# MeMAD Deliverable

## *D6.1 – Specification of the data interchange format, initial version*

*Version 2.0*

| | |
|---|---|
| Grant Agreement number | 780069 |
| Action Acronym | MeMAD |
| Action Title | Methods for Managing Audiovisual Data: Combining Automatic Efficiency with Human Accuracy |
| Funding Scheme | H2020-ICT-2016-2017/H2020-ICT-2017-1 |
| Version date of the Annex I against which the assessment will be made | 3.10.2017 |
| Start date of the project | 1.1.2018 |
| Due date of the deliverable | 31.03.2018 |
| Actual date of submission | 31.05.2019 |
| Lead beneficiary for the deliverable | Limecraft |
| Dissemination level of the deliverable | Public |

**Action coordinator's scientific representative**
Prof. Mikko Kurimo
AALTO – KORKEAKOULUSÄÄTIÖ, Aalto University School of Electrical Engineering, Department of Signal Processing and Acoustics
mikko.kurimo@aalto.fi

*MeMAD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780069. This document has been produced by theMeMAD project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content*

| Authors in alphabetical order | | |
|---|---|---|
| Name | Beneficiary | e-mail |
| Karel Braeckman | Limecraft | karel.braeckman@ limecraft.com |
| Simon Debacq | Limecraft | simon.debacq@ limecraft.com |
| Harri Kiiskinen | YLE | harri.kiiskinen@yle.fi |
| Lauri Saarikoski | YLE | lauri.saarikoski@yle.fi |
| Wim Van Lancker | Limecraft | wim.vanlancker@ limecraft.com |
| Dieter Van Rijsselbergen | Limecraft | dieter.vanrijsselbergen@ limecraft.com |
| Maarten Verwaest | Limecraft | maarten.verwaest@ limecraft.com |
| Kim Viljanen | YLE | kim.viljanen@yle.fi |

| Document reviewers | | |
|---|---|---|
| Name | Beneficiary | e-mail |
| Tiina Lindh-Knuutila | Lingsoft Language Services | tiina.lindh-knuutila @lingsoft.fi |
| Jörg Tiedemann | Helsinki University | jorg.tiedemann @helsinki.fi |

| Document revisions | | | |
|---|---|---|---|
| Version | Date | Authors | Changes |
| 1.0 | 30/01/2018 | All authors | First version of this deliverable. |
| 1.1 | 16/05/2019 | Dieter Van Rijsselbergen, Kim Viljanen | Major revision of the document, added clarification of methodology, revised user stories, separated metadata requirements and made improvements to prototype specification. |
| 1.9 | 29/05/2019 | Dieter Van Rijsselbergen | Processed internal review of version 1.1. |
| 2.0 | 30/05/2019 | Dieter Van Rijsselbergen | Revision of metadata requirements section and tables and final markup. |

**Abstract**

This deliverable defines the functional and non-functional requirements of the MeMAD prototype system, based on input concerning the tools developed in WP2, WP3, WP4 and WP5 and based on the project's overall use cases from which many requirements will be dictated.
This initial version introduces the methodology that will be followed to construct the project's prototype requirements and exchange format specifications. We then identify relevant stakeholders and processes for the project's prototypes from the media production and consumption chain. We define an initial set of user stories for the prototype system, from which the functional, non-functional and data interchange format requirements for the first version of the MeMAD prototype are then derived.

# Contents

# 1 Introduction

This deliverable represents the first result of work done on task T6.1, formally named the "*Specification of the data interchange formats*", but it also includes preparatory work done to reach this outcome, meaning defining overall prototype requirements, as described in the DoA. As such, this deliverable includes descriptions of the methodology and intermediary steps to reach conclusions on the interchange formats, including use case definitions, even though the title only describes the definition of the interchange formats as its topic.

In this deliverable, the first of three iterations, we define a first set of requirements for the prototype MeMAD platform. The aim of this single platform is to form a coherently integrated system of underlying technical components with a single interface for users to interact with, making it easier to test end-user workflows and to help assess the quality provided by various automated analytics and processing tools. The platform offers a single entry point for audiovisual material ingestion, storage and workflow task dispatching, and provides a centralized metadata store and search index and interface.

This document defines the functional and non-functional (i.e., in terms of quality, processing performance or system resilience) requirements of the MeMAD prototype system, based on input concerning the tools developed in WP2, WP3, WP4 and WP5 and based on the project's use cases from which many requirements will be dictated. In addition to (non-)functional requirements, subsequent revisions of this document will also incorporate specific test scenarios and evaluation criteria to determine the performance of the prototype system.

The structure of this deliverable is as follows.
We explain the methodology followed to obtain the MeMAD prototype requirements and exchange format specifications, and provide details on the context of use for the project, i.e., the relevant media production and consumption process. We list the overall project use cases and then for each, a set of more specific user stories, along with the relevant metadata involved in realizing each user story. From this overview, we then deduct the required sets of metadata that will be exchanged between components of the prototype platform. We then validate the generation or consumption of this metadata by the technical components provided in various work packages by the consortium members. Finally, we define the functionality of the first iteration of the prototype MeMAD platform, including its functional and non-functional requirements.

# 2 Methodology for determining prototype requirements and metadata exchange formats

This section describes the methodology followed for obtaining the MeMAD prototype requirements, of which the functional requirements and the data exchange format definition are a part. Our methodology will be built on two pillars:

1. **As a guiding principle for defining the functional requirements of the project's prototype and its underlying individual components we use the four project use cases (PUCs) defined in the project's DoA.**
   The four PUCs define in broad terms the functional objectives of the project and give us a context to build more detailed functionality specifications from, even though they have been defined in a very generic fashion.
2. **For the definition of actual functional requirements, we follow the Human-centred Design methodology[1].**
   Applying human- or user-centered design (UCD) is a good match for the MeMAD project, as the project aims to build a prototype that will be actively used and interacted with by end users. Not only that, but when using the prototype system these users will need to adapt to changes in the execution of contemporary production processes, because the MeMAD prototype will offer improved or new ways of tackling problems or it will deliver automated solutions of which the results need to be incorporated in existing production processes. Examples of such changes include: users who need to manually correct automatic suggestions for video clip descriptions instead of typing all descriptions manually, or users who are presented with automatically generated transcriptions of interviews in electronic format while they formerly only had this information available in paper print-outs.
   Using the UCD methodology will help the consortium build better user experiences because end users will be involved throughout the design and development, and additionally, designs will be iterated on and refined by user-centered evaluations.

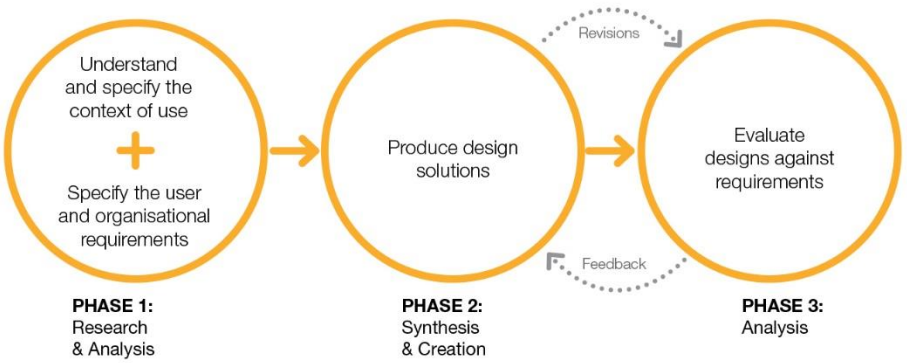


*Figure 1: Synopsis of the User-Centered design process (from O'Grady, 2008[2]).*

[1] Cf. ISO Standard 9241-210:2010 – Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems.
[2] Cf. Visocky O'Grady, J. & Visocky O'Grady, K. (2008) The information design handbook. Mies: RotoVision.

## 2.1 Implementing User-centered design for MeMAD

The execution of UCD in MeMAD will occur in several steps, as illustrated by Figure 1.

1. Phase 1, step 1, is the research phase to understand and determine the context of use across the entirety of media production and consumption process, and from the viewpoint of the various intended stakeholders of the project's results. These stakeholders include:
    - End users who will be using the system first-hand.
    - Other stakeholders who have a stake in the implementation of the system, e.g., producers who manage the budget for the process execution.
    - Technology developers and researchers who need to understand the user requirements in order to implement them, and who need to provide feedback on the feasibility of succeeding in the implementation of user requirements.

    The context of use is further explored in Section 3.

