deutsche Welle)

We intend to improve diversity and representation in media by counting numbers and appearances with the help of Text Mining and NLP. This will be done in 30+ languages (using machine translation), covering the range of DW languages.

Concept

We will measure both text and audiovisual content, published on dw.com and beyond. The table below indicates the range of content targeted for this analysis.

Table 1: Targeted content for the diversity analysis

#	What (representation)	How (technology)	Description, Examples
1	Quantity Representation of dimensions in media • Gender (Woman, Man, "Diverse") • BIPOC (black, indigenous and people of color) • People with disability • Age Groups • Sexual Orientation • Social Background Representation in specific roles • as interview partners • as experts • as moderators	Named Entity, NLP, Speaker Diarization, Speaker Recognition	In Germany, people with international backgrounds are 25% of the population. But in news coverage they hardly appear. Some dimensions will be easier to identify than others (e.g. first names give quite an indication of the assumed gender). Other dimensions might only be made recognizable by looking at context, e.g. such as: “is a non-binary artist”, "is a gay activist" or "is focused on accessibility".
2	Quality / Semantics What are people talking about?	Topic labeling, Topic clustering	Content analysis: Clustered along journalistic categories such as: politics, science, culture, sports, ... Example: Right or wrong? Do men only talk about beer and cars? Do women only talk about beauty and clothes?

3	General Diversity Topic Analysis How often are diversity topics addressed?	Topic labeling, Topic clustering	Gendered and diversity topics: Gender, Feminism, Racism, Colonialism, Ableism

Automatic reporting will make the analysis output available in the form of summarizations and graphical visualizations.

3.1.2 Press Agency Analysis

Goal

This use case application, led by Priberam, targets a particular user set of Media Monitoring platform users, Press Agencies (PRs). Typically, these users serve a set of clients, operating in different areas of business.

Concept

To specify the monitoring activities, the user sets their preferences, which may include which topic areas should be monitored (specific entities/organizations, different topics - sports, politics, economy, etc.), which sources can be shown in their original languages, and which should be displayed in English (default) or another chosen language. All these settings are saved in a so-called “View” - in other words: how you look at things - which can be reviewed and adapted at any time in a dedicated area.

Based on these settings, the system provides search results and other visualizations which the user can view either as a larger cluster or as individual items. Both clusters and items can be edited or adapted while working with them: users can mark items as favorites, save or curate them, provide feedback to the system, mark them as read or unread, get translations or put them through the fact checking process. These adaptations and configurations are used to train the system to improve following activities of the monitoring process.

Based on all these activities, the user can produce various outputs, like reports, dashboards, etc., and translate them into any desired language supported by the system. These outputs can either be shared internally, e.g. with colleagues or the head of department who had asked the analyst or journalist to do the monitoring, or they can be exported to pdf to share with clients.

3.2 News production

The *News Production* use case uses the NLP technologies that are developed in the project to support the creation of media content. The focus lies on the creation of text (summaries for newsletters and video subtitles) and synthesized speech (video voice-over and audio podcasts).

In all cases, the Media Monitoring component can provide input material. This is especially true for the newsletter scenario, where relevant news material needs to be identified and summarized first.

However, it is also possible to upload input material directly into the system for further processing. This is especially true for the video subtitling and video voice-over scenarios, where it is sometimes necessary to process hand-selected videos. It also applies to the Podcast translator scenario, where selected podcasts are translated into other languages using speech synthesis.

We will experiment with different user interfaces to collect user corrections of the corresponding SELMA components. We will also experiment with the UI to show the automated post-processing suggestions (in WP3) and to accept or reject user suggestions, including hints on how the correction should be used for learning (e.g. topic-dependent, one-time only, global), We will explore how to benefit from user suggestions at multiple levels – (1) on the document level, the UI should show corrections based on user correction of terms in the same file and (2) on the user level, for other documents edited by the same user and (3) globally to improve the overall suggestions to all users.

3.2.1 News Podcast Creation

This section proposes a use case application where audio news bulletins are produced with the support of SELMA. This use case application is optional and “system should be able to” wording in this section should be interpreted as optional functionality.

Goal

This use case application is based on a workflow observed in DW's Brazilian language department. The goal is to increase the workflow's efficiency by supporting the journalist in the production of daily audio news bulletins through SELMA.

The Reuters Digital New Report for 2020 highlights the rise of popularity of news podcasts over the last two years. One of the most well-known daily news podcasts is *The Daily* by the New York Times which attracts 2 million daily listeners. Based on this trend, DW's Brazilian department started its own daily news podcast in August 2020. While the monthly usage rises consistently with more than 50.000 impressions in December 2020, it becomes equally clear how resource-intensive its production is.

DW expects that the demand for news podcasts in other languages will rise. Consequently, it is highly desirable to be able to produce audio news content in one of DW's other 29 (soon 31) languages with a short ramp-up time and with minimal personnel effort.

Concept

We want to build an application to support journalists in the Brazilian department to produce daily audio news bulletins by automating some of the tasks involved. The Brazilian News Bulletin, on average 6 minutes long, is structured as follows:

Introduction, followed by 5 news stories, followed by an outro.

The editorial department scans several sources (both internal to DW as well as other broadcast or press agency material) to compile the stories. Examples are: https://www.dw.com/pt-br/not%C3%ADcias/s-7111, https://www.dw.com/en/germany-third-conservative-mp-quits-as-pressure-grows-over-lobbying-scandals/a-56843514

Targeted Platforms

The bulletins are delivered to 8 platforms:

Table 2: Targeted platform for the news podcast creation

Platform	**Access**
Spotify	Link to spotify.com
DW's Media Center	Link to dw.com
Apple Podcast	Link to Apple Podcast Platform
Google Podcast	Link to Google Podcast Platform
YouTube	Link to YouTube playlist Link to YouTube RSS-Feed
Amazon Alexa	Link to Alexa Store
Google Assistant	Link to Google Action Store in Brazilian
Facebook Messenger	Link to Facebook Messenger

Production Steps

The manual production of a single news bulletin can be subdivided into the following steps:

Table 3: Production steps of a single news bulletin

Step	**What**
1	Research 5 stories
2	Write 5 stories
3	Check stories by colleague
4	Record, edit, upload into the system
5	Add metadata in CMS, create YouTube video, publish on YouTube

6	Create bit.ly links and publish on Social Media (Twitter & Facebook)

Where could SELMA support the production?

Below is a proposed overview of the workflow, involving a number of AI components integrated into SELMA, that could be used to support the journalist with the aim to increase efficiency. Most likely, SELMA needs to be coupled with a variety of other systems, such as

- Podcast production platforms
- feed generators
- the DW CMS
- Social Media platforms (YouTube, Instagram, TikTok, ...)

It remains to be defined which steps should be carried out by SELMA and which steps by external systems. Once the split-up is defined, the appropriate interfaces need to be looked at, including the associated APIs.

Step 1: Research

- SELMA should be used to identify a configurable number of suitable storylines and propose them to the journalist.
- The individual storylines should be identified
    - by a title,
    - an abstract summary,
    - a list of the sources that report on the storylines
    - statistics that illustrate why the storyline is judged to be relevant
- The storylines should be ranked by relevance
- The considered time period should be configurable, e.g. 'the last 6 hours' or 'the last 12 hours'
- The system should be able to judge the newsworthiness of considered storylines based on several observations, amongst them:
    - the number of related stories published during the observation period
    - the change (first derivative) of the number of related stories published during the observation period with respect to preceding observation periods
- The journalist should be able to choose the storylines she or he wants to include in the news bulletin

Step 2: Writing stories

- For each of the selected storylines, the system should be able to produce a first draft of a related news report using summarization techniques
- The draft news report should be configurable:
    - number of words / duration if read out aloud by an average news reader / by average news readers
- the type of news report - should be read by a single news reader in a more formal (neutral) style or possibly by two readers in a more casual style
- The draft report should be editable by one or more journalist(s)

Step 3: Checking Stories

- The journalist should be able to store the edited text in a structured format at a central location where colleagues can provide feedback
- Once the texts are checked, the journalist should be made aware of that fact

Step 4: Recording, editing, upload into the system

- The journalist should be able to choose whether to record the bulletins with her or his own voice or automatically, via voice synthesis
- The system should be able to add intro, background and outro music automatically, possibly in conjunction with an external podcast production system
- The journalist should be able to edit the result, for example to adjust the individual levels and duration of the segments
- The export of the final audio track should be possible with the touch of a button

Step 5: Add metadata in CMS, create social video, publish

- The system should generate automatic metadata where possible:
    - title
    - short description
    - playlist with timestamps
    - publication date
- The system should be able to publish to various platforms automatically, via RSS feeds that it generates

- The system should be able to produce a video containing the audio track with a single fixed image
- As an alternative, the system could offer the possibility to add an image to each segment of the news podcast and create the corresponding video
- The system should be able to upload the video to YouTube
- As an extension, the system could be able to produce additional videos with short duration (<= 1 minute or even 30 seconds) for platforms such as TikTok, or Instagram Reels
- Another extension would be to produce image tiles with audio, to be assembled into Instagram Stories

Step 6: Create Social Media Links

- The system could create template texts that could be posted on Social Media (Twitter & Facebook) via copy & paste

This application will allow journalists to automate some of the steps, standardize and make the process much more efficient, while keeping full editorial control over the final output.

