Compendium Use Case Requirements Analysis

Deliverable Nr Title:	7.1.1 & 8.1.1 Use Case Analysis: Crowd Sourcing and Video Sharing
Delivery Date:	30 April 2014
Author(s):	David Riccitelli, Andrea Volpini, Patrick Aichroth, Grant Miller and Chris Lintott
Publication Level:	Public

Table of Contents

- Table of Contents
- Document History
- Documentation Information
- Introduction
- MICO requirements analysis methodology
  - Showcases (SC)
  - User Stories (US)
  - Technology Enablers (TE)
  - Datasets (DS)
  - Requirements (UC, NF) as links to system design / architecture
- MICO Showcases - Zooniverse
  - Background of UC partner(s)
  - Current status of project(s)
  - Overall MICO goals and exploitation
  - User Roles
  - Zooniverse - Generic issues [SC-ZOO]
  - Zooniverse showcase - Galaxy Zoo [SC-14]
  - Zooniverse showcase - Snapshot Serengeti [SC-16]
  - Zooniverse showcase - Plankton Portal [SC-10]
  - Zooniverse showcase - Worm Watch Lab [SC-11]
  - Zooniverse showcase - Crisis Response [SC-17]
  - Zooniverse showcase - Asteroid Zoo [SC-15]
  - Zooniverse showcase - Whale FM [SC-05]
- MICO Showcases - InsideOut10
  - Background of UC partner(s)
  - Current status of project(s)
  - Overall MICO goals and exploitation
  - High Level Application Flow
  - User Roles
  - InsideOut10 Music Showcase [SC-01]
  - InsideOut10 News Video Showcase [SC-02]
- Cross-Cutting user stories and technology enablers

Document History

Version	Name	Date	Remark
V0.1	Patrick Aichroth, Grant Miller, Andrea Volpini, David Riccitelli	20.03.2014	Initial versions (separate documents) for methodology, D7.1.1 and D8.1.1
V0.2	Andrea Volpini, David Riccitelli	24.04.2014	Merge, cleanup and extension of D7.1.1 parts
V0.3	Grant Miller, Patrick Aichroth	24.04.2014	Merge, cleanup and extension of D8.1.1, and methodology parts
V0.4	Grant Miller, Andrea Volpini, David Riccitelli, Patrick Aichroth	30.04.2014	Modifications wrt methodology, structuring of D7.1.1 and D8.1.1 descriptions, cleanup of user stories and datasets
V0.5	Patrick Aichroth, Rafa Haro	05.05.2014	Restructuring to include tech enablers and links to SotA
V0.6	Grant Miller, Patrick Aichroth, Emanuel Berndl, Thomas Kurz	07.05.2014	Introduction, table completion, insertion of new TE
V0.7	Patrick Aichroth, Grant Miller, David Riccitelli	09.05.2014	Renumbering of IDs, TE completions, minor modification wrt methodology description, inclusion of cross-cutting stories and enablers

Documentation Information

Item	Value
Author(s)	Grant Miller, Andrea Volpini, David Riccitelli, Patrick Aichroth
Document Title	Combined Use Case Requirements Deliverable
Actual Distribution Level	Public

Document Context Information

Project (Title/Number)	MICO - “Media in Context” (610480)
Work Package / Task	WP7 Use Case: Crowd Sourcing Platform WP8 Use Case: Video Sharing Platform
Responsible person and project partner	Grant Miller / University of Oxford Andrea Volpini / InsideOut10

Quality Assurance / Review

Name / Partner / QA Control /	John Pereira Salzburg Research
Comment	Review successful
Release	9 May 2014

Introduction

The following document summarizes the high-level requirements for the MICO project, which serve as a basis for all further R&D steps. The requirements result from an exchange between all partners within the consortium, aligning the user perspective and technical possibilities.

The document is structured accordingly:

- A description of the **requirements analysis methodology**, including a description of how “targets” for the project were selected, and how requirements analysis, system design and architecture are connected; this section will also introduce relevant key terms such as showcases (SC), user stories (US), datasets (DS), technology enablers (TE), use-cases (UC) and non-functional requirements (NF)

- Two **main chapters on the showcases, one for Zooniverse and one for InsideOut10**, include
    - description of **partner background and goals**, and existing partner platforms
    - current and future projects (“**showcases**”, SC)
    - respective high-level user requirements (“**user stories**”, US)
    - available content (“**datasets**”, DS)
    - technology demands (“**technology enablers**”, TE) resulting from the above

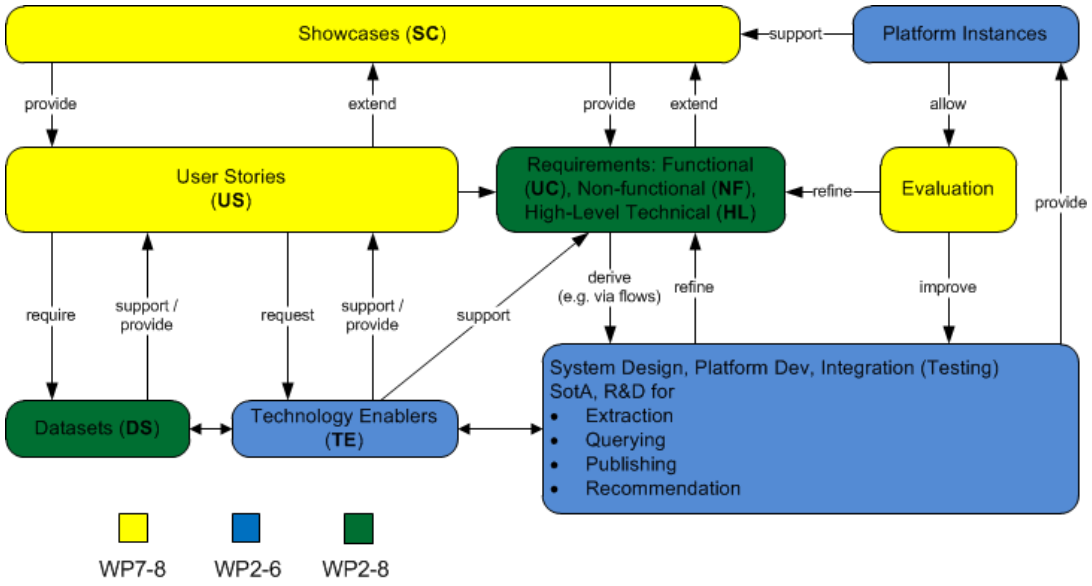
- One chapter describing a few **cross-cutting user stories and respective technology enablers**, and a glimpse on non-functional requirements, which will be discussed in more detail in the further R&D process.