2. In Phase 1, step 2, the actual functional requirements are defined. More detailed user requirement will be defined as user stories to describe more specific sets of desired functionalities, each of them fitting within the definition of one of the PUCs on the one hand, and with the context of use we determine from step 1. We use user stories for this purpose, from an end-user perspective (implying desired functionality from the back-end system indirectly) because they are easy to grasp by project stakeholders, which will facilitate their evaluation. At the same time, they form a good basis to refine further requirements from.
   To ensure the user stories we define are relevant to all project stakeholders, we will validate them in a project review process as follows:
    a. To make sure we cover the entire spectrum of possible applications for this project and the project's use cases, we will consider the media production and consumption process end-to-end (as introduced in Section 3) to determine possible innovative functionalities that the MeMAD project can provide. To give us a first baseline to start from and allow for easier discussions, the consortium partners define this first set of stories, aided by industry media professionals employed by consortium partner YLE. These user stories are defined in Section 4 of this deliverable.
    b. While YLE's representatives count as experts in their domain, to avoid undesired bias by including a single organisation's viewpoints, the first iteration of user stories will also be evaluated by members of the project's external collaborators and by contacting professionals through industry channels such as EBU and the EU Mediaroad sandbox project.
       These stakeholders will be provided with a survey where they can indicate their interest in each user story, along with the possibility to suggest additional scenarios for implementation by MeMAD. This survey will allow us to prioritize on those functionalities deemed most imported across the industry participants, remove those that were backed with very little support, and potentially add those functionalities that we didn't consider yet within the project's consortium.
       This revised list of user stories will be reported as part of the next iteration of this document.

3. In Phase 2, based on prioritized and relevant user stories, we can further flesh out the exact requirements involved for each story. The result from this effort will be the following:
    a. Detailed requirements that can guide the development of the prototype and underlying technologies by the consortium. This defines the functionality required from the system to realize the goals stated. This will encompass textual descriptions, as well as visual designs and potentially interactive mock-ups of application interfaces.
    In addition, non-functional requirements will also be deduced from the context of the use (e.g., timing or processing speed constraints, availability constraints, accuracy requirements, etc.) along with criteria to measure the success of each non-functional requirement.
    b. The types of process inputs and outputs (i.e., its relevant metadata) involved in each user story.
    We elaborate on this in Section 5.
    c. The kinds of automated processing, both in the visual and auditory domain that will be required to implement the listed user stories.
    We touch upon this in Section 6.

    This definition will be done using in-depth interviews and interactive design sessions with relevant end users, based on contemporary production processes and tools.

4. Finally, for Phase 3, we will define exact test cases and evaluation procedures to measure the implementation of prototype features against the functional and non-functional requirements defined in Phase 2.

5. Once an initial cycle has been completed from Phase 1 to Phase 3, further iterations between Phase 2 and Phase 3 will occur to tweak the functional requirements of the systems implemented in the project, based on user-centered feedback. This is illustrated in Figure 1.

As explained above, because the UCD process will encompass the entire duration of the project, not all of its results are available in this deliverable yet. We summarize how each piece will subsequently be completed in which deliverable of Work Package 6 in Table 1.

| Interchange format specification and requirements definition | The MeMAD prototype | Evaluation of the MeMAD prototype |
|---|---|---|
| **D6.1:** Definition of the context of use and an initial set of high-level user requirements. In addition, this deliverable maps out a first revision of required metadata and sets the requirements for the first prototype iteration (M3). | **D6.2:** A report on the first implementation of the prototype, executed per the specifications of D6.1 (M12). | **D6.3:** An evaluation of the first prototype and its requirements, to the extent possible with the limited implementation. This report also includes feedback concerning the use cases and requirements for exchange format specifications (M12). |
| **D6.4:** Refinements of the initial set of high-level user requirements based on feedback from external advisors. This second version will define more detailed requirements for the second MeMAD prototype, including test cases and scenarios (M18). | **D6.5:** A report on the implementation of the second prototype, executed per the specifications of D6.4 (M24). | **D6.6:** An evaluation of the second prototype and its requirements (M24). |
| **D6.7:** Definition of the final requirements and test criteria for the MeMAD project prototype, along with final specifications of all metadata exchange formats (M27). | **D6.8:** A report on the implementation of the final MeMAD prototype, executed per the specifications of D6.7 (M36). | **D6.9:** A report on the evaluation of the final MeMAD prototype, which will be done by interested parties outside the project consortium (M36). |

*Table 1: Orientation of MeMAD Work Package 6 deliverables.*

# 3 Processes and stakeholders in the media production and consumption chain

To better understand the context of use for MeMAD technologies, we need to take a closer look at the media production chain, and identify each of its processes, along with the users and other stakeholders who participate in these processes and for which implementing MeMAD applications could make sense.

Two large process phases are relevant for the MeMAD project: the production and consumption of media. The consumption phase can be considered by itself as a single process executed by the consumer end user, who can be a viewer, listener, reader or a combination depending on how the media is delivered.

The production phase on the other hand encompasses many processes that we need to map out in detail to understand the MeMAD context of use. The production chain can be roughly divided in the following sequential phases:

1. **Pre-production and conceptualization**, which involves the phase of story building, conceptualization of programs and the planning of the further production processes;
2. **Production**, which involves the production of original audiovisual material;
3. **Post-production**, which involves the assembly and finishing of various pieces of audiovisual material, which is either originally produced in phase (2) or reused from existing sources or archives;
4. **Distribution**, which involves the preparation of the distribution of audiovisual content, including taking care of accessibility and delivering programs in certified formats to distribution outlets.

Concerning the pre-production and conceptualization processes, those are out of scope for MeMAD. We hope to re-use some of the results that are produced at this stage, e.g., production scripts, but as no audiovisual content exist at this point, it is not of particular interest to this project.
Looking further at the production phase, in which new content is being recorded on-site, on-set or in studios, there is also limited benefit to be obtained from the MeMAD project: the acquisition process for new content is performed based on input from pre-production and is executed using highly optimized and specific equipment and procedures. These can be influenced by feedback from the post-production process (e.g., a crew needs to shoot another piece of material because a particular viewpoint was still missing when assembling the program) but it would not inherently be improved by MeMAD tools (as defined in the PUCs).

The interest for MeMAD lies with the post-production and distribution processes, as listed in Table 2.

| Production Phase | Task/Process | Users and stakeholders |
|---|---|---|
| Post-production | **Material collection:** users gather interesting audiovisual content to build programs from. This content can come from original acquisition (obtained from the production phase) or from archives. Documentaries often source much content from archives, while current affairs programs make combinations of both sources (e.g., newly recorded interviews mixed with archive content), and drama production is often exclusively comprised of newly produced material. Users will find material guided by conceptual outlines or stories that were made during pre-production. | **Documentary/current affairs production team, news reporters, journalists**: share similar roles (depending on the specific program format they work at), which is to gather new materials (they might actually be responsible for production part of the material itself) and assemble them according to the program concept. The will also assist editors in making the best editing decisions (cf. Editing process). **Producers**, will oversee the production of programs or items and will guide the creation process towards a desired outcome. **Researchers**, will gather materials from archives based on research done on a specific topic. |
| | **Editing:** this process involves the fine-grained assembly of the program by taking individual pieces from the gathered materials and 'cutting' them into a montage using interleaving pieces of content. Editing can happen for video and audio separately, or combined, depending on the program format. | **News editors, Documentary editors, Archive editors, Drama editors, sound editors**, depending on the program format, each editor has a similar function, but with particular expertise for best delivering a certain format to the screen, will cut and assembly various pieces of audiovisual content into a final presentation. **Directors**, who posses the creative control over the production will assist the editor in making the correct editing decisions. |
| | **Production office**: which involves a variety of task of making plannings, keeping track of progress of the overall production process, and ensure the program is delivered on time and budget. This means that the production office oversees all other processes in post-production and delivery, and is hence a stakeholder in those processes too. | **Producers**, who in this role supervise the budget of the production and ensure that the program is delivered for the smallest budget possible, while enabling creative visions to be expressed wherever possible. |
| Distribution | **Subtitling:** which involves the creation of textual subtitles or | **Subtitlers**, who create subtitles for programs that will be broadcast. This can involve |

| | | |
|---|---|---|
| Distribution (continued) | closed captions to help audiences understand the program's content. | creating same-language subtitles for accessibility purposes, or translated subtitles for enabling the material's access in a given language market. Depending on the subtitling context, the procedures and tools used will be different: live subtitling is done in real-time, often using re-speaking ASR technologies, while off-line subtitling is done in batch using dedicated subtitling tools. |
| | **Audio description**, which involves the creation of auditory descriptions of the content depicted in the program, often interleaved with original content audio, e.g., character dialogue or sound effects. | **Audio describers**, who create the audio descriptions, first by writing an script, then by resolving timing such that the descriptions properly interleave the original audio content, and finally by recording and assembling the script into a final audio-described mix. |
| | **Archiving**, which deals with managing material coming into archives and ensuring content is placed in the archive such that it can be retrieved as efficiently as possible. | **Archivists,** who curate the metadata that is input into the archive to ensure all content is annotated in a uniform fashion to ensure maximum retrievability of archived content. |
| | **Delivery**, which deals with the logistics of delivering programs (typically as digital files nowadays) to distribution chains, broadcasters, OTT services, etc. Alternatively, it also deals with delivering content to consumers as efficiently as possible (e.g., by maximizing the content that is being consumed, or by maximizing revenue by promoting content which delivers higher monetary returns or drives better received advertisements). | **Production officers**, who deal with material logistics when finishing and delivering programs, in the right format and with the proper metadata associated (e.g., program identifiers, order numbers, etc.).<br>**Marketing executives**, who deal with optimizing the revenues vs. costs of the delivery services, and who want to promote as much of their service's content as possibly while maximizing revenue, e.g., by enabling content-related advertisements. |

*Table 2: Processes and stakeholders in the media production and consumption chain.*

# 4 Use cases and user stories for an integrated MeMAD prototype system

In order to determine the functional requirements for an integrated MeMAD prototype system, the consortium members have expanded the original four project use cases into an extensive set of candidate user stories which describe the actual potential functionality required from the MeMAD system in more detail. We have done this by intertwining the PUCs with the processes (and stakeholders) identified in the previous section.