3.2.2 Video Subtitling

This section describes a DW use case application where SELMA output is further processed as subtitles.

Goal

The system creates subtitles based on the audio track, potentially translates them and makes them available for manual editing. Subsequently, it applies them to the video.

Concept

We start from content that comes from the SELMA monitoring use case. In that output, the user may find a video of interest for further publication with subtitling in source language or in another language.

Step 1: Select video item

- The system allows the user to select one or more videos from the SELMA monitoring output (use case 1) – or possibly from another source- and send it to a module that allows further processing of selected items

Step 2: Apply ASR and manually correct transcription in original language

- The user applies automated speech recognition (ASR) or uses the ASR output that was used during the monitoring phase. The system may use third-party services for this
- The user corrects the ASR output manually. This provides a correct text transcription in the source language
- User corrections are analyzed by the system and used to train the system at different levels

Step 3: Select target language, apply MT and manually correct translation

- The user selects a target language of his/her choice from the languages available in the system
- The system applies machine translation (MT). The system may use third-party services for this
- The user corrects the MT output manually. This provides a correct text translation in the target language
- User corrections to the MT output are analyzed by the system and used to train the system at different levels

Step 4: Apply subtitling

- Subtitles are produced using a separate module
- The transcription is shown as synchronized subtitles in the source language
- The translation is shown as synchronized subtitles in the selected target language(s)

3.2.3 Video Voice-Over

This section describes a use case application where SELMA output is further processed as voice-over.

Goal

The extracted and edited subtitles are converted into an audio track via text-to-speech. The audio track is subsequently added to the video.

Concept

We use the same procedure as in the subtitling use case application, and follow the steps outlined there. In addition, we use a synthetic voice to produce a voice-over. The objective is to make the speech output in the target language as expressive and natural as possible, resembling the expressiveness of the original version.

A possible extension of this use case application is to apply the speech-to-speech translation module that is part of experimental research in WP3.

Step 1: Select video item

- See subtitling use case application for details

Step 2: Apply ASR and manually correct transcription in original language

- See subtitling use case application for details

Step 3: Select target language, apply MT and manually correct translation

- See subtitling use case application for details

Step 4: Apply subtitling

- See subtitling use case application for details

Step 5: Apply voice-over

- Voice-over is produced through a separate module using speech synthesis
- The user selects the preferred voice from the list of available voices
- SSML (speech synthesis markup language) may be applied to further correct and optimize the synthetic voice output
- An audio or video file can be downloaded with the target voice output

4. Personae

This section presents the personae applicable to the SELMA project in relation to broadcast monitoring and news production, as well as press agency monitoring. This section details the personae, their role and relationship.

4.1 Broadcast monitoring and news production

In this section, we focus on use cases as applied to DW. Five personae have been identified by DW and defined (see table below).

Table 4: DW Personae relevant to SELMA

Persona ID	**Role**	**Organization**	**Relationships**
DW P1	Editor-in-Chief	Programme Directorate	Leads overall DW production team
DW P2	Head of Region	Regional Department	Represents DW language departments belonging to a particular region
DW P3	Head of Language Department	Language Department	Leads a specific DW language department
DW P4	Journalist	Language Department	Reports to Head of Language Department Produces content for a specific language requirement

DW P5	Editor	Language Department	Writes, edits or reviews content

4.2 Press Agency monitoring

In this section, 6 personae have been identified by Priberam and defined (see table below).

Table 5: Priberam Personae relevant to SELMA

Persona ID	**Role**	**Organization**	**Relationships**
PBA P1	Sales Rep	Press Agency	Sells services (mainly to Media Publishers)
PBA P2	Head of Business Development	MMO	Monitors company, competitors and business-related subjects
PBA P3	CEO	SME	Monitors company, competitors and business-related subjects
PBA P4	Head of Marketing and Commerce	Medium/Large Company	Monitors company, competitors
PBA P5	Analyst	Government agency	Monitors a number of industries and markets
PBA P6	Communication Manager	Industrial cluster	Monitors specific industry for associates

5. Workflows

This section presents the workflow diagrams for the two main use cases, media monitoring and news production. These provide an overview of the process for each case. The diagrams are divided into system and user actions.

5.1 Media monitoring

In the Media Monitoring use case, the system ingests preselected feeds and performs the necessary steps to provide the user with a result in line with his preferences. These steps include NLP processing of the media items, clustering and indexation. Resulting data is used for returning search results, alerts or translation of individual items.

The user can access items through different user interfaces/visualizations:

- *Storylines*, where news items are aggregated into clusters
- *Trending,* where the user is presented with the trending topics based on filters (date, feed, etc...)
- *Panels,* where the user can curate and discover relevant documents
- *Search/Views*, where the user can search and filter through the contents
- *Dashboards* where the system returns a series of visualizations to the user based on specific parameters

The user can then select any individual item, access its details and perform several actions on the item (e.g., mark as read/unread, translate and save). Finally, the user can produce an output under the form of a report, a dashboard, a *view* or a *panel.* The output can then be shared within the platform, exported to other systems or to the News Production use case.

This workflow is illustrated in more details in Figure 1.

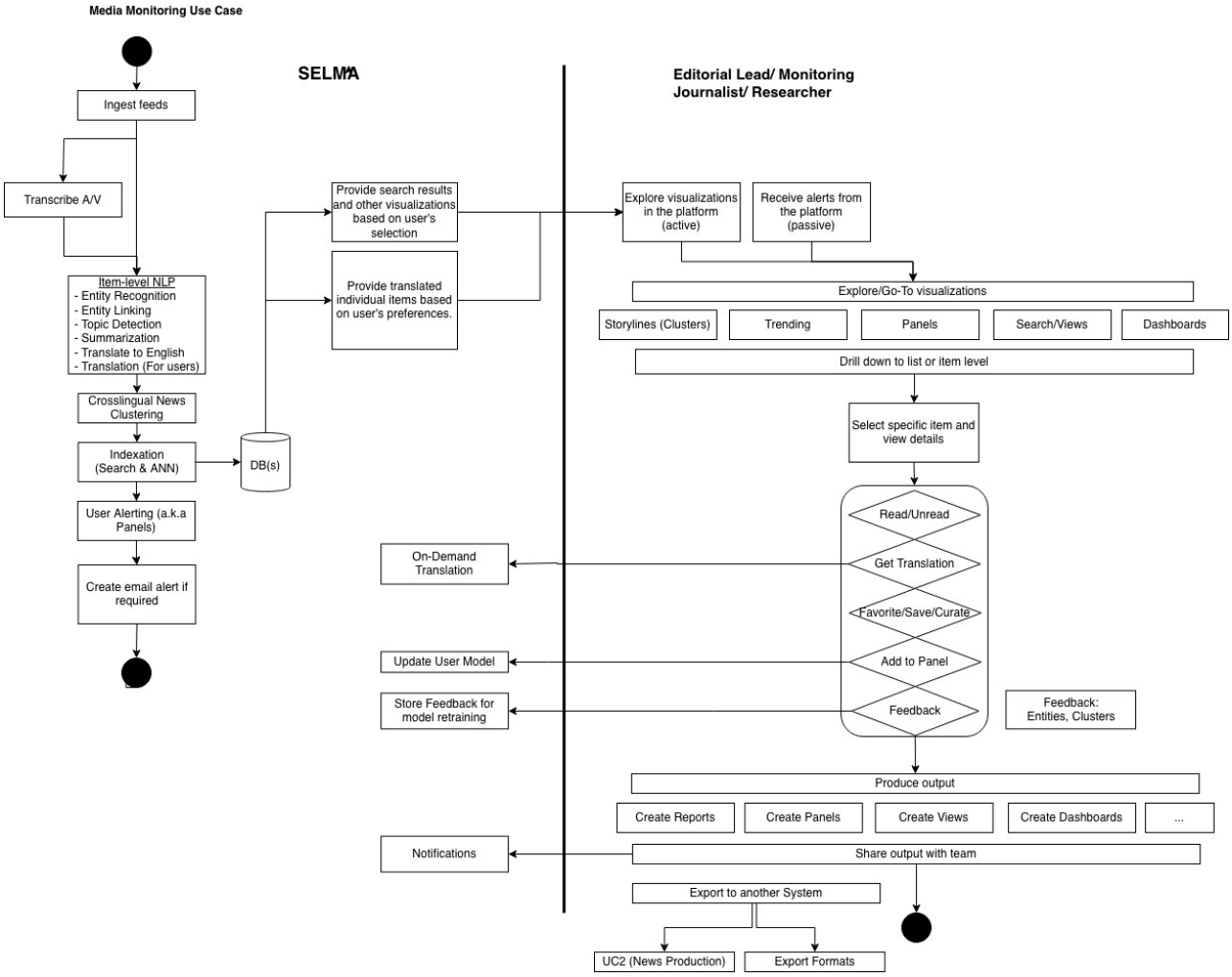


Figure 1: Workflow diagram for the UC: Media Monitoring

5.2 News production

The workflow for the News Production use case is built upon the workflow of Media Monitoring use case, described in the previous section. Here, the user can decide to upload an item or to select an item from the Media Monitoring use case for editing. The user can then request a transcription, translation, subtitling and/or voiceover for an individual media item. The user can edit the transcription, translation and subtitling text. Preferences regarding the engines used, options and voice-over customization can also be selected. Finally, the user can download the media item. Edits made on the transcription and translation text are used for re-training purposes by the system.