MICO requirements analysis methodology

Responsible Partner / Author: FHG (Patrick Aichroth)

As for many other projects, MICO faces the challenge of identifying requirements, filtering and streamlining them in an interactive process between use case partners and technical partners, and breaking them down into actionable items that can be used for all R&D work in the project - in a sense, linking requirements to system design and integration.

The approach chosen is described in the diagram below:



The relevant entities showcases (SC), user stories (US), datasets (DS), technology enablers (TE), use-cases (UC) as functional requirements, and non-functional requirements (NF) depicted above will be discussed in the following sections. The sections will outline the specific methodology for the MICO project. This methodology combines “top-down” (abstract, high-level requirements are derived into technical requirements) as well as “bottom-up” aspects (e.g. feedback regarding technical feasibility leads to the modification of high-level requirements).

Showcases (SC)

The main starting point for requirements analysis are ***showcases (SC)***. While not representing requirements as such, MICO showcases represent current or planned projects of Zooniverse (ZOO) and InsideOut10 (IO). They provide the context for all further requirements, and represent slightly different combinations of goals, content sets, and user communities involved. After the brainstorming phase within the MICO consortium, a set of 18 possible showcases were identified. They were then prioritized based on technical feasibility, impact and availability of datasets. This resulted in a total of 9 showcases as potential “targets” for the MICO project.

These were further prioritized into groups (A and B) considering

- expected impact and industrial, scientific, and societal value
- variety and volume of cross-media content
- existence of "MICO aspects", e.g. potential of accuracy and robustness improvements due to cross-modal analysis
- technical feasibility regarding availability of textual and audio-visual media extraction, publishing, querying, recommendation technologies, and cross-cutting technical platform requirements.

Group A was considered for a first iteration under the precondition that some might be dropped during the process considering the R&D risks involved:

1. SC-02 News Videos [IO]
2. SC-10 Plankton Portal [ZOO]
3. SC-14 Galaxy Zoo [ZOO]
4. SC-16 Snapshot Serengeti [ZOO]
5. SC-01 Music Promotion [IO]

Group B provides "backup" showcases that might or might not be used in the further process:

6. SC-11 Worm Watch Lab [ZOO]
7. SC-05 Whale FM [ZOO]
8. SC-15 Asteroid Zoo [ZOO]
9. SC-17 Crises Response [ZOO]

These 9 showcases are further described in the following main chapters for D.7.1.1 (Zooniverse) and D8.1.1 (InsideOut10).

User Stories (US)

The ***user stories (US)*** are derived from the showcases. At the same time, new user stories may extend (existing) showcases. User stories serve as a starting point for requirements analysis. Defining the high level requirements are essentially a technology-free process, for example:

```
As a <role>, I want <goal/desire> (so that <benefit>), e.g.
```

- "As a user, I want to search for my customers by their first and last names, so that I can find their contact information."
- "As a non-administrative user, I want to modify my own schedules but not the schedules of other users."
- "As a user closing the application, I want to be prompted to save if I have made any change in my data since the last save."

Throughout this document, the user stories are captured using the structure below to track the providing partner (PP) and the entries themselves via ID:

ID	PP	Description
<ID>	<providing partner>	<description>

The advantage of such a high-level approach is to ensure proper tracking and ownership of each user story. At the same time, a lot of information can be derived in the further requirements analysis process. However, at least in the case of MICO, such richness *only* emerges if the respective context, e.g. the showcase description, is aligned with the user story.

In MICO, user stories can take the following forms:

- a showcase level, applying to specific InsideOut10 or Zooniverse showcases
- a use-case partner level, applying to all InsideOut10 or all Zooniverse showcases
- a platform-level, applying across the board; such user stories were often motivated by technology enablers that are generic, and provided by technology partners

User stories can always be tied back to specific “demands” from a user perspective, and can be used to align all downstream development work. A updated and prioritized list of user stories can also be aligned and cross-checked with technology enablers and datasets, and detailed into one or several system use cases (for all: see next subchapters). With this we can ensure that all R&D work serves a defined purpose, and reduces discrepancies between user and technical perspectives.

Technology Enablers (TE)

Technology enablers (TE) are the technical counterpart to user stories. They are used to express which technologies are required to support user stories and showcases. By linking TE to US, they make sure that what is “wanted” from a user perspective will be matched by enabling technologies, thereby creating the link that all further R&D work can refer to.

The first practical example of this is the SotA analysis, which was developed with the US and TE in mind: The SotA sections mention the related TE. In this manner, it was ensured and cross-checked that (a) the SotA analysis is aiming at things actually required (US), and (b) that demands (TE) stemming from US were met with corresponding SotA sections. In a similar manner, TE serve for the upcoming system design and architecture decisions to cross-check technical work against demands. Of course, this requires that the lists be updated regularly, as minor deviations are to be expected in any R&D project.

It is important to note that, in the following sections, the mentioning of a TE does not imply that it is possible to implement, but must first be evaluated against project

constraints. Indeed, it is to be expected that during the further R&D process, we will learn that a number of TE cannot be realized, at least not within the project scope.