Not all candidate user stories listed in this first document revision will be retained for implementation in this project, as they are currently still under revision by the project consortium and the third-party project stakeholders, such as the project's External Collaborators Group. However, the recurring aspects and requirements should become clear from this initial set of user stories, even if some are deemed unfeasible or of lower interest and will hence not be implemented directly.

The following set of tables describe the user stories and the resulting (high-level) functional requirements, along with the relevant users for whom the functionality is provided.

## 4.1 Project Use Case 1: "Content delivery services for the re-use by end-users/clients through media indexing and video description"

Online media delivery platforms rely heavily on media metadata in supplying, recommending and grouping digital media to clients. This use case aims to enhance the end-user experience of such services by creating and making use of rich metadata and hyperlinking created by automated media analysis and multimodal media indexing.

As a result, users of such delivery services should be able to discover and watch media that are meaningful to them from a spectrum of starting points and interests that is significantly broader than what can be achieved by current methods of metadata creation. Users should, for example, be able to browse and discover themes, people and places from media, and parts of media containing these even when the information has not been entered by production staff or the original media product has been designed for a different purpose.

With respect to the media production process, this use case focuses on the consumption process, when actual production has completed. As such, we consider only requirements that deal with content consumer end users.

### 4.1.1 Sub-Use Case 1.1: The user can discover media content about a specific theme, person, place.

Considering entire programs, this sub-use case deals with how end users can discover related content through a variety of dimensions of metadata that is associated with the media content.

| User Story | Description | Users |
|---|---|---|
| 1.1.1 – Searching for consumer content. | Users can search directly for content thanks to metadata associated with all consumable content. The associated metadata exists across several dimensions and topics, incl.: persons, locations, time periods, subjects, etc.<br>This way, users can, for example, locate content dealing with a certain topic such as furniture design, German politics, 1920's lifestyle, cycling, etc. | Consumers |
| 1.1.2 – Finding related content. | Users can discover related content thanks to metadata enrichments added to consumed content. Properly typed relations can further refine the accuracy of these relations.<br>Examples include: a user is interested in other content related to the current by way of a place of living or a time period, or a user is interested in related content because it shares the presence of relatives or prominent figures. | Consumers |

*Table 3: User stories for sub-use case 1.1.*

### 4.1.2 Sub Use Case 1.2: Getting the relevant parts from the program.

Not only entire programmes can be cross-related and searched for, but also parts of a program. By relating programme parts, users can be given even more flexibility in consuming relevant content.

| User Story | Description | Users |
|---|---|---|
| 1.2.1 – Finding related program segments | Users can access individual program segments, instead of accessing entire programs through which they then have to filter the relevant sections manually.<br>Examples of this story are the following:<br>• From a lifestyle or current affairs programme, users can find only those segments which are of interest to them, e.g., dealing with the correct topic, discussing a person of interest, etc.<br>• Users would like to find all quotes on a certain topic pronounced by a public figure and be able to listen and/or to see them. | Consumers |
| 1.2.2 – Skipping program segments | Users can skip those segments from a programme that are not of interest to them. This could include also skipping the end credits and opening graphics automatically between episodes. | Consumers |

*Table 4: User stories for sub-use case 1.2.*

## 4.2 Project Use Case 2: "Creation, use, re-use and re-purposing of new footage and archived content in digital media production through media indexing and video description"

This use case aims to improve discoverability and re-usability of digital-born as well as pre-existing media for the purpose of crafting new stories and audiovisual concepts. Media professionals are provided with rich and relevant relationships between archive media, scripts and raw footage during different stages of digital media production, enabling them to develop a digital story and concepts with the help of automated metadata extraction and media analysis. Relevant media fragments are automatically recommended, which saves significant amounts of editorial work compared with conventional methods of research in media archives.

With respect to the media production process, this use case focuses exclusively on the actual creation process, which for our project begins from the moment audiovisual content is created or recuperated and the assembly process can start, right up to

finishing content for delivery. The focus of the requirements hence lies with the professional media producers.

### 4.2.1 Sub Use Case 2.1: Ingest, organization and editing of new footage.

After the very first stages of the media production process where the program is conceptualized and its story elements are defined, the first opportunity for MeMAD to provide meaningful added value is presented: newly created material enters the production facility at an initial stage, and it can then be used for editing and shaping the story into an actual program. This is the subject of requirements for this sub-use case.

| User Story | Description | Users |
|---|---|---|
| 2.1.1 - Real-time analysis and indexing of ingested content. | Reporters return from the field with interviews and other footage. They ingest the material into the production system which indexes the files with rich metadata. The indexed data offers quickly several alternatives for interviews and footage to be used in a very short time span, to ensure the resulting program can be completed the same day. | News and current affairs reporters. |
| 2.1.2 – Extensive analysis of ingested content. | Documentary production teams return with a large collection of raw footage, which they ingest into the production system. The system indexes the files so that the production team can move on with scripting and editing their program. Typically, the amount of media is quite large, but the production schedule is not as tight as on day-to-day news production. | Documentary and current affairs producers. |
| 2.1.3 – Users browse ingested content for editing. | News editors go through news feed material without pre-existing knowledge about the content and chooses an interesting topic to edit a news story on.<br>Instead of starting from a given set of search terms of topics, the ingest library could offer a list of random or popular topics for which content has recently been ingested and | News editors, documentary makers, journalists. |

| | | |
|---|---|---|
| | processed, to kick-start the content discovery process. | |
| 2.1.4 – Ingest feed notifications. | News feeds are constantly monitored and analysed as they feed into the production system. Real-time processing provides speech recognition and keyword spotting, allowing for trend analysis of the detected results. Thanks to such analysis, incoming feed topics can be tracked and potentially newsworthy content can be detected. | News editors, journalists. |
| 2.1.5 - Editing assistance using multi-model metadata. | Editors can take advantage of multimodal annotations of content to help speed up the editing process by quickly triaging material before editing. Examples include:<br>• Occurrences of detected persons in the image are indicated on the editing timeline;<br>• Transcript annotations are available on the editing timeline;<br>• Automatic classification of shot types (close-up, two-shot, over-the-shoulder). | All editors, incl. news editors, documentary editors, current affairs editors, etc. |
| 2.1.6 – Use of autotranslated content for editing. | Editors who are editing interviews conducted in a foreign language unknown to them can get to work immediately, without the need for any available interpreters because the content has been automatically transcribed and machine-translated. | All editors, incl. news editors, documentary editors, current affairs editors, etc. |

*Table 5: User stories for sub-use case 2.1.*

### 4.2.2 Sub Use Case 2.2: Discoverability of archive content.

Not all content used for creating audiovisual programs is newly created for that single program. Often, material is reused from archives, where the challenge is to disclose as much relevant content from these archives. This is not a trivial task, as contemporary processes can only rely on manually entered metadata for searching. MeMAD can help

resolve this issue by retro-actively processing and enriching archived content such that it becomes easily discoverable for re-use in new productions.

| User Story | Description | Users |
|---|---|---|
| 2.2.1 – Searching for content in archives. | When searching through the archive, researchers can find material using metadata that has been added automatically, and optionally been corrected by archivists. As with user story 1.1.1, this associated metadata exists across several dimensions and topics. Researchers can browse using detected topics, persons, speech fragments, etc. using named entities or free text queries. Examples include:<br>• Looking for content about a given celebrity who has recently deceased;<br>• Looking for footage of an Airbus A380 taking off from Charles de Gaulle on a foggy morning;<br>• Looking for specific quotes uttered by a politician who was in the news yesterday. | Researchers, journalists, editors. |
| 2.2.2 - Searching for segments of content in archives. | As an extension to 2.2.1, researchers can find also locate just those sections that are relevant to the search query of the user.<br>E.g., news editors wants to find the correct one sentence quote from the video recordings of parliamentary meetings. | Researchers, journalists, editors. |
| 2.2.3. Notifications from the archive about selected topics. | Researchers can set up notifications such that they are alerted to new content in the archive that matches their search criteria.<br>E.g., a news editor instructs the system to watch content from the city council meeting to watch if something interesting came up. | Researchers, journalists, editors. |

| | | |
|---|---|---|
| 2.2.4 – Intuitive manual correction of automatically generated metadata. | Archivists can correct automatically tagged and enriched archival items. This must be done using an intuitive user interface such that this process will take much less time than inputting all metadata manually. | Archivists. |

*Table 6: User stories for sub-use case 2.2.*

### 4.2.3 Sub Use Case 2.3: Managing material and footage between multiple production parties

After an audiovisual program is completed there remain a variety of opportunities for the MeMAD project to help facilitate in helping with exchanges of material, and their associated metadata, between different production parties, e.g., between the production house and the broadcaster, or between the production and an archive.