These steps are illustrated in more detail in the workflow diagram in Figure 2.

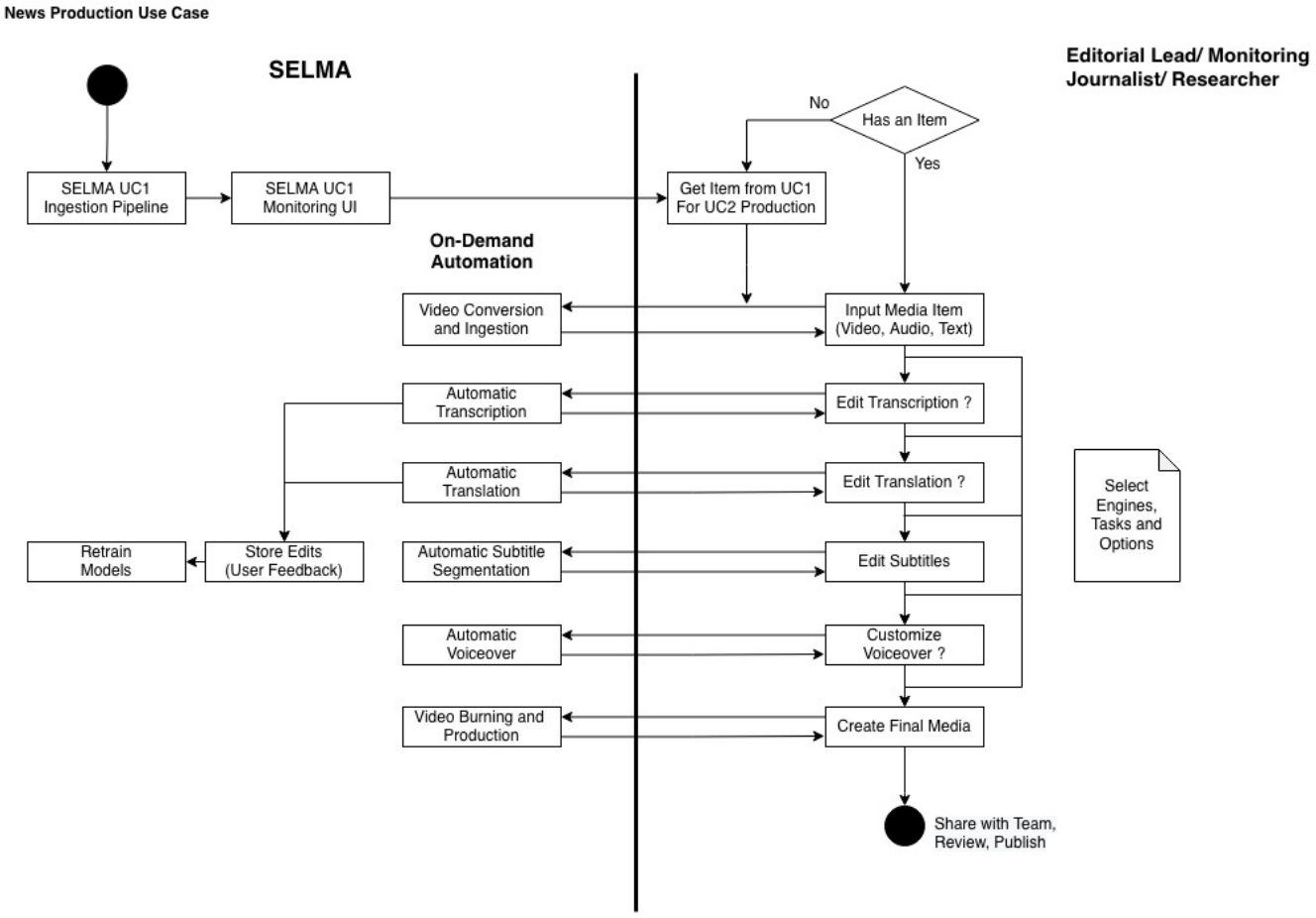


Figure 2: Workflow diagram for the UC: News Production

6. Scenarios

Based on the personae and workflow descriptions described previously, there are several functional areas that can be identified as being relevant to SELMA. The Scenario Model defines these scenarios and their interaction with both SELMA and the relevant personae, whilst the individual scenario descriptions define the functional path within each of these scenarios.

6.1 Overview

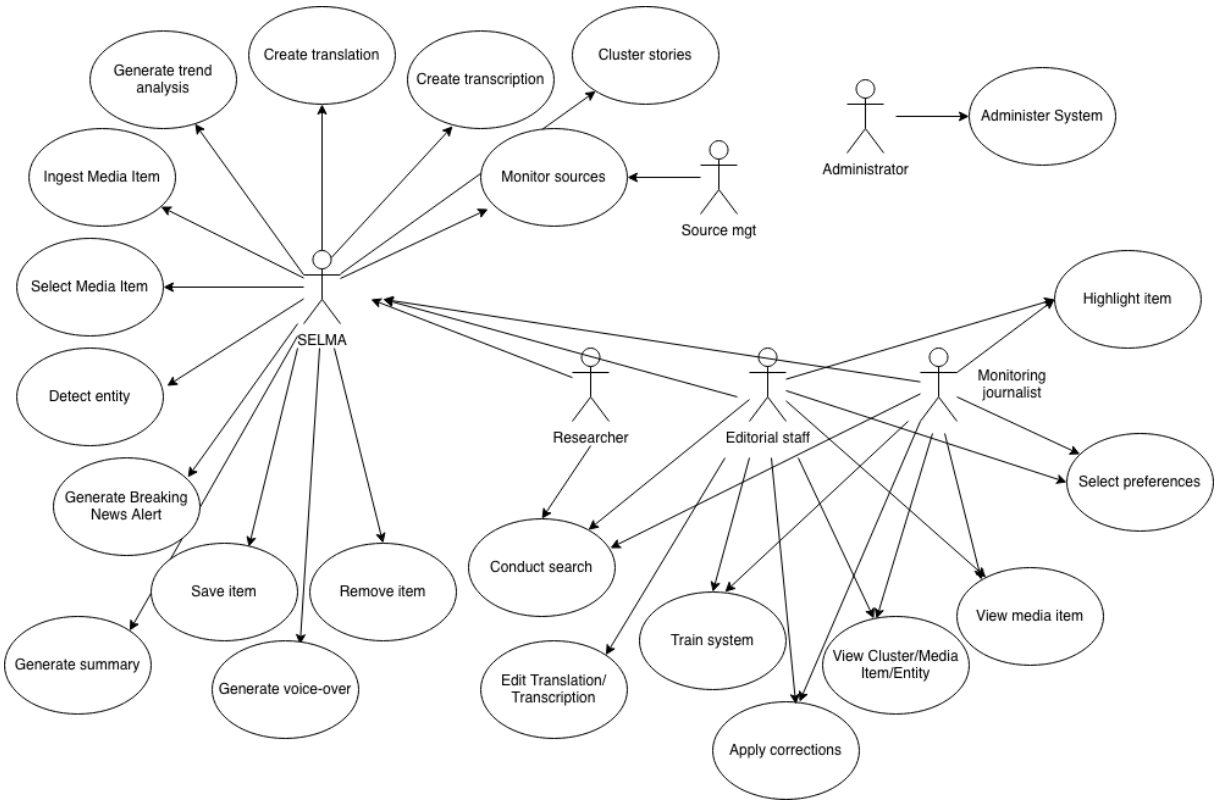


Figure 3: Scenario model

6.2 Detailed descriptions

Detailed in this section are 22 user scenarios, each describing a specific activity/step in the targeted use cases. The individual functional scenarios are defined as follows:

1- Monitor Sources
2- Ingest Media Item
3- Select Media Item
4- Detect and Link Entity
5- Generate Breaking News Alert
6- Create Transcription
7- Create Translation
8- View Cluster/ Entity
9- View Individual Media Item
10- Select Preferences
11- Conduct Search
12- Save Cluster/ Individual Media Item
13- Remove Item
14- Train System
15- Highlight Item
16- Generate Trend Analysis
17- Administer System
18- Group Media Items into Clusters
19- Generate Summary
20- Generate Voice-Over
21- Edit Transcription/ Translation
22- Apply Corrections

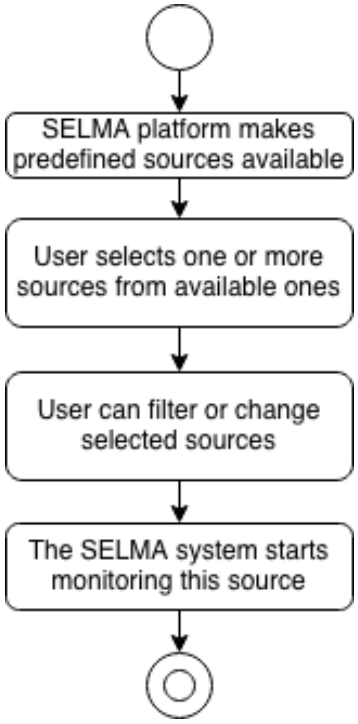
Scenario name	Monitor Sources	
Unique reference	SEL-Sc-001	
Scenario Summary	The user and language team specifies the input source(s) they wish to monitor through the system	

Figure 4: Scenario 001 - Monitor Sources

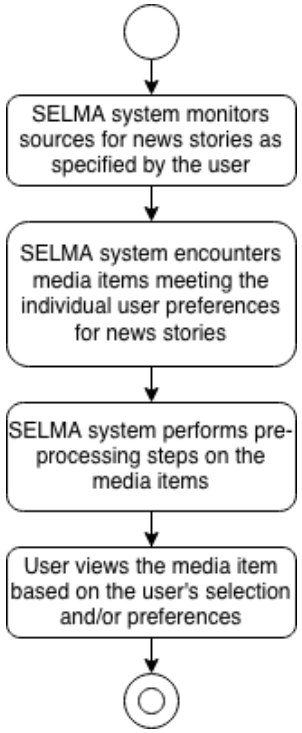
Scenario name	Ingest Media Item	
Unique reference	SEL-Sc-002	
Scenario Summary	The system ingests media items from the sources	

Figure 5: Scenario 002 - Ingest Media Item

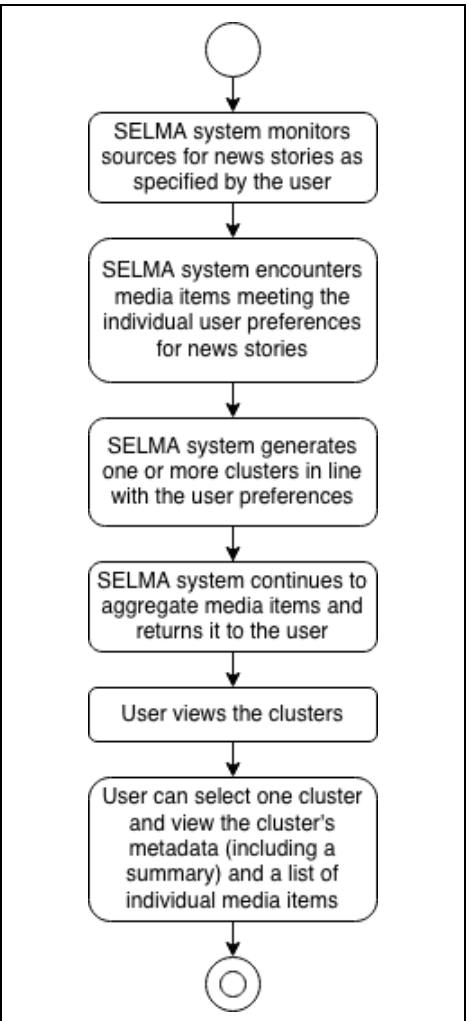
Scenario name	Select Media Item	
Unique reference	SEL-Sc-003	
Scenario Summary	The system selects media items and shows them to the user based on the specified preferences	

Figure 6: Scenario 003- Select Media Item

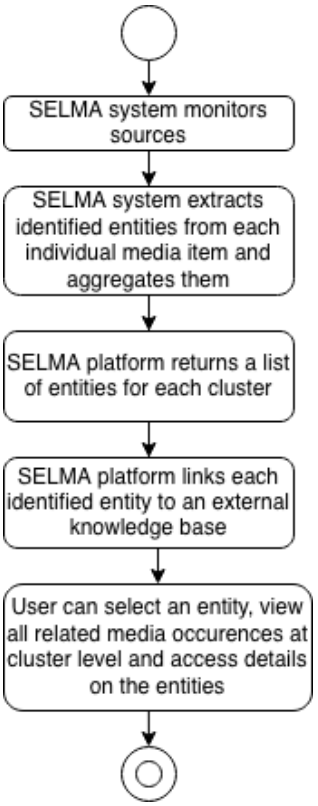
Scenario name	Detect and Link Entity	
Unique reference	SEL-Sc-004	
Scenario Summary	The system detects an entity and links it to other media items and/or clusters from the sources being monitored based on preferences specified by the user	

Figure 7: Scenario 004- Detect and Link Entity

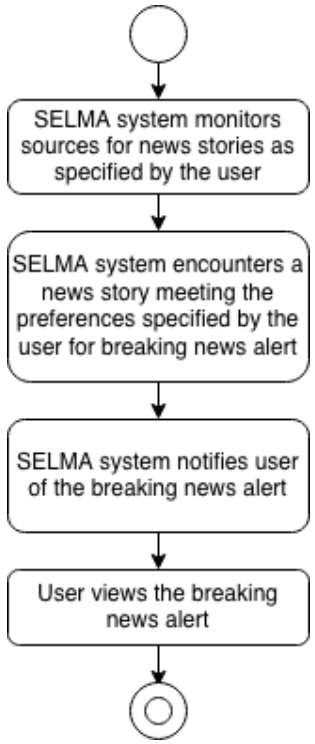
Scenario name	Generate Breaking News Alert	
Unique reference	SEL-Sc-005	
Scenario Summary	The system generates a breaking news alert based on the preferences set by the user	

Figure 8: Scenario 005- Generate Breaking News Alert

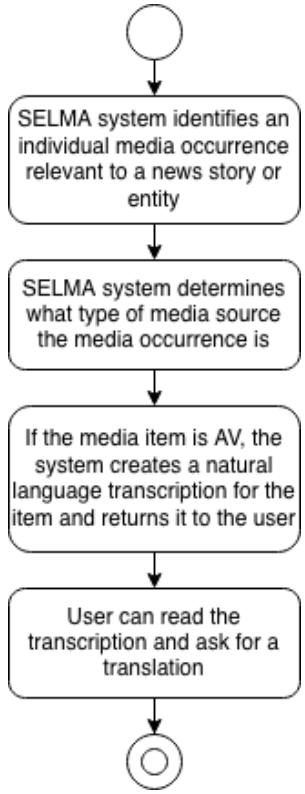
Scenario name	Create Transcription	
Unique reference	SEL-Sc-006	
Scenario Summary	The system creates a transcription for an individual AV media item	

Figure 9: Scenario 006- Create Transcription

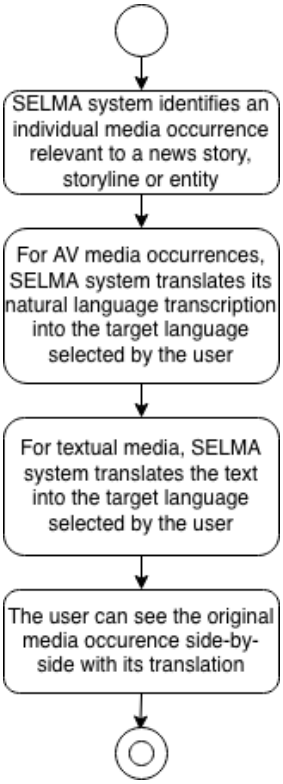
Scenario name	Create Translation	
Unique reference	SEL-Sc-007	
Scenario Summary	The system creates a translation for an individual media item	