Datasets (DS)

Datasets (DS) are crucial not only to support user stories and use cases, and hence showcases, but also for training and testing purposes. Apart from supporting a showcase as such, content is crucial to train, and in some cases, develop textual and audio-visual extractors for the specific demands of a showcase. Datasets can also trigger new ideas regarding specific media extractor, publishing, querying and recommendation approaches.

Requirements (UC, NF) as links to system design / architecture

For further development purposes, MICO uses the following types of ***requirements*** for the system:

1. Functional requirements aka System ***Use Cases (UC)***: Use Cases capture the functional requirements that serve as the core starting point for system design and implementation, for instance by providing basic flows which will then be detailed into UML sequence diagrams, which specify interactions between actors and components, thereby also helping to identify component roles and descriptions, APIs/interfaces, etc. They also provide a main reference for system evaluation. In contrast to User Stories (US), they can and probably will include technology implications, e.g. imply that certain components or subsystems do exist.

2. **Non-functional requirements (NF)**: Non-functionalities can be specific to a specific Use Case (hence the entry on non-functional requirements in the use case tables below), but in most cases, they are more generic, applying to subsystems or the whole system. Examples: Usability, Testability, Performance, Scalability, Security, Privacy, Portability, Interoperability, Maintainability, Modifiability, Integrability, Extensibility.

Hence, these requirements provide the glue to get from requirements to system design and architecture, and respective aspects will be picked up in D6.1.1.

MICO Showcases - Zooniverse

Responsible Partner / Author: UOX / Grant Miller

Background of UC partner(s)

Zooniverse is the largest and most popular online citizen science platform. The team is based in the astrophysics department at the University of Oxford and the Adler Planetarium in Chicago. It started in 2007 with the Galaxy Zoo project and now operates over 20 separate projects across many fields of research such as astrophysics, climatology, ecology, biology and history. Each project is built around the idea that volunteers can access the website and classify data (images, video, audio) by performing basic recognition tasks that cannot be easily performed automatically by computers. The form of the task and size of the dataset is unique to each project so they normally take between six months and one year to to build. Zooniverse receives 30-40 proposals for new projects every year but at the moment we are only able to build a small fraction of them.

Current status of project(s)

The are currently 22 active projects running on the Zooniverse platform, with a further 5 at various stages of production. To-date 1.1 million people have registered accounts with the Zooniverse, however we do not require that our volunteers registers an account to be able to take part in our projects.

Overall MICO goals and exploitation

Zooniverse is focused on validating MICO results within the context of citizen science platforms and the scientific research community.

By definition, Zooniverse projects have large amounts of data which require many volunteers to analyse. Anything that can refine the process is extremely useful as it will lead to the science goals being met faster. MICO technologies should be able to help in various ways

- pre-filtering and removing files that do not need to be viewed by the volunteers
- image/video/audio/textual analysis on the data, metadata and associated text comments to retrieve information that will contribute to the classifications
- grouping of files that will allow certain types to be delivered to specific volunteers

The main goals are to increase the speed, accuracy and efficiency of the analysis and also to create a system that will stimulate higher levels of motivation among volunteers.

User Roles

This and the following subsections provide user stories for Zooniverse. The associated user roles used will be

- *Zooniverse Administrators*: are the IT managers with complete access to the solutions.
- *Volunteers*: are the final end-users, i.e. the people who are classifying the data on the various Zooniverse project websites.

Zooniverse - Generic issues [SC-ZOO]

Responsible Partner / Author: UOX / Grant Miller

Description

Due to the common goals for all Zooniverse projects, there are generic user stories that apply to all Zooniverse showcases. These are described in the following, including the respective technology enablers.

Generic User Stories

ID	PP	Description
US-27	UOX	As a Zooniverse admin, I would like to be able to assess how interesting / appealing / complex a picture is based on automated analysis, citizen annotations, and comments on 'Talk'
US-28	UOX	As a Zooniverse admin, I would like to be able to detect when a scientist should be prompted to look at a subject, based on annotations and information from 'Talk' comments
US-29	UOX	As a Zooniverse admin, I would like to identify volunteer types
US-49	UOX	As a Zooniverse admin, I'd like to know when I should interrupt a volunteer perhaps based on the recent subjects they have viewed, or how many classifications they have performed
US-50	UOX	As a Zooniverse admin, I'd like to know whether I should interrupt a volunteer with text, an image, or a video
US-51	UOX	As a Zooniverse admin, I'd like to know when I should educate a volunteer

US-52	UOX	As a Zooniverse admin, I'd like to know whether I should educate a volunteer with text, an image, or a video
US-53	UOX	As a Zooniverse admin, I'd like to know which piece of education I should give to a volunteer
US-54	UOX	As a Zooniverse admin, I'd like to know when a volunteer has made an interesting comment on a subject
US-55	UOX	As a Zooniverse admin, I'd like to know when Zoonibot (our bot that interacts with our volunteers in the 'talk' areas of the projects) should comment on a subject
US-56	UOX	As a Zooniverse admin, I'd like to know when Zoonibot should give an explanation
US-57	UOX	As a Zooniverse admin, I'd like to know what Zoonibot should say to a volunteer
US-58	UOX	As a Zooniverse admin, I'd like to be able to group subjects (i.e. images, videos or audio files) by similarity
US-59	UOX	As a Zooniverse admin, I'd like to be able to recommend different projects to volunteers based on their previous experiences

Datasets required and available

Name	Description
ID	DS-04
Partner/Person	Zooniverse / Grant Miller
Description	Data is provided in the form of a MongoDB dump for each use case. The entry for each file in the database contains the metadata associated with it and an AWS url link for the actual file itself. The files for all use cases are JPG images, except the Worm Watch Lab and Whale FM use cases which are MP4 video and MP3 audio files respectively.
Formats	JPG, MP4, MP3