| User Story | Description | Users |
|---|---|---|
| 2.3.1 – Tracking media assets in final programs. | Archive researchers look up promising video clips for a TV production and deliver them to a production house responsible for the production. Later on, the production house wants to track which segments from which archive clips were used in the finished program. | Researchers, producers, rights managers. |
| 2.3.2 – Delivering relevant production metadata downstream. | After finishing a joint production, a production company delivers the finished TV series to a media archive and the production officers sending the files needs to add content description and metadata to them based on the guidelines from the receiving archive. | Production officers, archivists. |
| 2.3.3 – Processing and harmonizing delivered production metadata. | Archivists at a media archive receive finished TV programs from multiple production companies. Some programs may have partial metadata or content descriptions, but archivists need to produce coherent metadata for all to enable consistent further archive use. | Archivists. |

*Table 7: User stories for sub-use case 2.3.*

## 4.3 Project Use Case 3: "Improving user experience with media enrichment by linking to external resources."

A video program may be edited using a complex narrative, but viewers have different background and interests and may not be familiar with all the elements being presented, triggering the need to go more in depth for some aspects being presented. Video programs also trigger social media reactions (e.g. on Twitter or Facebook) where sometimes viewers clip and repurpose some original parts of the video program. One way to improve the user experience is to provide individual users the possibility to access and explore related material (e.g. videos, news articles or set of facts extracted from encyclopedia) that will contain additional information that they personally need or are interested in to better understand the narrative of the video program.

External material may be essential for understanding the audiovisual content. For example, when republishing decades old audiovisual content from the archives, to understand the meaning of the archive content, additional material may be required that gives the historical context and information on how to interpret the content.

With respect to the media production process, as with use case #1, this use case focuses on the consumption process, when actual production has completed. As such, we consider only requirements that deal with content consumer end users and other stakeholders in this process. To the extent that the user stories defined here require metadata to be made available during the media creation process, they will have a counterpart user story from use case #2 or #4.

### 4.3.1 Sub Use Case 3.1: Promoting relevant cross-platform media content.

In this sub-use case, related content, both linearly audiovisual but also interactive and cross-platform experiences are recommended to users as part of navigating content libraries.

| User Story | Description | Users |
|---|---|---|
| 3.1.1 – Libraries of Audiovisual content hyperlink to various related media during browsing. | Users browsing on-demand OTT services can select interesting topics or headlines, which refers them to audio and video clips related to the first broadcast and textual content describing how the different media clips are related to the topic, in addition to containing references to news articles that were produced about this topic. | Consumers. |

*Table 8: User stories for sub-use case 3.1.*

### 4.3.2 Sub Use Case 3.2: Extending the user-experience with more details and background information about the content.

Providing users watching audiovisual content with a rich-media experience of linked content that relates to the consumed content can provide valuable insights for those users and could also lead additional disclosure of existing but seldomly accessed rich media content.

| User Story | Description | Users |
|---|---|---|
| 3.2.1 – During playback of Audiovisual content hyperlink to various related media are shown. | Users watching documentary or current affairs content through an on-demand service are provided with background information – and even relevant links - on topics that are addressed in the program. For example:<br>• For users watching a program about animals in Sahara, an on-demand service displays information about the currently visible objects, such as ants, birds and plants;<br>• Users listening to radio programs about birds are presented with information about the birds being discussed on a second screen.<br>• Users watching current affairs programs in which politicians are features are presented with linked content to clarify each politician's background and affiliation, along with party programme points that this politicians party stands for. | Consumers |
| 3.2.2 – During playback of Audiovisual content hyperlink to various interactive media are presented. | We provide three examples to sketch possible scenarios for this user story:<br>• Users watching sports content through an on-demand service are provided with statistics about the players and game, along with relevant historical statistics and links to further information.<br>• Users watching lifestyle programs through an on-demand service can participate in discussions with like-minded consumers related to the topics addressed in the program. | Consumers |

| | | |
|---|---|---|
| | • Users watching a health program about diabetes are also shown an interactive experience, “are you in risk of getting diabetes?” to test their own risk of obtaining the disease. | |

*Table 9: User stories for sub-use case 3.2.*

### 4.3.3 Sub Use Case 3.3: Validating the content, e.g. the truthfulness.

Providing users with links to related rich media content without curation can present hazards, as the linked content might not always present truthful and accurate information. At the same time, the original content might suffer from the same issues. We can envision a potential role for the MeMAD project in enabling insights into the truthfulness of the content that is consumed and linked to.

| User Story | Description | Users |
|---|---|---|
| 3.3.1 – Truthfulness validation of audiovisual content. | Users watching news, current affairs or political programs are presented with results from a truthfulness analysis based on the content’s analysed speech and externally linked resources, giving an indication whether what is being said on screen is plausible to represent the truth, or is likely fake news. | Consumers. |

*Table 10: User stories for sub-use case 3.3.*

### 4.3.4 Sub Use Case 3.4: Show relevant TV or other advertisement in context of the current content.

Content providers can benefit from targeted advertising, which is related to the content being distributed, because it is more relevant to consumers than generic advertising. Leaving in the middle whether the user’s personal preferences are taken into account, or the advertising is based only on the profile of the content, MeMAD-generated and managed metadata can assist in advertisement recommendations.

| User Story | Description | Users |
|---|---|---|
| 3.4.1 - Content-related advertisements | Free OTT distribution services send out targeted content-related advertisements. Instead of showing generic commercials, the OTT service can benefit from associated | Consumers, OTT distribution producers. |

| | | |
|---|---|---|
| | media item metadata to show advertisements that are likely more relevant and of interest to viewers.<br>At the same time, the OTT service can sell this advertisement space in a targeted way, e.g., bicycle manufacturers can bid on advertisement slots associated with sporting events or cycling documentaries. | |

*Table 11: User stories for sub-use case 3.4.*

## 4.4 Project Use Case 4: "Automated subtitling/captioning and audio description. Speech and sounds to text and also visual content to text, both with multiple output languages, for general purpose use and for the deaf, hard-of-hearing, blind, and partially-sighted audiences."

This use case addresses an urgent requirement to enhance as much content as possible with complementary subtitles and aural audio description. Conventionally these are created by human subtitlers and translators, and at a total production cost of 1000-1200 Euro per hour (for subtitling) up to 3000 Euro per hour (for audio description). Also, manual subtitling and audio description requires a significant cycle time from one to two weeks. For this use case, we will undertake to maximize productivity of both subtitling (same language as well as language to language) and audio description processes, through "supervised automation".

This is the single PUC which is clearly represented both in the production and consumption process. The 'consumption' of subtitling and audio description, especially if targeted toward minority groups of audiences for accessibility purposes, needs to have a consumer counterpart such that the project can properly take into account the consumption environment and the consumer quality requirements that will be posed on any generated subtitles or audio descriptions. Meanwhile, of course, the actual production processes involved in making these elements are an important focus for MeMAD.

### 4.4.1 Sub Use Case 4.1: Live / semi-live captioning, subtitling and audio description.

MeMAD has the potential to assist in optimizing contemporary accessibility production processes such as subtitling and audio description. In this first sub use case, we consider the processes that already exist today and that could be helped by the MeMAD components, without profoundly impacting common production practices.

| User Story | Description | Users |
|---|---|---|
| 4.1.1 – Assistance in live subtitling. | Subtitlers who are live subtitling (i.e., with a minimal delay wrt. the broadcasted program, measured in seconds) could be aided live ASR results that provide suggested subtitles which only need correction. | Subtitlers. |
| 4.1.2 -Assistance in live audio description | Similarly, audio describers who are describing live broadcasts to aid visually impaired people could be helped with suggested automated descriptions of the content (e.g., automatic identification of people in the image). | Audio description producers. |
| 4.1.3 – Assistance in near-live subtitling. | In near-live situations, the time pressure to deliver subtitles is much less than in live scenarios. This could provide a different dynamic and allow MeMAD tools to help in this process. The suitability compared to live subtitling should be investigated in this case. | Subtitlers. |
| 4.1.4 – Automated same-language subtitling. | Users, and in particular, hearing-impaired users are provided with automatically generated same-language subtitles for content such that they can consume the content without the audio being available or audible. | Subtitlers. |

*Table 12: User stories for sub-use case 4.1.*

### 4.4.2 Sub Use Case 4.2: Extending coverage of audio descriptions

Finding ways to extend the coverage of audio descriptions, without proportionally increasing the effort to create these descriptions for more content will be an important aspect of the MeMAD project, but at the same time a very challenging one: creating audio descriptions which correctly capture the semantics of the audiovisual content and transcend the level of plainly describing what is visible and audible to take into account the editorial context of the content will be hard to do.

| User Story | Description | Users |
|---|---|---|
| 4.2.1 – Content consumption with auto-generated audio descriptions. | Visually impaired consumers can still experience all episodes of their favorite shows, thanks to the audio descriptions that have been made available using an additional audio track. | (Visually impaired) consumers |
| 4.2.2 – Manual corrections improve auto-generated audio descriptions. | Audio description producers in charge of delivering audio descriptions for documentaries can deliver audio descriptions more efficiently thanks to automatically generated audio descriptions, that are reviewed and corrected manually. | Audio description producers. |

*Table 13: User stories for sub-use case 4.2.*

### 4.4.3 Sub Use Case 4.3: Automatic translation of existing subtitles to other languages to increase minority or general audience accessibility.