Figure 10: Scenario 007- Create Translation

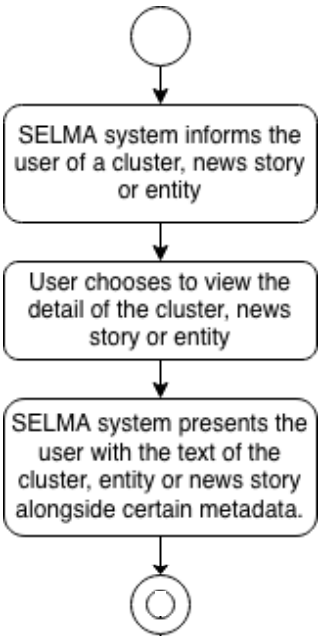
Scenario name	View Cluster/Entity	
Unique reference	SEL-Sc-008	
Scenario Summary	The user views the details of a cluster and/or an entity	

Figure 11: Scenario 008- View Cluster/Entity

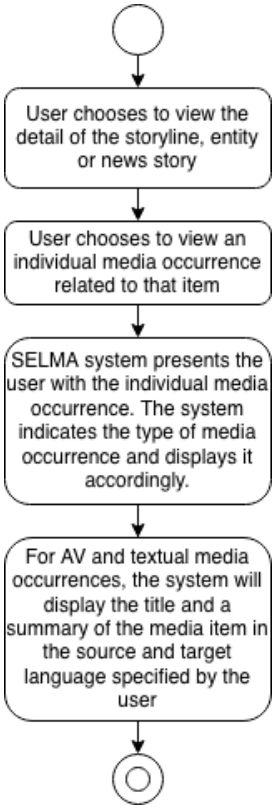
Scenario name	View Individual Media Item	
Unique reference	SEL-Sc-009	
Scenario Summary	The user views an individual media item in relation to a cluster or entity	

Figure 12: Scenario 009- View Individual Media Item

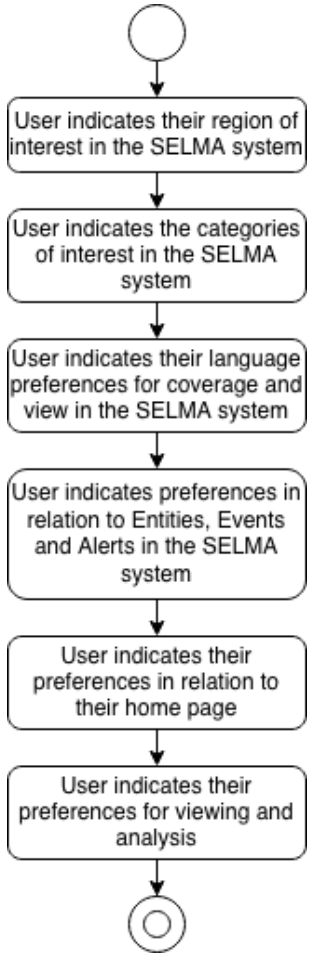
Scenario name	Select Preferences	
Unique reference	SEL-Sc-010	
Scenario Summary	The user sets their preferences in the system	

Figure 13: Scenario 010- Select Preferences

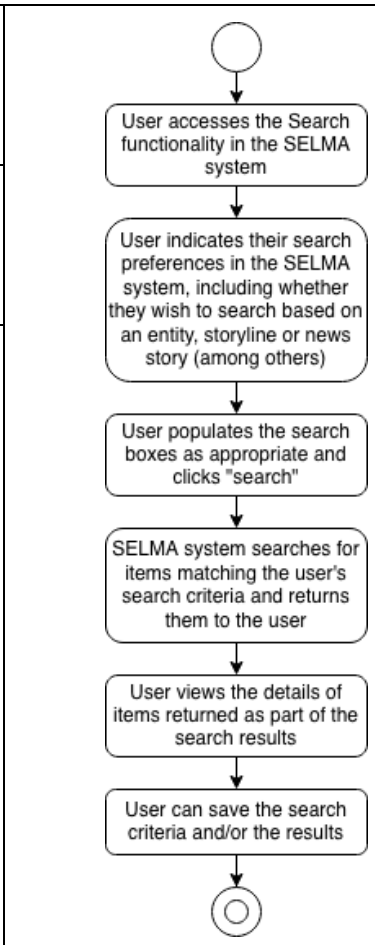
Scenario name	Conduct Search	
Unique reference	SEL-Sc-011	
Scenario Summary	The user can search for an item in the system	

Figure 14: Scenario 011- Conduct Search

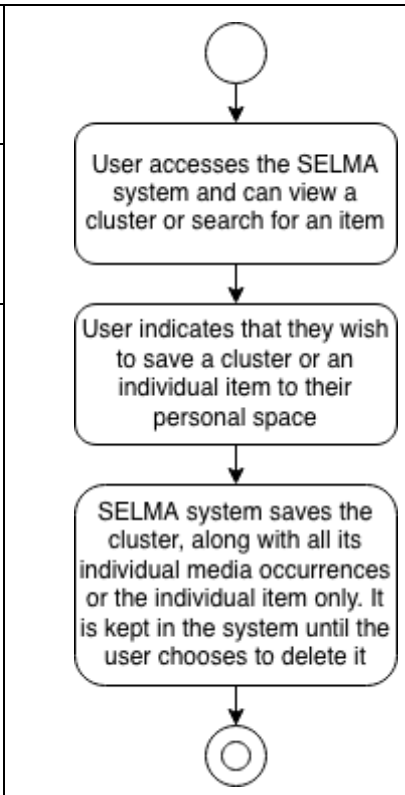
Scenario name	Save Cluster/Individual Media Item	
Unique reference	SEL-Sc-012	
Scenario Summary	The user can save a cluster or an individual media item in the system where it is stored for more than predefined set of time	

Figure 15: Scenario 012- Save Cluster/Individual Media Item

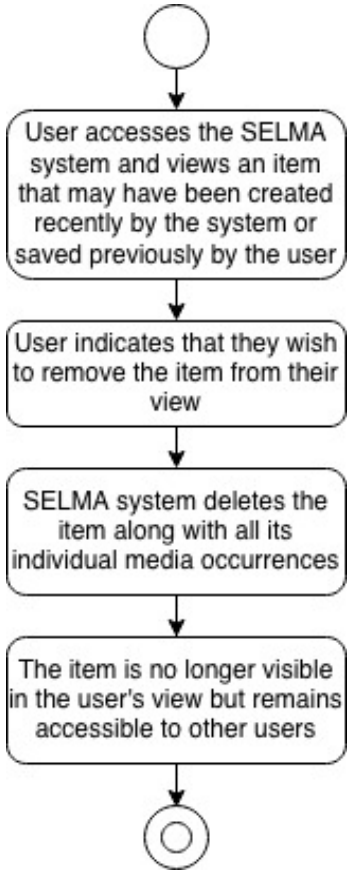
Scenario name	Remove Item	
Unique reference	SEL-Sc-013	
Scenario Summary	The user can remove an individual item and/or a cluster (with all its associated media items) from their view	

Figure 16: Scenario 013- Remove Item

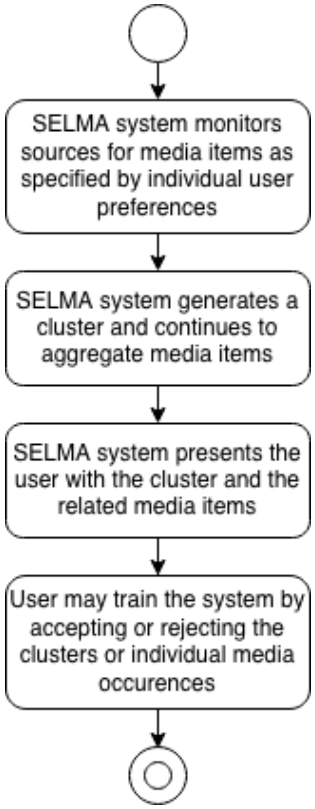
Scenario name	Train System	
Unique reference	SEL-Sc-014	
Scenario Summary	The user can train the system in relation to the cluster generation	

Figure 17: Scenario 014- Train System

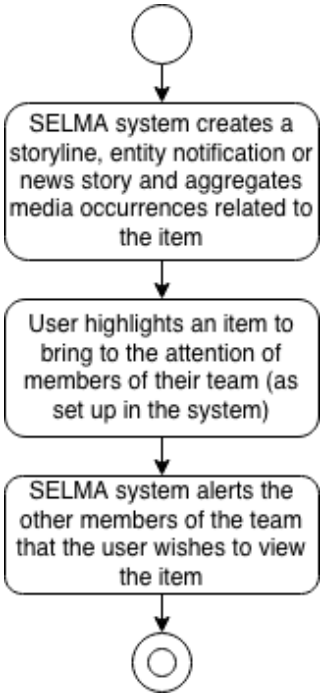
Scenario name	Highlight Item	
Unique reference	SEL-Sc-015	
Scenario Summary	The user can highlight an item to make it visible to other members of their team	