Generic Technology Enablers

TE#	TE application	Applies to
TE-201	Feature extraction for image appeal and similarity	US-27, US-58
TE-211	Image appeal and similarity are classified by means of machine learning or feature space distance based on extracted low level features	US-27, US-58
TE-212	Detect questions directed to researchers and other forum members	US-27, US-28, US-51, US-55, US-56
TE-213	Distinguish between different types of discussions and classify image appeal	US-27, US-28, US-29, US-54, US-55, US-58
TE-215	Derive features from text fields to be used in cross-media classification	US-58
TE-216	Text cleaner to remove markup and standardize citation, punctuation, etc. Necessary as a preprocessing step for e.g. the phrase-structure parser	US-27, US-28, US-55
TE-217	Phrase-structure parser	US-27, US-28, US-55
TE-218	Interactive wrapper generator	US-27, US-28, US-29
TE-219	Graph operation toolkit to model semantic features	US-54
TE-220	Extract keywords related to e.g. species or activities	US-27, US-28, US-29, US-55, US-56, US-59
TE-401	Support User with image snippets (e.g. that shows a specific animal / galaxy) for training	US-52, US-57
TE-403	Support image metadata retrieval	US-52
TE-404	Allow users browsing the database for images, that shows specific scenes (e.g. a group of animals,	US-53

	galaxies) and support them with useful metadata (e.g. what are the characteristics of this galaxy)	
TE-405	Support user with graphical user interface	US-53
TE-407	Train the users by showing them images and metadata with a high similarity	US-53
TE-501	User activity and context monitor; collect user, usage and context information	US-29, US-49, US-50, US-51, US-52, US-53, US-54
TE-502	Project similarity calculator	US-59
TE-503	User Similarity calculator; determine the similarity of volunteers based on their activities on the projects	US-29
TE-504	Volunteer-type analysis; determine the characteristics of a volunteer based on their activities on the projects	US-29, US-49, US-50, US-51, US-52, US-53, US-56, US-57
TE-505	Subject-type analysis; determine the characteristics of the subjects (images, audio, video files) based on their content	US-57, US-58
TE-506	Cross-modal content recommender; determines which content should be delivered to a specific volunteer	US-49, US-50, US-51, US-52, US-53, US-55, US-56, US-57
TE-507	Item similarity calculator; determines the similarity of media items	US-29, US-58

Zooniverse showcase - Galaxy Zoo [SC-14]

Description

The task involves classifying the morphology of galaxies in images from various ground and space-based telescopes. Help scientists better understand the evolution of galaxies and our Universe. Volunteers are presented with an image containing a galaxy at its centre and then follow a decision tree by clicking on-screen buttons to classify the galaxy.

There are some existing publications and previCUS work on automated analysis of Galaxy Zoo data:

- http://labs.adsabs.harvard.edu/adsabs/abs/2012MNRAS.421.2277L/
- http://labs.adsabs.harvard.edu/adsabs/abs/2011A%26A...532A..74B/
- http://labs.adsabs.harvard.edu/adsabs/abs/2005ApJ...635L..29P/
- http://labs.adsabs.harvard.edu/adsabs/abs/2010ApJS..186..427N/

Kaggle recently ran a challenge with a prize for the best algorithm for image analysis of older Galaxy Zoo data - http://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge

Specific User Stories (in addition to the generic Zooniverse user stories)

ID	PP	Description
US-30	UOX	As a Zooniverse admin I would like to be able to pre-classify Galaxy Zoo images using simple low-level image features (brightness, size, symmetry, concentration, spirality, number of objects, clumpiness of the galaxy)

Datasets required and available

Name	Description
ID	DS-05
Partner/Person	Zooniverse / Grant Miller
Description	Data is provided in the form of a MongoDB dump. The entry for each file in the database contains the metadata associated with it and an AWS url link for the actual file itself. There are 365,000 individual images.
Formats	JPG

Specific Technology Enablers (in addition to the generic Zooniverse enablers)

TE#	Specifics of TE application	Applies to

TE-201	Low-Level Visual Feature Extraction: Some basic features can be combined to applications specific feature in order to give an estimate; additional annotations for the properties might be required	US-30

Zooniverse showcase - Snapshot Serengeti [SC-16]

Description

The task involves identifying various animals and their behaviour from camera trap images in the Serengeti National Park. Helping scientists better understand how the species interact with each other. Volunteers have to identify the animals from a list of 48 species and give information of their numbers and activities.

Specific User Stories (in addition to the generic Zooniverse user stories)

ID	PP	Description
US-31	UOX	As a Zooniverse admin I would like to be able to automatically detect Snapshot Serengeti images with no classifiable animals in them
US-32	UOX	As a Zooniverse admin I would like to be able to perform automatic image series detection for the case of timestamping malfunction in Snapshot Serengeti images
US-33	UOX	As a Zooniverse admin I would like to be able to perform automatic animal species pre-classification in Snapshot Serengeti (48 species)
US-34	UOX	As a Zooniverse admin I would like to be able to perform automatic animal attribute pre-classification in Snapshot Serengeti
US-35	UOX	As a Zooniverse admin I would like to be able to perform automatic animal number detection in Snapshot Serengeti

Datasets required and available

Name	Description
ID	DS-06

Partner/Person	Zooniverse / Grant Miller
Description	Data is provided in the form of a MongoDB dump. The entry for each file in the database contains the metadata associated with it and an AWS url link for the actual file itself. There are 1.5 million images available.
Formats	JPG

Specific Technology Enablers (in addition to the generic Zooniverse enablers)