The availability of subtitles associated with audiovisual content is often the most straightforward way of lowering barriers towards new audiences: textual subtitles can be delivered via side-channels and provide meaning to any foreign-language content. Making additional subtitles in other languages available to new audiences at marginal cost is an important topic for the MeMAD project.

| User Story | Description | Users |
|---|---|---|
| 4.3.1 – Automatically translated subtitles for foreign users. | Users that have moved in from abroad can select the automatically generated subtitling in a language familiar to them such that they can follow what goes on in the program. | Consumers. |
| 4.3.2 – Automatically translated subtitling of foreign content. | Users browsing foreign European media libraries can consume this content even if this content produced in other languages. Thanks to automatically translated subtitles or audio descriptions, users can experience and | Consumers. |

| | | |
|---|---|---|
| | understand content otherwise inaccessible to them. | |
| 4.3.3 - Translated subtitles based on translated transcripts | Subtitlers can create translated subtitles using translated transcripts as a starting point, avoiding the need for translators in many aspects of the subtitling process. | Subtitlers. |
| 4.3.4 - Manual correction of auto-translated subtitles | Subtitlers need to manually correct automatically translated subtitles because the automated translation will generate errors, and subtitle timing or wording sometimes need to be changed to deliver subtitles of sufficient quality. | Subtitlers. |

*Table 14: User stories for sub-use case 4.3.*

This concludes the initial set of user stories defined for each of the PUCs of MeMAD. This set of high-level functionalities will be revised and refined now as follows:

- Media professionals from the consortium will review and provide additional insights, with the aim of preparing a survey that can be distributed to participants external to the project.
- This survey, conducted with the industry external collaborators and interested parties, will further define which user stories to prioritise or retain for actual implementation of the project.

The refined list of the user stories will then be turned into detailed user requirements and back-end specifications, as explained in Section 2.

# 5 Metadata Exchange Format requirements

While the breakdown of functionalities to be implemented listed in the previous section is not yet finalized, the user stories and their descriptions defined in the previous section can already give us a better understanding of the various metadata that will be required to form the backbone of the MeMAD ecosystem. Generating, editing and managing this metadata such that it brings significant added value through new or improved applications is the core of the MeMAD project. As such, exchanges of well-structured metadata will be vital to the implementation of the MeMAD prototype platform, as they will allow individual components to contribute to a growing graph of metadata which in turn will support enhancements to the production and consumption processes described before.

We can already learn many of the features those metadata – and the formats they are exchanged in – need to support. This will give the consortium the opportunity to start preparing the development of back-end tools (cf. the work done in all WP2 tasks: T2.1, T2.2 and T2.3) even while end-user requirements are still being refined.

For each of the user stories in the previous section, we now provide a summary of metadata that is relevant to be processed and interacted with. We also list a number of considerations that were devised by the consortium partners as the user story candidates were being drawn out. We don't provide any conclusions about these considerations just yet; but they will be taken into account when further refining the project requirements in future version of this document.
From the exhaustive list of required metadata, we then summarize each of the types of metadata that the project will be required to support.

## 5.1 Required metadata for consumer-oriented user stories

| **User Story** | **Relevant item metadata** |
|---|---|
| 1.1.1 – Searching for consumer content.<br>1.1.2 – Finding related content. | Content tagged with named entities, including:<br>• objects/nouns and time periods (input);<br>• persons.<br>• places.<br>• time periods.<br>• topics, objects/nouns and time periods.<br>• topics, persons, political orientations and affiliations. |
| 3.1.1 – Libraries of Audiovisual content hyperlink to various related media during browsing.<br>3.2.2 – During playback of audiovisual content hyperlink | The above, plus the addition of metadata with spatial coordinates per described entity. |

| | |
|---|---|
| to various interactive media are presented. | |
| 3.2.1 – During playback of audiovisual content hyperlink to various related media are shown. | Content enriched with links to rich metadata of which the structure can be a variety of formats, e.g., for use in custom application's GUI;<br><br>Content tagged with links to interactive applications instead of static content sources. |
| 1.2.1 – Finding related program segments.<br>1.2.2 – Skipping program segments. | Content tagged with time-based metadata of named entities. This metadata forms a timeline in which each relevant description is temporally constrained to just the section for which it is relevant;<br><br>Similarly, content can be tagged with time-based annotation of typed sections (e.g., credits, intro, etc.);<br><br>Spoken (time-coded) transcript text fragments of spoken audio. |
| 3.3.1 – Truthfulness validation of audiovisual content. | Metadata that links to sources which prove or disprove the assumptions stated in the content item. |
| 3.4.1 - Content-related advertisements | Content tagged with named entities for objects/nouns and actual products with links to advertisements, or resources where items can be purchased. |

*Table 15: Required metadata for consumer-oriented user stories from PUC 1 and 3.*

**Considerations for the generation and use of this metadata.**

Concerning search functionality offered to consumers:

- Depending on the flexibility required from the Search functionality, explicitly typed named entity search parameters could be provided (e.g., searching for "Paris" which is a "Location").
- Search functionality might also require combinations of parameters, i.e., to enable AND and OR operators on search queries, and also ranking of search results and inclusion of related items in search results.

Concerning the detection of objects in content offered to consumers:

- Depending on the subjects that require automatic detection, a variety of trained models will need to be employed to detect domain-specific subjects, e.g., exotic animals, but even more obscure elements might be present in the audiovisual content.

- Adding spatial coordinates (in addition to temporal ones) will allow users to distinguish objects of interest visible at the same moment in time.

Concerning program segmentation:

- By ensuring that all technologies used to produce metadata are capable of delivering time-coded metadata, we can solve this challenge. All tools should have a grasp of the temporal aspect of audio and video content and this should be well-represented by the metadata they produce.
- Spoken text fragments, or transcript fragments, can be a valuable source of metadata to gain insights into spoken audio, especially if they can also be located at the precise point in time that they occur in the content segment.

Concerning the advertising metadata:

- Depending on the type of content being associated with advertisements, it could be necessary to also add exclusion lists to the tagged named entities. This way sensitive content could be avoided, e.g., don't show water commercials when the topic is about devastating tsunamis.

## 5.2 Required metadata for production-oriented user stories

| **User Story** | **Relevant item metadata** |
|---|---|
| 2.1.1 - Real-time analysis and indexing of ingested content.<br>2.1.2 – Extensive analysis of ingested content.<br>2.1.3 – Users browse ingested content for editing.<br>2.1.5 - Editing assistance using multi-model metadata.<br>2.1.4 – Ingest feed notifications. | Transcripts of spoken audio;<br><br>Video content descriptions;<br><br>Disambiguated named entities concerning a variety of subjects: persons, locations, etc. |
| 2.1.6 – Use of auto-translated content for editing. | Translated transcripts of spoken audio. |
| 2.2.1 – Searching for content in archives.<br>2.2.2 - Searching for segments of content in archives.<br>2.2.3. Notifications from the archive about selected topics. | Content tagged with named entities, including:<br>• objects/nouns and time periods (input);<br>• persons.<br>• places.<br>• time periods.<br>• topics, objects/nouns and products.<br>• environmental characteristics (outdoor, indoor, scenery, etc.)<br>• topics, persons, political orientations and affiliations.<br><br>Transcripts of spoken audio. |

| | |
|---|---|
| 2.2.4 – Intuitive manual correction of automatically generated metadata. | As with 2.2.3, but with video content descriptions. |
| 2.3.1 – Tracking media assets in final programs. | Time-based metadata in the form of a rich EDL traces a composed content item back into its origins (possibly spanning multiple generations). |
| 2.3.2 – Delivering relevant production metadata downstream.<br>2.3.3 – Processing and harmonizing delivered production metadata. | A variety of metadata can be the result of the production process, including:<br>• Subtitles;<br>• Speech transcripts;<br>• Production scripts;<br>• Logging data which corresponds to named entities (e.g., person names given to video clips to identify interviewees, or locations where video was shot).<br>• Production office metadata concerning characters and actors, locations used, research gathered during the production process, etc. |

*Table 16: Required metadata for production-oriented user stories from PUC 2.*

**Considerations for the generation and use of this metadata.**

Concerning the automated indexing and analysis of ingested content:

- For analysis given a very short lead time, the generation of the metadata will likely occur automatically, using 1) ASR, 2) video analytics and description algorithms, 3) named entity recognition and disambiguation tools. The metadata must be indexed and then related to other material available such that relevant related content can quickly be located.
- In related cases, but given the longer lead time, it is feasible that longer-running but more thorough processing techniques are used to deliver more accurate results that require less manual curation before being useable or signed off as completed.
- Finally, for cases where real-time monitoring is desired, the timing constraints are even more specific, the analytics need to be performed in real-time, but also cannot run faster than real-time. Depending on the types of analytics available, some tools or advanced processing modes might have to be skipped in this case.