Figure 18: Scenario 015- Highlight Item

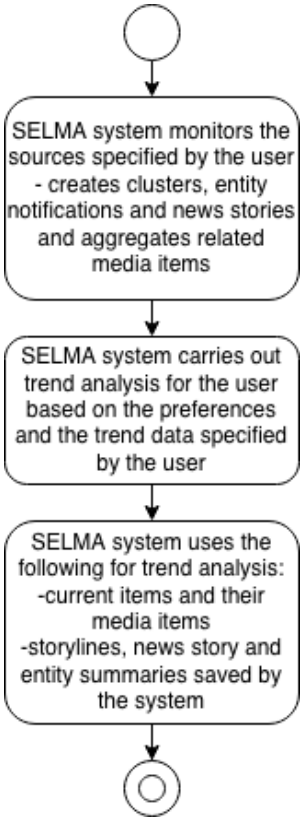
Scenario name	Generate Trend Analysis	
Unique reference	SEL-Sc-016	
Scenario Summary	The system carries out trend analysis and presents the results to the user	

Figure 19: Scenario 016- Generate Trend Analysis

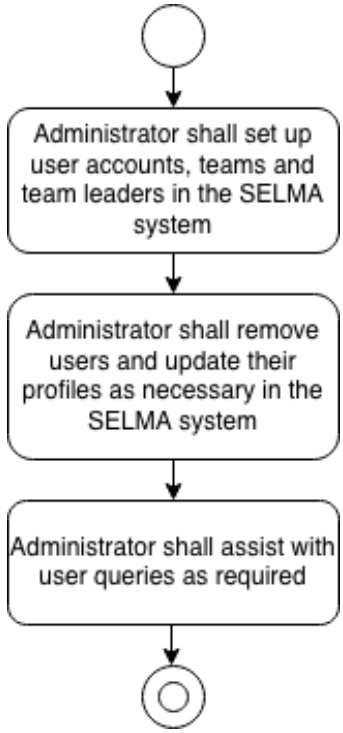
Scenario name	Administer System	
Unique reference	SEL-Sc-017	
Scenario Summary	The System Administrator carries out various activities to administer system	

Figure 20: Scenario 017- Administer System

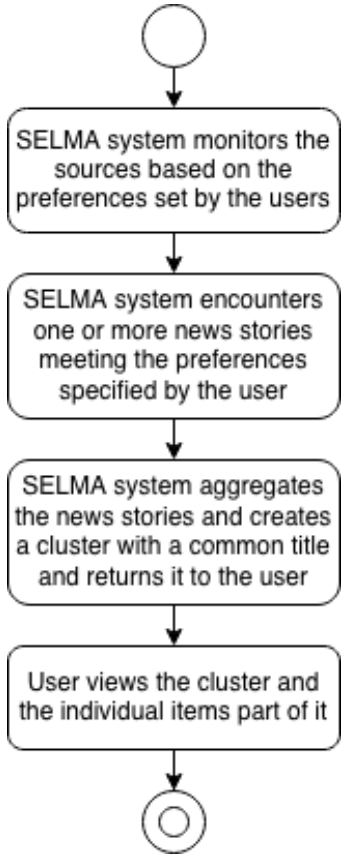
Scenario name	Group Media Items into Clusters	
Unique reference	SEL-Sc-018	
Scenario Summary	The system clusters media items based on the preferences set by the user	

Figure 21: Scenario 018- Group Media Items into Clusters

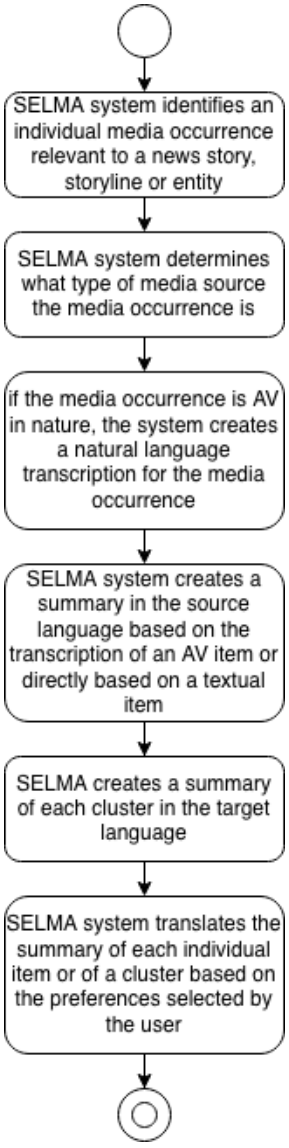
Scenario name	Generate Summary	
Unique reference	SEL-Sc-019	
Scenario Summary	The system generates a summary for each media item	

Figure 22: Scenario 019- Generate Summary

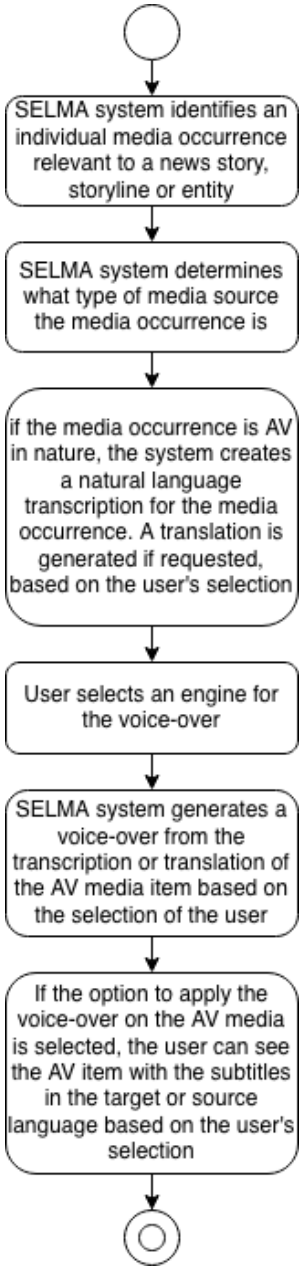
Scenario name	Generate Voice-Over	
Unique reference	SEL-Sc-020	
Scenario Summary	The system generates a voice-over for a transcription and/or translation of a media item on demand	

Figure 23: Scenario 020- Generate Voice-Over

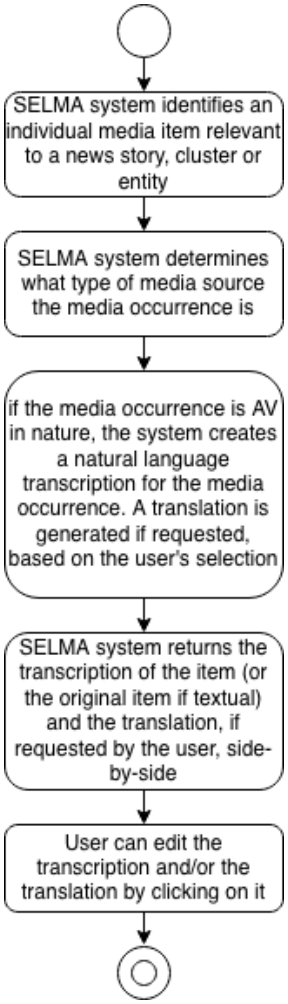
Scenario name	Edit Transcription/Translation	
Unique reference	SEL-Sc-021	
Scenario Summary	The user can edit and correct the transcription and the translation. It is possible for 2 users to edit a transcription/ translation simultaneously.	

Figure 24: Scenario 021- Edit Transcription/Translation

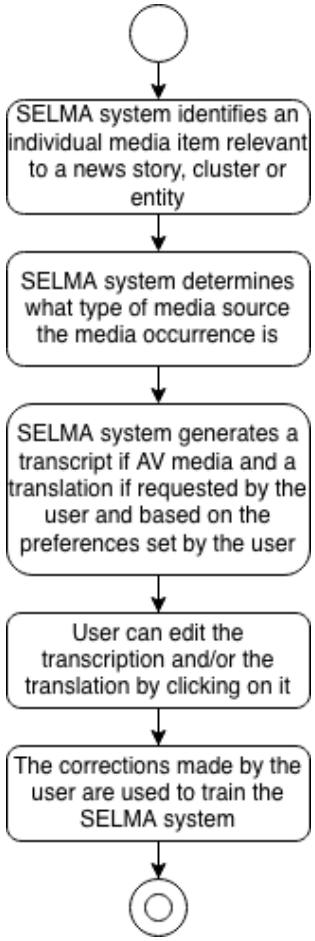
Scenario name	Apply Corrections	
Unique reference	SEL-Sc-022	
Scenario Summary	The system applies the corrections made by the user to the rest of the single media item or its cluster as defined by the user	

Figure 25: Scenario 022- Apply Corrections

7. Requirements

The following tables list the requirements for the SELMA system, as they can currently be foreseen. The overall requirements have been extracted from the workflows and scenarios, and have been divided into User requirements and Technical requirements.