TE#	Specifics of TE application	Applies to
TE-201	Low-level feature extraction for RoI detection	US-35, US-31
TE-202	Automatic detection of image with no classifiable animals in it (Semi-)automatic animal detection (Semi-)automatic animal number detection	US-31, US-33, US-35
TE-203	Automatic animal species pre-classification Automatic animal attribute pre-classification	US-33, US-34
TE-210	Automatic image series detection for the case of timestamping malfunctions	US-32
TE-401	Support spatial media fragment e.g for counting the number of animals on query-time	US-35
TE-403	Support regional query functions to identify and aggregate regional fragments e.g. return a lion right beside a gazelle	US-32, US-35
TE-411	Support for pivot vocabularies (diverse datasets for animal classification)	US-33

Zooniverse showcase - Plankton Portal [SC-10]

Description

The task involves classifying small sea creatures by species, size and orientation in images taken at different depths in the ocean. Helping scientists measure the health of the ocean. Volunteers have to measure the dimensions of any creature they see in the image and then identify the species with help from a decision tree. There are 23 possible species to choose from.

Specific User Stories (in addition to the generic Zooniverse user stories)

ID	PP	Description
US-36	UOX	As a Zooniverse admin I would like to be able to perform automatic segmentation / localization of plankton (for manual classification) in Plankton Portal
US-37	UOX	As a Zooniverse admin I would like to be able to automatically detect Plankton Portal images with no classifiable plankton in them
US-38	UOX	As a Zooniverse admin I would like to be able to perform automatic detection of plankton size / orientation in Plankton Portal images
US-39	UOX	As a Zooniverse admin I would like to be able to perform automatic pre-classification of plankton species (23 species) in Plankton Portal images

Datasets required and available

Name	Description
ID	DS-07
Partner/Person	Zooniverse / Grant Miller
Description	Data is provided in the form of a MongoDB dump. The entry for each file in the database contains the metadata associated with it and an AWS url link for the actual file itself. There are 338,000 images available.
Formats	JPG

Specific Technology Enablers (in addition to the generic Zooniverse enablers)

TE#	TE application	Applies to
TE-201	“Empty” image detection based on low level features	US-37
TE-202	Object detection for pre-segmentation	US-36, US-38, US-39
TE-203	Species classification of plankton (may depend on proper segmentation though)	US-39
TE-401	Provide images and localisations	US-36
TE-403	Support search on image and metadata	US-36
TE-408	Search for media that may contain something of interest to a certain probability	US-36
TE-410	Present search results (media items and fragments)	US-36

Zooniverse showcase - Worm Watch Lab [SC-11]

Description

The task involves watching videos of nematode worms and marking when in the video they lay an egg. Data gathered helps scientists better understand how genes work. The volunteer has to watch a 30 second video of a single worm and press the z-button on their keyboard when they see an egg appear.

The videos are long and often nothing interesting happens. If the MICO technologies could be used to identify interesting segments and only show these parts to the volunteers it would save time and possible increase motivation to take part in the project.

Specific User Stories (in addition to the generic Zooniverse user stories)

ID	PP	Description
US-40	UOX	As a Zooniverse admin I would like to be able to perform automatic worm motion analysis to reduce video duration in Worm Watch Lab

US-41	UOX	As a Zooniverse admin I would like to be able to automatically detect the time in a Worm Watch Lab video where the worm lays an egg

Datasets required and available

Name	Description
ID	DS-08
Partner/Person	Zooniverse / Grant Miller
Description	Data is provided in the form of a MongoDB dump. The entry for each file in the database contains the metadata associated with it and an AWS url link for the actual file itself. There are 74,000 videos available.
Formats	MP4

Specific Technology Enablers (in addition to the generic Zooniverse enablers)

TE#	TE application	Applies to
TE-201	Temporal global feature for motion analysis	US-40
TE-228	Moving shape analysis	US-41
TE-401	Spatio-temporal search (is the worm moving, when and where is it moving)	US-40, US-41
TE-412	Has the exact position of the worm changed to a certain percentage within a time period	US-40

Zooniverse showcase - Crisis Response [SC-17]

Description

The task will involve looking through satellite image data in close to real-time in the aftermath of a humanitarian crisis (such as a major flood or earthquake) to identify any features that will aid the rescue efforts. Features could include things such as blocked roads, fires, temporary shelters, groups of people. The annotations made by the volunteers will be rapidly assessed for probability and that information will be passed onto the emergency services (and possible through other avenues like social media) as soon as possible.

Specific User Stories (in addition to the generic Zooniverse user stories)

ID	PP	Description
US-42	UOX	As a Zooniverse admin I would like to detect important features such as areas of flooding, damage, temporary shelters, blocked roads etc. in the satellite images
US-43	UOX	As a Zooniverse admin I would like to be able to assess the probability/weight of a volunteer's classification based on their experience

Datasets required and available

Name	Description
ID	DS-09
Partner/Person	Zooniverse / Grant Miller
Description	Data will be in the form of satellite imagery.
Formats	JPG

Specific Technology Enablers (in addition to the generic Zooniverse enablers)

TE#	TE application	Applies to
TE-202	Object and Animal Detection; with (registered) reference imagery, this might be possible	US-42

TE-407	Find users that match a certain skill pattern	US-43
TE-408	Return the confidence how good a pattern matches a certain (metadata-) structure	US-43
TE-504	Volunteer-type analysis; determine the characteristics of a volunteer based on their activities on the sites	US-43

Zooniverse showcase - Asteroid Zoo [SC-15]

Description

The task involves searching through optical images to discover and track near-Earth asteroids. Volunteers look at multiple images of the same part of the sky taken minutes apart and are asked to identify anything that is moving with respect to the fixed stars.