Concerning using indexed ingested or archival content for editing:

- The metadata used is similar as for the user stories that focus on searching, except that it would need to be present conveniently during editing, for example, for filtering batches of materials, of for grouping materials based on shot type, subjects, etc.

Concerning searching and indexing of archival content:

- A variety of approaches can be considered for the generation of this metadata. One way is that, for content that has been tagged manually by archivists, the

manual tagging metadata has been augmented by named entity recognition and disambiguation. In more advanced scenarios, instead of starting from manually annotated content, augmentations are done completely based on automated detection processes.
- Whenever possible, analytics tools should try to augment manually tagged content not only from item-wide annotations, but also for temporal descriptions, e.g., actions that take place at a given moment in the content.

Concerning manual corrections of automatically generated metadata:

- The main goal is to automate as much as possible in the content analytics and enrichment process, but manual corrections will remain necessary in many cases, for example, for speech transcripts where exact quotes are important.
- When corrections need to happen, the aim is to observe that manual interventions take significantly less time than performing the annotation process entirely manually.
- Similarly, when ingesting incoming media into an archive, after being augmented automatically, proper curation is required to ensure correct data is archived, ideally using convenient editing and verification tools.
- In any case, delivered metadata (whether it is to an archive or another production party) needs to be presented in a clearly defined format, ideally based on a standard or industry best practices and procedures.
  This, along with the availability of proper identification mechanism will be crucial in case multiple renditions of the same content are enriched with metadata independently, and then need to be consolidated in a single data repository. Additionally, the validation of incoming metadata should be considered, potentially by comparing this incoming metadata with results from automated metadata extraction processes.

Concerning the generation of EDLs for tracing back content's provenance:

- The creation of this EDL can be done through an explicit audit trail during the production process (where actual EDL are processed, exchanged and stored) or it can be performed post-factum through video and audio analytics by comparing the composed program content with indices of potential source content.

## 5.3 Required metadata for accessibility-oriented user stories

| **User Story** | **Relevant item metadata** |
|---|---|
| 4.1.1 – Assistance in live subtitling.<br>4.1.3 – Assistance in near-live subtitling. | Transcripts of spoken audio;<br><br>Subtitles. |
| 4.1.2 – Assistance in live audio description | Video content descriptions;<br><br>Audio descriptions and/or timed transcripts. |

| | |
|---|---|
| 4.2.1 – Content consumption with auto-generated audio descriptions.<br>4.2.2 – Manual corrections improve auto-generated audio descriptions. | Time-based audio descriptions;<br><br>Timed transcripts. |
| 4.3.1 – Automatically translated subtitling for foreign users.<br>4.3.2 – Automatically translated subtitles of foreign content | Subtitles in multiple languages. |
| 4.3.3 - Translated subtitles based on translated transcripts<br>4.3.4 - Manual correction of auto-translated subtitles | Subtitles in multiple languages;<br><br>Optionally, automatically generated transcripts and content descriptions. |

*Table 17: Required metadata for accessibility-oriented user stories.*

**Considerations for the generation and use of this metadata.**

Concerning subtitling or audio description assistance:

- Even if subtitling remains a manual process, the resulting subtitles are very valuable for further processing actions. For example, subtitles that contain correct individuals' names will greatly help in disambiguation of named entities and hence allow for better content relations.
- Similarly, manually generated audio content descriptions can help other detection processes as manual interpretation will aid any automated disambiguation tasks.
- The perceived quality of automatically generated audio descriptions will depend on the program format and the relation between what is visible or audible in the content and its relevance to the narrative. Depending on the format, it could be relevant to involve the use of production scripts to guide automated description processes in extracting relevant descriptions that might not be obvious when only observing visual content.

Concerning auto-generated audio descriptions:

- Manual corrections are likely to form an important part of the process of delivering auto-generated audio descriptions for a large variety of content formats which previously often went without audio descriptions.
  End users will have to decide whether automatically generated audio descriptions are of sufficient quality for end user consumption, or whether human intervention is preferred to deliver adequate aids for visually impaired viewers.

Concerning auto-translated subtitles:

- Two scenarios are relevant:
  - Existing subtitles are translated into the target language of the consumer.

    - Subtitles in the target language are derived from automatically generated and translated transcripts of spoken text.
- Implementations might combine original subtitles with computer-generated transcripts to aid in the translation process. Additionally, the way in which subtitles are generated could differ; generation could strictly follow original subtitle timing and fill in translations as best as possible, or the new subtitles could follow transcripts in first order, and then be optimized for optimal temporal placement (e.g., not crossing shot boundaries).
- In any case, the option must be explored to use multimodal techniques to help the improve the accuracy of the translation of subtitles or transcripts to the target language.

## 5.4 Consolidated required metadata

If we now invert the presentation of these metadata requirements, and group similar functionality, we obtain the various metadata elements that the MeMAD project components will use, what their requirements are, and provide an early suggestion regarding formats we can adopt to exchange and process this metadata.

| # | Type of metadata | Metadata requirements | Suggested format |
|---|---|---|---|
| 1 | Subtitles | Subtitles must be associated to a language and maybe also to a revision identifier. Additionally, all current capabilities used in the industry must be supported, incl. subtitle positioning, subtitle markup, subtitle coloring, etc. | Formats such as W3C TTML[3] or EBU-TT[4] are standardized formats and contain all required attributes listed. |
| 2 | Timed transcripts | Timed transcripts must also define a language, up to a granularity of individual words if applicable. Additionally, the following information should be included:<br>• Named or placeholder speaker segmentation;<br>• Per-word timing information;<br>• Per-word confidence scores;<br>• Optionally: per-word or per-transcript segment | A starting point for representing timed transcripts could be the subtitle format, however, this would likely need to be extended with many of the attributes listed on the left. We will evaluate which approach works best, extending a subtitle |

[3] Cf. Timed Text Markup Language 1 (TTML1) (Second Edition), W3C Recommendation, 24 September 2013, at https://www.w3.org/TR/ttml1 .
[4] Cf. EBU-TT Part 1 Subtitling format definition Version 1.0, EBU Tech. 3350, 2012, at https://tech.ebu.ch/docs/tech/tech3350v1-0.pdf .

| | | | |
|---|---|---|---|
| | | alternatives (each with confidence scores). | format, or defining a new and targeted format for time transcripts. |
| 3 | Production scripts | Production scripts are typically formatted according to a common screenplay format for drama production, or more loosely define formats for other types of programmes, e.g., spreadsheets or text documents. Actual formats used should be parsable into a machine-processable and accessible format such that individual typed parts of the script can be extracted and used if applicable in a given context (e.g., only use dialogue for speech-related processing tasks). | Movie Script Markup Language[5] defines the common elements of a drama screenplay and could be used as basis for exchanging production script data. On the other hand, EBUCore[6] version 1.8 has recently been extended with elements that could likely also represent sections of a script. |
| 4 | Timed natural language video content description | In its most basic form, this encompasses natural language content descriptions, i.e., sentences that describe the content in the image (either directly visible as would be detected by video analytics tools, or indirectly as inferred from an available script or synchronized audio). In any case, timing and language information is also vital. Some processes require only this form of metadata to work, e.g., as output from analytics tools or as input to an audio description rendering process. | Again, a subtitle format could serve as basis for basic natural language video content descriptions as they would need to store only text without much additional markup. On the other hand, spatial features could be represented by defining subtitle regions. |

[5] Cf. Dieter Van Rijsselbergen, Barbara Van De Keer, Maarten Verwaest, Erik Mannens, and Rik Van de Walle. 2009. Movie script markup language. In Proceedings of the 9th ACM symposium on Document engineering (DocEng '09). ACM, New York, NY, USA, 161-170.

[6] Cf. EBU CORE METADATA SET (EBUCore) Version 1.8, EBU Tech. 3293, 2017, at https://tech.ebu.ch/docs/tech/tech3293.pdf .

| | | | |
|---|---|---|---|
| 5 | Semantically linked content descriptions with identified and disambiguated named entities. | In many user stories above, a simple set of relevant disambiguated named entities (disambiguated in such a way that each is linked on a semantic level with external resources that provide the right context for relating the item in question with other similar items or resources available on the internet or other content repositories).<br>As seen from the user stories, the following are entities types required to be described:<br>• persons;<br>• objects/nouns;<br>• time periods;<br>• places;<br>• affiliations and political orientations;<br>• actions;<br>• environmental characteristics.<br><br>Whenever possible, when storing this metadata, references should be made to widely used and state-of-the-art controlled ontologies, vocabularies and repositories of disambiguated named entity definitions. In this way, the correct context is provided for further processing of content items (e.g., for recommendations and assistance in translations), and this reduces the chances for errors in metadata processing based on false assumptions provided by named entity definitions. | RDF[7] is the most commonly used format for representing this type of metadata (either in XML form or a lightweight alternative such as N3) and will likely be a good match in MeMAD also.<br><br>Further research is required to determine which exact ontologies will be used to represent and define various aspects of the metadata. |
| 6 | Timed and semantically linked natural language content descriptions with | In order to fully enable the capabilities required in the most advanced use cases in this project, a combination of metadata #4 and #5 will be required. A combination | EBUCore is a viable candidate to represent this type of metadata. EBUCore can be represented using RDF, |