They have been prioritized in collaboration with the technical partners, in terms of need, challenge and feasibility. The MosCow method has been used for prioritization (Must - Should - Could - Won't).

Each requirement has been assigned a value as following:

- *Must:* the requirement is non-negotiable, and the product is not viable without it
- *Should:* the requirement is important, but not vital for the product
- *Could:* the requirement is wanted/ desirable, but is less important than a *Should*
- *Won't:* the requirement will not be delivered

Each requirement has been assigned to the Media Monitoring use case or the News Production use case, as appropriate.

7.1 User requirements

Table 6: User requirements for SELMA

	Requirements	Use Case	MoSCoW
1	The System allows content to be ingested via standard interfaces as used by news organizations where available	1	Must
2	It is possible to use APIs to add content to the SELMA platform	1, 2	Should
3	The System ingests on-demand AV content in MP4 format	1,2	Must
4	The System monitors selected social media feeds	1	Should
5	The System scrapes news articles from websites	1	Must

Module Functionality			
6	The system processes content in the 30+ SELMA languages	1, 2	Must
7	The system transcribes from audio	1,2	Must
8	The system provides statistical analysis of ingested material	1	Must
9	The system provides automated translation of all SELMA languages into English as default	1	Should
10	The system provides automated translation into other SELMA languages upon request	1, 2	Should
Output			
11	The system offers the possibility to create dashboard user interfaces	1	Should
12	The system provides a summarization of individual media items in original language and/or English	1	Must
13	User selects ingested channels to monitor	1	Must
14	User subscribes to notifications when relevant content arrives	1	Must
Content			
15	The system uses clustering technology to group individual media items into over-arching high-level clusters	1	Must
16	The system provides a visualization which contains a list of all high-level news stories that are relevant to the user, according to the preferences they have set	1	Must
17	On detection of a high-level news story, the system provides a default name for this story based on the clustering technology parameters	1	Could

18	The system offers the user the ability to follow a specific story and subscribe to updates regarding that story	1	Could
19	The system allows the user to select a high-level news story and view the individual media items that are relevant to it	1	Must
20	For each high-level news story, the system displays a timeline. The system indicates where each individual media item fits on that timeline	1	Could
21	When viewing an individual media item, the user has the ability to link back to the over-arching high-level news story to which it is related (in order to view the other media items related to that cluster)	1	Must
22	The individual media items for a cluster continue to accrue for a selected period of time. However, the user can indicate to the system that the cluster is no longer relevant before this time has elapsed	1	Could
23	The system retains a record of clusters that it has identified along with an indication of how many media items were identified by the system	1	Must
Entity Requirements			
24	The system uses entity identification technology to group individual media items by entities	1	Must
25	For each selected entity, the system displays a timeline	1	Must
26	The system carries out entity identification (person, organization, locations and events) using the original language and/or English translation	1	Must

27	The system provides a list of identified entities that are relevant to the user, according to the preferences they have set	1	Must
28	The system allows the user to highlight an entity identified by the system and view the individual media items that include the entity	1	Must
29	When viewing an individual media item, the user can link back to the entity to which it is related (in order to be able to view the other media items related to that entity)	1	Could
30	The system retains a record of entities identified along with an indication of how many media items were identified as relating to those particular entities	1	Must
News Story Requirements			
31	The system uses preferences set by the user to detect news stories of interest to the user	1	Must
32	For each cluster, the system displays a timeline. The system indicates where each individual media item fits on that timeline	1	Should
Breaking News Alert Requirements			
33	The system provides breaking news alerts that will correspond to individual media items in accordance with the preferences set by the user	1	Must
34	A breaking news alert consists of a textual description of the associated media item in the original language and/or in English, along with some specified meta-data (such as date and time)	1	Must

35	The user selects their preferences for the type of clusters for which they wish to receive alerts (this may be related to particular event or entity types)	1	Must
36	The user selects the manner and frequency at which they receive event alerts	1	Must
37	Breaking news alerts are as close to live as is technically possible	1	Must
General Functional Requirements			
38	The system monitors all input sources selected by the user	1	Must
39	The user can turn English translation on or off	1	Must
40	It is possible to associate a user with their team in the System	1, 2	Must
41	It is possible to indicate the role of a user in the System	1, 2	Could
42	A user can share a cluster or an individual media item with their team	1	Must
43	Once a user has indicated that a particular news story or cluster is no longer relevant, individual media items relating to that entity or cluster can be removed from the user's view	1	Must
44	The user can flag a particular individual media item, entity or cluster and its related media items and indicate that they wish to save them for future reference	1	Must
45	The user has an option in the system where they can view individual media items, entities or clusters which they chose to save (alongside all the individual media items related to these)	1	Must
Media Item Requirements			

46	The system provides a clear visual indicator as to the nature of an individual media item (social media, blog, website, AV etc.)	1	Must
47	The user can view the detail of an individual media item (when applicable)	1	Must
48	For an individual AV media item, the user views the video and its original transcription, a translation in the prespecified languages (where applicable) and the meta-data associated with the item	1, 2	Must
49	For individual AV media items, the system supports a player and editor with tools to 'scrub' through the video, rewind and download	1, 2	Must
50	For individual AV media items, the system supports a player and editor with tools to select elements to 'clip'	1, 2	Could
51	It is possible for the user to 'clip' an individual AV media item by means of text selection from the transcript	1, 2	Could
52	For other media occurrences (i.e. textual), the user views the text, its translation in the prespecified language (where applicable) and any meta-data associated with the item	1, 2	Must
53	The system provides a 'confidence level indicator' which will indicate how strongly an individual media item relates to the existing or suggested clusters	1	Must
User Preference Requirements			
54	It is possible to set up a set of default sources that will be frequently used by a particular team	1, 2	Must
55	The system contains a predefined list of sources by region	1	Must

56	The user can specify entities of particular interest to them	1	Must
57	The user can choose their region of interest in the System	1	Must
58	The user can prioritize countries of interest within their region of interest	1	Must
59	The system contains a predefined list of regions and countries	1	Must
60	It is possible for the user to set time parameters in the system using an indicator such as a time slider to indicate the time period in which they are interested	1	Must
61	In general, the system supports input of preferences in a number of ways: From predefined lists, using data being encountered in the system (the system will create new entities, events etc.), using free-format text (i.e. search boxes)	1	Must
62	The system is adaptable and configurable to the user's preferences	1	Should
Administration Requirements			
63	The user can log into the system	1, 2	Must
64	The user can log out of the system	1, 2	Must
65	The system supports a super user account	1, 2	Must
66	The system supports an administrative user (for account management)	1, 2	Must
67	The administrator can create teams in the system	1, 2	Should
68	The administrator has typical administrator rights including add new user, remove users, update user profiles as well as the ability to manage data held by the System	1, 2	Should

Search Requirements			
69	It is possible for the user to conduct a search based on an entity or entities	1	Must
70	It is possible for the user to search based on the type of the media item (e.g. social media, AV etc.)	1	Must
71	It is possible for users to search based on event types	1	Won't
72	It is possible to take a screen shot (or still frame) from an individual AV media item (rights to be considered here)	1, 2	Could
73	Entity search is able to handle a range of variable spellings for the same entity	1	Must
74	It is possible to train the system in relation to alternative spellings for searches. For example, it should be possible to link to alternative spellings and indicate that they relate to the same thing	1	Won't
Input Source Requirements			
75	The system informs the user if a source stops broadcasting	1	Could
76	The system informs the user if the frequency at which a channel is broadcast, changes	1	Could

Trend Analysis Requirements			
77	The user has the ability to set their preferences in the system with regard to the types of trend analysis they wish to see in the system	1	Must
78	The system offers the user various options around trend analysis including maps incorporating hotspots, graphs and timelines showing hotspots	1	Must
79	The system utilizes saved clusters and media items (that the system has saved along with the number of media items) to conduct trend analysis over a period longer than one week	1	Could
Generate Voice-Over			
80	The system generates a voice-over on request for individual AV media items	2	Must
81	The user chooses whether the voice-over is performed in the original or in the selected translation language	2	Must
82	The system provides a list of available synthetic voices for the user	2	Must
83	The user can choose and/or change which synthetic voice is used for the voice-over	2	Must
84	The user can change the synthetic voice per segment	2	Must
85	The user can amend the voice-over output including phonetics, pauses and pitch	2	Could