Specific User Stories (in addition to the generic Zooniverse user stories)

ID	PP	Description
US-44	UOX	As a Zooniverse admin I would like to be able to perform pre-filtering of the images to remove artefacts such as bright stars and camera read errors
US-45	UOX	As a Zooniverse admin I would like to be able to detect moving and transient objects in the images

Datasets required and available

Name	Description
ID	DS-10
Partner/Person	Zooniverse / Grant Miller
Description	Data is provided in the form of a MongoDB dump. The entry for each file in the database contains the metadata associated with it and an AWS url link for the actual file itself. There will be millions of images available.

Formats	JPG

Specific Technology Enablers (in addition to the generic Zooniverse enablers)

TE#	TE application	Applies to
TE-201	Temporal Low-Level Visual Feature Extraction, to possibly detect moving and transient objects (moving asteroids across pictures)	US-45
TE-205	A/V Error Detection and Quality Assessment. especially for camera read errors, to remove artefacts (bright star / spike, overexposure also from stars in other images, camera read errors, variable starts changing in brightness)	US-44
TE-412	Detect moving objects of a certain structure	US-45

Zooniverse showcase - Whale FM [SC-05]

Description

The task involves listening to recordings and look at frequency graphs of whale sounds and match them by similarity. Trying to understand whale 'dialect'. Volunteers listen to a short sound clip, for which they can also see a graph of frequency, and are asked to match it to the most similar clip from a list of 36 that a computer algorithm has identified as being the most similar.

Specific User Stories (in addition to the generic Zooniverse user stories)

ID	PP	Description
US-46	UOX	As a Zooniverse admin I would like to be able to remove background noise and distracting sounds from the audio files, e.g. noise from boats
US-47	UOX	As a Zooniverse admin I would like to be able to group whale calls that are similar

US-48	UOX	As a Zooniverse admin I would like to be able to identify the number of whales heard in a single audio file

Datasets required and available

Name	Description
ID	DS-11
Partner/Person	Zooniverse / Grant Miller
Description	Data is provided in the form of a MongoDB dump. The entry for each file in the database contains the metadata associated with it and an AWS url link for the actual file itself. There are 16,000 audio files available.
Formats	MP3, WAV

Specific Technology Enablers (in addition to the generic Zooniverse enablers)

TE#	TE application	Applies to
TE-222	Noise removal	US-46
TE-223	Audio classification (based on length, spectral density, etc.) for single whales and whale groups	US-47, US-48
TE-406	Support boolean queries and counting of informational properties	US-48

MICO Showcases - InsideOut10

Responsible Partner / Author: IO10 / Andrea Volpini, David Riccitelli

Background of UC partner(s)

InsideOut10 is an Italian start-up and consulting firm with an extensive experience on web publishing and media delivery platforms.

InsideOut10 is major shareholder of Interact Egypt L.L.C. (soon to be InsideOut10.Today) a startup based in Cairo, Egypt and focused on content management solutions for broadband and mobile networks.

The intrinsic mix of Internet technologies and communication skills is at the very base of the projects carried out by InsideOut10 for a wide range of clients in Europe, US and the Middle East.

InsideOut10 client-base include mobile operators, ISPs, broadcasters, media companies, energy utilities and content providers as well as education & research bodies and here follows a selection of clients interested in the development of the MICO platform:

1. **Mobile Operators & TLC sector**
    a. Wind (Italy)
    b. A1 (Austria)
    c. Mobinil (Egypt)
    d. LinkDotNet (Egypt)
    e. Tunisiana (Tunisia)
    f. Djezzy (Algeria)
    g. Mobilink (Pakistan)
    h. Mobily (Saudi Arabia)
    i. FastTelco (major ISP in Kuwait)
2. **Media**
    a. 5FM Radio (South Africa)
    b. Kuwait National Radio & Television (Kuwait)
    c. Melody TV (Egypt)
    d. West African Content (Ghana)
    e. Saudi Research and Marketing Group (Saudi Arabia)
    f. Digital Media Co. (Egypt)
3. **Marketing & Advertising**
    a. AirMob (Ghana)
    b. JWT (Ghana)
4. **Energy** Enel (Italy)
5. **Sports** Al-Hilal club (Saudi Arabia)
6. **Education**
    a. German University in Cairo

    b. Berlitz
    c. Università La Sapienza
7. **Tourism** Savoy hotels
8. **Technology**
    a. RealNetworks (US)

InsideOut10 is active in the Research and Development with joint programmes developed in partnership with Sapienza Innovazione (Università La Sapienza), the Italian Council for Research (CNR) and the Department of Information Engineering, Computer Science and Mathematics at Università degli Studi dell'Aquila (DISIM UNIVAQ).

InsideOut10 main competencies include:

- Development of digital asset management solutions,
- Dynamic Semantic Publishing,
- Linked Data publishing,
- Design, planning and development of strategic online positioning for brands.

InsideOut10 as MICO's industry partner will validate the project's results in the products and solution: Helix Cloud (video middleware developed in partnership with RealNetworks), WordLift (a semantic plug-in for WordPress) and a mobile instant video recording application for user generated content being developed by the Egyptian team and named "شوف" (it means "Look here" in arabic and it is pronounced "*Shoof*").