[7] Cf. RDF 1.1 Concepts and Abstract Syntax, W3C Recommendation, 25 February 2014, at https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/ .

| | | | |
|---|---|---|---|
| | identified and disambiguated named entities. | of both natural-language content descriptions (i.e., #4), in which each relevant named entity has been disambiguated and linked to its correct definition (i.e., #5), will provide the best possible context for further automated processing, e.g., to produce better translations, or to deliver better search results in a cross-lingual search. | including a mix of non-EBUCore elements if required, but taking advantage of EBUCore's media-oriented elements to define metadata using spatial and temporal constraints.<br>If EBUCore is used, we will likely define an application specification to provide a framework of how EBUCore can be utilized to represent metadata for a number of our use cases. |
| 7 | Linked content. | We assume the primary mechanism of describing linked and related content to be done through a URL/URI mechanism, but it is hard to specify specific details early in the project. Formats to describe such linked content will likely build upon those that are used to describe metadata #4, and links will point to specific resource destinations instead of 'logical' definitions of named entities. | As with #5, RDF is a good candidate to be used for this type of metadata. |
| 8 | Edit Decision Lists (EDLs) | Those use cases that require tracing the provenance of composed assets, can refer to formats and models used in the industry for exactly this purpose: describing how editing of an item should be done. This is typically defined in terms of audio and video tracks, each of which can contain a temporal sequence of sub clips from a set of source materials. | Full-featured edit decision lists are typically conveyed using de facto industry file formats, such as flavors of XML (used by Apple's Final Cut Pro) or Advanced Authoring Format[8] (AAF, used by Avid Media Composer and a variety of other editing tools). |

[8] Cf. Advanced Authoring Format Object Specification, Version 1.1, AAF Association/AMWA, 2005, at http://aaf.sourceforge.net/docs/aafObjectModel.pdf .

| | | | |
|---|---|---|---|
| | | | Unfortunately, these formats feature only limited extensibility and in some cases are stored in binary form, complicating their use. However, they can serve as an inspiration for modeling other formats, or guide the use of formats that do specify composition details, such as SMIL[9] or SMPTE IMF[10]. Again, a middle ground could potentially be found in EBUCore, in which references to parts provides a way to express content provenance. |

*Table 18: Consolidation of required metadata types for MeMAD prototypes.*

We can already make an important observation about the required metadata, namely that several types of metadata are relevant for many user stories, even across PUCs and user contexts (e.g., subtitles and named entity tags are vital supporting metadata during both the media production and consumption processes). Hence, we can expect the re-use of metadata to become an important aspect of the MeMAD prototype, such that metadata that is generated once can serve a new purpose in another usage context.
While this re-use is actually the subject of user story 2.3.2, it will also be important to bear in mind for other user stories and for the choice of metadata formats: they should support the maximum of requirements to support all valid use cases such that re-use can happen seamlessly.

During the remainder of the first year of the MeMAD project, the aim is to settle on one particular format for each type of metadata, which will then be exchanged between all relevant technological services in the integrated system.

[9] Cf. Synchronized Multimedia Integration Language (SMIL 3.0), W3C Recommendation, 01 December 2008, at https://www.w3.org/TR/SMIL/ .
[10] Cf. Interoperable Master Format — Overview for the SMPTE 2067 Document Suite, OV2067-0:2017, SMPTE, 2017, at http://ieeexplore.ieee.org/iel7/7864295/7864296/07864297.pdf .

Note finally, that we have not explicitly discussed or defined recommended formats for actual audiovisual content in this document. The aim of MeMAD is not to further the state of the art in audiovisual encoding or storage, and hence, MeMAD will use commonly used formats such as the ISO (MP4), Quicktime (MOV) or Material Exchange Format (MXF) as container formats, and will determine a limited set of commonly supported audio and video codecs to be used for distributing content to and from the project's processing services.

# 6 Mapping MeMAD metadata to MeMAD technology components

After surveying the technology components provided by project partners based on the project plan, the following relations between these components, and the exchanged metadata identified in the use cases can be observed, as depicted in Figure 2. The figure shows elliptical nodes that represent processing tools (provided by the indicated project task), interconnected by the inputs and outputs accepted and delivered by each component (the rectangular metadata nodes, each of which refers to a metadata type defined in the previous section).

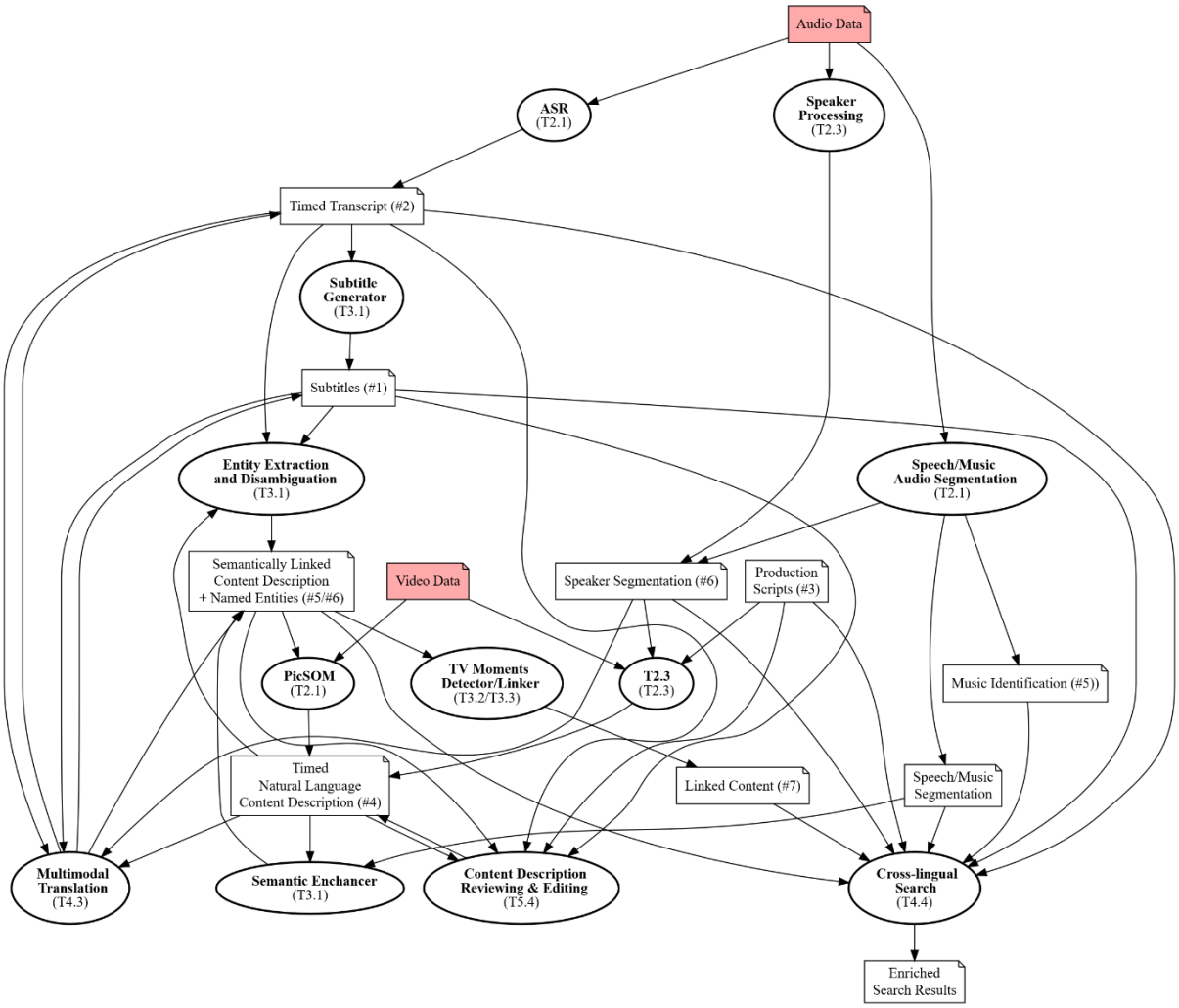


*Figure 2: MeMAD components and exchanged metadata.*

As can be observed from the image above, all relevant components have been associated in the graph, and the exchanges of metadata can be properly distinguished. This exercise forms the basis for defining actual exchanges and for combining individual processing services into usable high-level functional workflows, which is the subject of the next section. Additionally, it highlights which components (available or to develop) will input or output which kinds of metadata. This information will be briefed to component developers in the various applicable work packages. We will further refine these relations and this image as the user requirements for the project are finalized.

# 7 Defining the first prototype system implementation

The main goal of the first prototype iteration, as defined in the project's DoA, is to combine existing technology components (provided by WP2-5), in the state that they are at the beginning of the project, as a single coherent system. This system will offer end-users an initial set of workflows based on their creative authoring processes, providing a glue between each of the underlying technologies. Experimental new components will not yet be considered in this release, as the focus will be placed on integration and building a consensus on the exchange formats and APIs used for the integration.
This approach allows us to commence the implementation of the prototype platform early in the project, despite that fact that the exact requirements are still a moving target, and that only limited progress will have been made on project-specific developments in individual work packages. Still, it will allow the consortium to already prepare for the integration efforts that need to be done and will allow the platform to be made ready to support the previsioned integrations.