Edit Transcription/Translation			
86	The user can edit the transcribed text	2	Must
87	The user can edit the translated text	2	Must
88	The user saves the edited versions of the transcribed and/or translated text	2	Must
89	The user changes the engine and perform the transcription again for the whole text and/or by segment	2	Should
90	The user changes the engine and perform the translation again for the whole text and/or by segment	2	Should
System Learning and User Feedback			
91	The system is trained by the user. The system offers the user a selection of news stories and individual media items relevant to a cluster and the user accepts/rejects them as necessary, thus training the system	1	Must
92	The system is trained by the user. The system offers the user breaking news alerts and the user accepts/rejects as necessary – training the system to meet the user's preferences	1	Must
93	The system learns from the user's corrections and apply them throughout the text	2	Must
94	The system applies corrections on different levels (current and future items) based on the preferences set by the user	1, 2	Should
95	The system learns from the user's corrections for entities	1, 2	Must
Diversity Detection			

96	The system identifies the binary gender associated with the author's name (if present) of an individual media item	1	Should
97	The system provides the number of times each binary gender is mentioned in the media items	1	Must
98	The system identifies the gender of the protagonist in each individual media item (if applicable)	1	Should
99	The system provides the number of times each binary gender is mentioned in each topical cluster	1	Must
100	The system provides a visualization of the gender analysis	1	Must
101	The system provides all information pertaining to the diversity data (gender, age, sexual orientation, country of citizenship, medical condition) found on the Wikidata entry of relevant public figures where applicable	1	Must
102	The system identifies the gender associated with the named entities of type person, even if the gender information is not available in Wikidata (or the entity is not in Wikidata at all)	1	Should

7.2 Technical requirements

Table 7: Technical requirements for SELMA

	Requirements	Use Case	MoSCoW
Platform - Orchestration			
P1	Orchestrates processing jobs on the data stream, automatically	1	Must
P2	Orchestrates processing jobs on the data stream, on user request	1,2	Must
P3	Allows listening for job results via push notifications (e.g, web sockets)	1,2	Should
P4	Allows listening for job state changes (errors, job progress)	1,2	Must
P5	Allows consulting the state of a job request / jobs on an item via an API (e.g., REST)	1,2	Must
P6	Accepts new “job workflow” requests, which may entail running several jobs organized in a graph of dependencies, on an item	1,2	Must
P7	Orchestrated jobs are eventually applied, meaning, a job cannot be lost - it is either completed successfully or logs an error	1,2	Must
P8	Orchestration-related configuration changes happen without downtime	1,2	Could
P9	Resilience to the unavailability / downtime of specific workers. Jobs wait until the worker recovers	1,2	Should

P10	The system allows parallel jobs to execute on the same item if they can be run that way according to the orchestrated job graph	1,2	Could
Platform - Worker management			
P11	Processing workers use Docker (or equivalent) containerization for deployment	1,2	Must
P12	The system is prepared for kubectl-compatible deployment in Kubernetes clusters	1,2	Should
P13	The system manages the lifetime of different worker containers	1,2	Could
P14	The system scales the number of worker containers according to the corresponding task queue flux	1,2	Could
P15	Worker-related configuration changes (new workers, worker scaling, etc.) happen without downtime	1,2	Could
Platform - Replication and Sharding			
P16	The system allows replication at the worker level	1,2	Must
P17	The system allows replication at the orchestration controller level	1,2	Should
P18	The system allows replication at the database level	1,2	Should
P19	The system allows sharding at the orchestration controller level	1,2	Should
P20	The system allows sharding at the database level	1,2	Should

Component - Online News Classification and Clustering			
C1	For each ingested news item, the system attributes a cluster	1	Must
C2	For each ingested news item, the system attributes an IPTC topic	1	Must
C3	The system clusters documents in an online fashion, e.g., without having to revisit all past decisions	1	Must
C4	The system clusters documents natively in all 30 SELMA languages	1	Should
C6	The system leverages user feedback on clustering decisions to improve future decisions	1	Should
Component - Summarization			
S1	For each ingested news item, the system generates a summary	1	Must
S2	The system generates summaries natively in all 30+ SELMA languages	1	Should
S3	The system generates summaries either from original text article or video transcripts	1	Should
S4	The system leverages user feedback on summarization results to improve future summaries	1	Should
Component - Machine Translation			
M1	The system translates a textual document by demand	1,2	Must
M2	The system translates a video by demand	1,2	Must

M3	The system translates between all 30+ SELMA languages	1,2	Should
Component - Automatic Transcription			
R1	The system automatically transcribes an ingested video or audio file	1,2	Must
R2	The transcription is enriched by punctuation	1,2	Must
R3	The transcription is enriched by speaker information	1,2	Should
R4	The transcription is enriched by named entity labeling	1,2	Must
R5	The system transcribes all 30 SELMA languages	1,2	Should
Component - Entity Recognition and Linking			
E1	For each ingested news item, the system detects named entities	1	Must
E2	For each ingested news item, the system links named entities to a knowledge base	1	Must
E3	The system links entities natively in all 30 SELMA languages, leveraging crosslingual representations	1	Should
E4	For each ingested news item, the system attributes a gender for each person named entity detected in the news item.	1	Should
Component - Story Segmentation			
G1	Each ingested long audio segment gets split into meaningful units	1,2	Could
G2	Speaker clustering is used to create speaker independent units	1,2	Could

G3	Speaker recognition automatically identifies the original speaker in each segment	1,2	Could
G4	Topic segmentation is used to separate by spoken content	1,2	Could
Component - Voice Conversion Synthesis			
V1	The text-to-speech system automatically produces voices in Latvian, German, and French	2	Must
V2	The text-to-speech system will be improved to better handle foreign words	2	Must
V3	A speech-to-speech translation system works at least on one language pair	2	Should
V4	A speech-to-speech translation system works at least on one language pair and can generate a synthetic voice in the target language close to the natural voice in the source language	2	Could

8. Key Performance Indicators – KPIs

This section presents the KPIs for each main component of the system as identified by the technical partners. They will be used to measure the success of each component and to assess the progress throughout the project.

Table 8: Key Performance Indicators for SELMA

Component	KPI
Platform - Ingestion Scalability	Able to scale up to processing around 10M news articles/segmented video transcripts per day, given the computational resources.
Platform – User Scalability	The system should handle 500 simultaneous users for an installation of the platform.
Platform – Language Coverage	Processing models for Albanian, Amharic, Arabic, Bengali, Bosnian, Bulgarian, Chinese, Croatian, Dari, English, French, German, Greek, Hausa, Hindi, Indonesian, Kiswahili, Latvian, Macedonian, Pashto, Persian, Polish, Portuguese for Africa, Portuguese for Brazil, Romanian, Russian, Serbian, Spanish, Turkish, Ukrainian, Urdu.

Component - Crosslingual Representations and Entity Linking	We will evaluate the representations on downstream tasks that make use of them, such as entity linking and text classification. We seek to outperform scores of the state of the art on standard offline multi-pass methods, improving over 86.6 micro F1 scores on end-to-end linking on the CoNLL YAGo dataset (Kolitsas et al., 2018).
Component - Summarization	We expect gains of at least 5% in ROUGE-1, ROUGE-2 and ROUGE-L F-scores over the state-of-the-art scores and improvements over 5% on factual correctness for abstractive summarization. A common dataset for evaluation is the CNN/DailyMail (Hermann et al., 2015) and New York Times dataset (NYT) (Sandhaus, 2008).
Component - Clustering	We expect improvements of at least 5% in the F1 metric.
Component - Segmentation	We aim at an improvement of 10% relative of the Diarization Error Rate (Bredin 2017) over state-of-the-art segmentation systems such as pyannote or the segmentation provided by Kaldi (Povey 2011), measured on the in-domain data provided by Deutsche Welle.
Component - Speech Machine Translation	The system will be able to offer 2 new language pairs thanks to the use of

	state-of-the-art end-2-end spoken machine translation approaches, especially for low resource speech translation.
Component - Automatic Transcription	The SELMA ASR system will outperform state-of-the-art results on some high resource language benchmarks. For instance, for French, on the ETAPE benchmark dataset, a reduction of at least 5% of word error rate is expected.

9. Conclusion

This document presents an overview of the preparatory work for defining the user requirements for the SELMA project, for the development of the two primary SELMA use cases, i.e., Media Monitoring and News Production.

The different personae describe the actors identified for both use cases by Deutsche Welle and Priberam. Two workflow descriptions are provided for each use case, with detailed diagrams.

A scenario model provides an overview of relationships between actors, activities and the platform. Twenty detailed user scenarios describe foreseen activities through interaction with the SELMA system.

A list of user requirements and technical requirements is the result of the preparatory work as described in this report. Together with the user scenarios, the workflow diagrams and the personae descriptions, these requirements form the basis for the use case applications and prototype development over the three years of the SELMA project. The input of this report (D1.1) will trigger the next phase, i.e., wireframing, mockups and prototyping as part of D1.2, in which these use case applications and scenarios will be implemented. This deliverable will be followed by three other reports in this series: D1.2 (Initial Prototype Report), D1.3 (Intermediate Prototype Report) and D1.4 (Final Prototype Report).