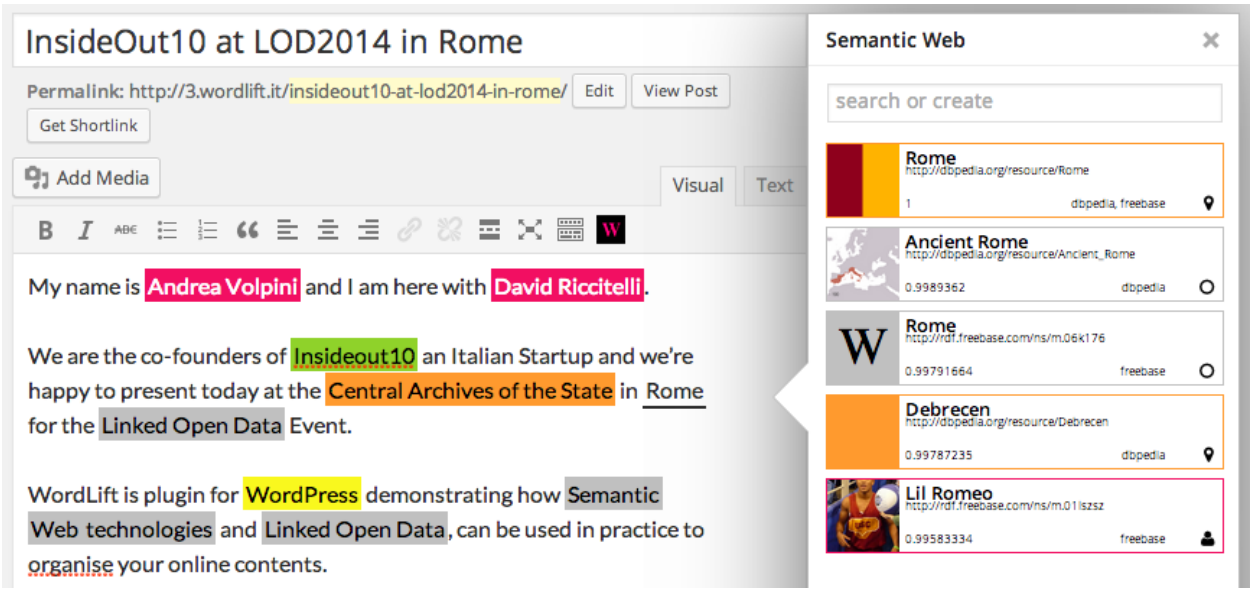
Current status of project(s)

InsideOut10 created Helix Cloud in partnership with RealNetworks Inc., a middleware platform to benefit existing and new audio/video operators that need to exploit the value of their content with streaming media. Helix Cloud integrates with video servers and exposes APIs to fully automate user management, content ingestion, content publication, channel management, conditional access and reporting[1].

InsideOut10 also created WordLift to provide on WordPress (one of the World's largest Open Source WCM) a uniform access infrastructure for data using the Semantic Web stack and Linked Data principles[2]. The goal of WordLift is to make content available for consumption by humans and machines using standardized formats and access mechanisms, providing reuse of open data easy and unified. WordLift connects different sources (both internal and external) providing interoperable data, forming a global graph that can be traversed by clients to discover new information. WordLift provides entity management facilities and integration with external rdf triple stores to expose the site contents as Linked Open Data. Content discovery functionalities

[1] An online presentation of Helix Cloud is available on SlideShare: http://slideshare.net/davidriccitelli/helixcloud-showcase
[2] Linked Data on the W3C website: http://www.w3.org/standards/semanticweb/data

are provided by interacting with the global graph and embedded microdata is added for search engine optimisation using schema.org.



A sample screen of WordLift plugin for WordPress

The hereby described use cases are formed by the experience gathered on large clients (mainly telecom operators providing value added services over broadband and mobile networks) and mid-sized organisations producing vast amount of contents on the Web using open source web content management systems. The general aim is to extend the capabilities of semantic analysis, currently limited in WordLift to textual contents, to audio/visual media and to complement the Helix Cloud platform with advanced semantic capabilities, extending the existing workflow of content analysis, publication, search and discovery.

An high-level flow of data is therefore the following:

The ingestion and publication stages take place inside the respective platforms. The other stages, analysis, search and discovery, see the integration of the platforms with the MICO platform.

The same flow applies to the user-generated content (UGC) application “Shoof”, with the main difference that it relates with users on the ground capturing short video clips with their handsets and publishing these videos using Helix Cloud. In this context it becomes crucial to introduce video-quality assessment functionalities, nudity detection and above all cross-referencing news contents to determine content relevance based on the geo-localization of the UGC clips.

Overall MICO goals and exploitation

With platform installations featuring terabytes of contents and millions of views per month, operators and organisations have an increasing need for boosting revenues by promoting content to a variety of users. They need to establish discovery patterns that engage casual/indifferent users leveraging on context knowledge, therefore matching the best content according to the user profiles and content similarity.

The pervasiveness of communication, massive shift towards mobile devices, requires the creation of intuitive natural user interfaces for content consumption which increase the use of offered media through simple actions related to natural, everyday human behaviors.

The current status quo where content is locked in different platform (silos) needs to evolve and content offerings shall be made seamlessly accessible across different channels. The ability to analyse ingested content, providing metadata, categorization and complementary information, will enable a holistic access to media that is currently missing. Enriched metadata can be combined with profiling data to further customize the user experience and extend content consumption by repurposing matching content.

InsideOut10 is focused on validating MICO results within the context of telecom operators, media broadcasters and web content publishers.

Solutions such as WordLift and Helix Cloud or UGC applications like Shoof traverse different market segments ranging from large network operators (both mobile and broadband) - that need to keep media contents within their datacenters - to web publishers and start-ups looking for a more flexible and a scalable approach and generally favorable to cloud-base solutions.

The MICO platform must be therefore modular and provide the ability to install “satellite” components on-premises. These components will allow content to be analyzed and annotated in the provider domain (hosted by the provider).

The information resulting from the content analysis shall be made available to the content provider in order to review the metadata, fine tune it and eventually integrate it.

Here is a high level overview of the goals and issues we can foresee at this stage:

- Platform scalability: content shall be analyzed and annotated in a distributed manner using on-premises “satellite” components to avoid transferring large files and to secure media contents; the general platform architecture shall make it possible to add additional nodes according to the platform load, with particular reference to the media processing nodes which might have a major load on the platform itself.