Conceptually, the broad picture of the intended integrated platform is shown in Figure 3. The platform forms the backbone of all processing tasks. Workflows are executed by the platform, and the tasks that comprise those workflows are then executed by components delivered by each of the work packages 2-5. To support the processing components, the platform offers storage of the source audio and video content, along with options to transcode to other audiovisual formats to help in easier processing, and it will also store audiovisual content and various forms of metadata.

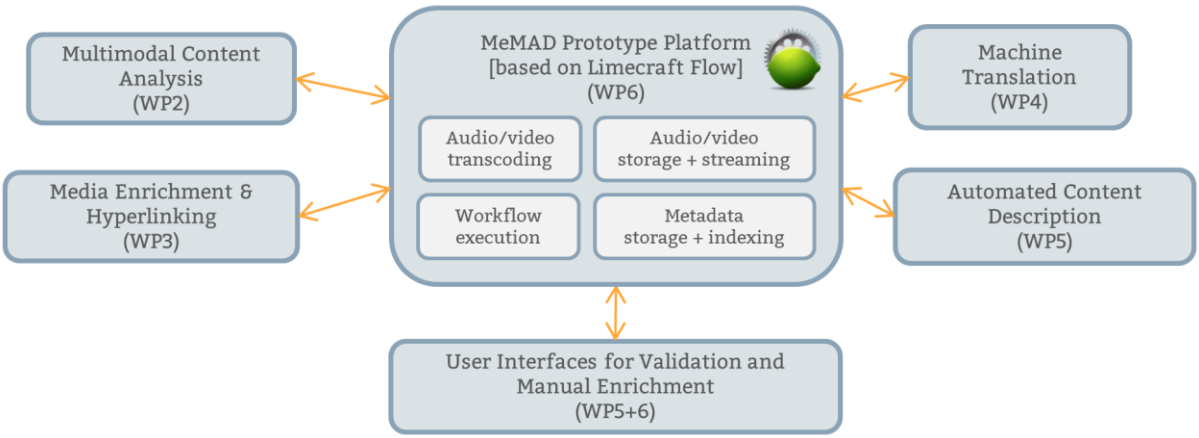


*Figure 3: Conceptual overview of the MeMAD integrated platform.*

We note that the prototype platform will be based on the Limecraft Flow product, which is already commercialized in a Software-as-a-Service model, primarily supported by a pay-as-you-go pricing model exploited by Limecraft. For most of its features, the platform requires no local installation of software applications and can be used from a standard web browser.
To ensure that the project's results do not overly depend on availability of the Limecraft software, we aim to ensure that all underlying components are loosely coupled to the Flow platform. The core of each component will be able to function independently (and

will in many cases also be available as open source software, cf. **D2.1 [TBC]**), and we will build a suitable architecture where components can easily participate in end users workflows that are orchestrated by the platform's API and that information is exchanged using open (and where possible) standardized metadata exchange formats (of which the definition will be defined in subsequent revisions of this document).

With respect to the first version of the prototype integrated MeMAD system, we now specify the functional requirements, in terms of workflows that will be executed. This iteration of the platform will not yet implement extensive support for the use cases listed in Section 2, especially considering that many components will not be ready or even implemented by the time of the first prototype delivery. However, the implemented workflows will provide a foundation of functionality on which we can later build to realize functionalities for supporting complete user stories.

The table below lists 4 initial functional scenarios that we will implement for the first iteration of the platform. We define each scenario in terms of the functionality required from the system, and we list the required components (cf. also Section 6) involved in this scenario.

| **Functional scenario** | **Required components** | **Addressed use cases** |
|---|---|---|
| *MeMAD Automated Speech Recognition*<br><br>An audiovisual item is ingested into the platform and various audio and speech operations are performed, with the aim of allowing users to search for content using this transcribed speech:<br>1. An audiovisual item is ingested into the platform;<br>2. the audio is extracted and;<br>3. then submitted to a speech segmentation component to extract metadata on sections that contain speech vs. non-speech;<br>4. This metadata is then sent back to the platform;<br>5. It is then also delivered, along with the original audio, to a selected ASR components available;<br>6. After speech-to-text processing, the delivered timed transcripts are submitted back to the platform and stored;<br>7. where they can be compared for accuracy by an end user, and be corrected if applicable. | • Platform ingest (T6.2);<br>• Platform transcodes (T6.2);<br>• Platform structured metadata storage (T6.2);<br>• Speech/music audio segmentation (T2.1);<br>• ASR (T2.1);<br>• Faceted Search (T6.2). | UC2.2: Researchers can search through ASR output transcripts derived from archive content to find interesting parts to use in a program.<br>UC2.1: Similarly, program makers can search through ASR output derived from newly recorded media to find those parts that are useful for usage in the program's edit. |

| | | |
|---|---|---|
| 8. Finally, users can perform searches upon the transcript results, retrieving items when search terms match words in the transcribed text. | | |
| *MeMAD Automated Subtitling*<br><br>This scenario builds upon the previous one, but extends it with a specific media authoring process, subtitle generation:<br>1. The timed transcript from a given audiovisual item is sent to a subtitle generation component.<br>2. After processing, the delivered subtitles are submitted to the platform and stored where they can be compared or edited by an end user.<br>3. As the last step in the process, the delivered subtitles can be exported into a standards-compliant file for playback outside of the platform. | All from scenario #1, plus:<br>• Subtitle generation (T3.1). | UC4.1, UC4.3: Automated subtitling can directly assist in realizing the requirements from UC4.1, and they form a stepping stone for implementing UC4.3. |
| *MeMAD NER on video descriptions*<br><br>This scenario brings together content analysis results from a variety of dimensions. In addition to speech-based results (we build upon scenario #1), video analytics and summarization is performed to give insights into the visual content of the media asset:<br>1. An audiovisual item is ingested into the platform and is then; | • Platform ingest (T6.2);<br>• Platform structured metadata storage (T6.2);<br>• PicSOM video analytics and content summarization (T2.1);<br>• Named entity recognition and | Several use cases benefit from this implementation, incl. UC1.1, UC2.1, UC2.2, UC3.1, UC3.2, UC4.2. Each these has an application for the results of named entity recognition, either for relating with other media items or web |

| | | |
|---|---|---|
| 2. submitted to a set of video and audio analytics tools for processing.<br>3. The resulting audiovisual summarization is submitted to the platform and is;<br>4. forwarded to the entity extraction processes to retrieve relevant named entities from the free-text content descriptions.<br>5. The result of this semantic enhancement is reported to the platform, where;<br>6. this metadata is indexed and can be used for faceted searches in the platform's user interfaces. | disambiguation (T3.1);<br>• Semantic enhancement (T3.1);<br>• Faceted search (T6.2). | resources, but also for improving content descriptions. This scenario puts the basic services in place to later build rich applications from. |
| *MeMAD first translation steps*<br><br>This scenario builds on the first scenario but adds to it a transcript translation step:<br>1. The timed transcript from a given audiovisual item is sent to the translation component.<br>2. After processing, the translations are submitted to the platform and stored where they can be compared, edited but most importantly be understood by an end user.<br>3. Finally, users can search in the translated transcript just as they can in the original transcripts. | All components required from scenario #1, plus:<br>• Transcript Translation from Multimodal translation (T4.3). | UC2.1, UC4.3: The most obvious cases for translation are these use cases, where either translated subtitles or transcripts can be used for content retrieval or understanding. Apart from subtitles or transcripts MeMAD will later also have other uses for translations, for example to support multi-language search operations. |

*Table 19: Functional scenarios to be built into the first MeMAD prototype.*

Obviously, each metadata exchange in a workflow listed above should take place using a common data format, of which the requirements and an initial set of suggested formats are listed in the previous section. The actual formats used will be determined in the first implementation phase of the MeMAD integrated platform.

For the first iteration of the platform, only a few non-functional requirements have currently been determined. The following have now been defined:

1. Each component should be integrated in such a way that it can autonomously accept a processing task from the platform, execute this task and report its result back to the platform, without requiring manual intervention;
2. Each component should be able to process multiple tasks without manual intervention, incl., cleaning up when tasks have finished. Ideally, multiple tasks

can be executed simultaneously, but at least tasks should be able to execute sequentially, providing a progress status to the platform.

For the first iteration of the platform, no test scenarios and evaluation criteria are defined yet. The functional requirements determine the scenario to follow, and the evaluation criteria will be defined in the next revision of this document.

# 8 Conclusions

In this deliverable, the first of three iterations, we have defined a first set of requirements for the prototype MeMAD platform. This document defines the functional and non-functional (i.e., in terms quality, processing performance or system resilience) requirements of the MeMAD prototype system, based on input concerning the tools developed in WP2, WP3, WP4 and WP5 and based on the project's use cases from which many requirements are dictated. In addition to (non-)functional requirements, subsequent revisions of this document will also incorporate specific test scenarios and evaluation criteria to determine the performance of the prototype system.