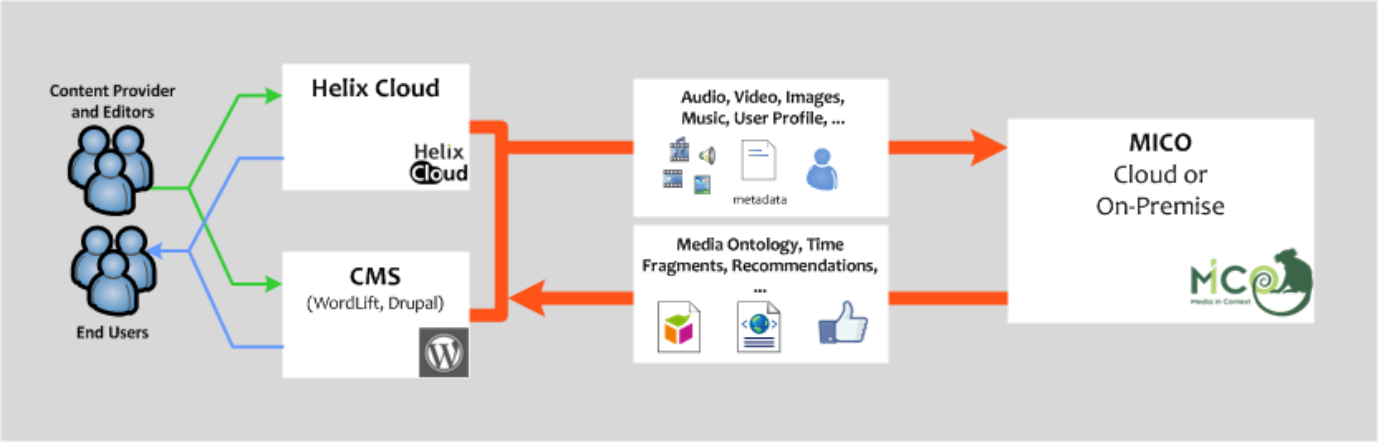
- Content annotation: the platform shall support (semi-)automatic analysis enabling content editors intervention to enrich metadata and interlink content; therefore content annotation is either automatic as a result of the analysis provided by the platform components, or manual as a result of editors describing the entity annotations on the media contents, or a combination of the two.
- Content recommendation: the platform shall discover and re-purpose content using content analysis and item similarity (item similarity shall include both the analysis of the metadata and the analysis of the actual media). Recommendation shall also be based on user profile (user similarity using the following parameters: gender, age, occupation and location).

High Level Application Flow

As an example, we describe a high-level application flow for Helix Cloud.

On the left side there are Content Providers and Content Editors ingesting content to the Helix Cloud platform. Helix Cloud integrates with MICO providing content analysis, search and discovery capabilities.

The semantic data is stored in the CMS data-store according to the specific CMS architecture and formats for presentation and consumption by the editors and the end-users.



User Roles

Several user roles are defined within the Helix Cloud solution. They are used throughout the following user stories. This is the list of roles:

- *Administrators*: The IT managers which complete access to the solution, they can create new organization accounts and new users, and assign roles,

- *Content Providers*: These participate in organization groups sharing the same disk quota and aggregates statistics, they receive a username and a password that enables them to ingest new content by using the integrated FTP Server - in the *user-generated content* case, the *content providers* and the *users* roles may overlap,
- *Report Users*: A special role for providers that can access the platform reports. They may access the full organization reports or only those related to their content,
- *Users*: These are the final end-users of the solution, that connect to a portal or an app which is publishing the media ingested by the content providers. Users are usually profiled by an application server and go through a payment gateway or subscription before they can actually access the content.

InsideOut10 Music Showcase [SC-01]

Description

The *music showcase* focuses on the needs of TLC operators, content portals, large content providers that create value by providing their subscribers, users and customers with diverse music datasets where music means audio-only clips, video clips and all their derivatives such as ringtones, ringback-tones and related materials such as phone wallpapers.

Most of the service platforms in use today deliver value added contents using multi-channel portals working over the Web, WAP, SMS and IVR. Content is organized vertically in categories and subcategories, and horizontally in bouquets that gather together related assets. End-users can use the content according to different business models, such as pay-per-user or unlimited monthly subscriptions.

The features provided by MICO will greatly enhance the users experience by providing tailored access to the vast amount of assets and actively proposing new content according to the user preferences and previous purchases. This shall increase both the customer satisfaction and media consumption.

User stories

ID	PP	Description
US-01	IO10	As a user, I want search and discover music using content identification or fingerprinting.
US-02	IO10	As a user, I want search and discover music based on my profile (gender, age, location expressed at city/region level and optionally occupation).

US-03	IO10	As a user, I want search and discover music based on other users similar to myself (user similarity).
US-04	IO10	As a user, I want search and discover music or ringback tones similar to a given music video (or to a given audio track) based on content similarity (item similarity); similarity is determined both by low-level features as well as by image tags.
US-05	IO10	As a user, I want search and discover music by exploiting enriched metadata and image tags to find related artists, new genre, or other relevant listening path. Few examples: ● all artists born in the location I'm visiting ● all artists connected to the artist I'm listening to by a common music background or by previous collaborations ● all tracks from the same genre recorded during the same period (1-12 months range) and around the same area (region)

Following is an concrete example case, where a content provider ingests content to the platform:

Context	Content providers upload audio, video on the platform - RBTs are uploaded on a specialized system (in this case only metadata is available on the platform).
Ingestion: metadata	Content providers ingest metadata along with files, e.g.: ● title (1) ● author (1) ● description (optional) ● genre (optional) ● content category (1+) ○ content subcategory (1+) ● copyright holder (optional) ● price Metadata is typically messy and incomplete, suffering for bad manual curation. In most cases only title, content category, content subcategory and price are available.

Ingestion: Audio/Video	Helix Cloud/CMSes send to MICO an audio/video asset for analysis along with the available metadata.
Ingestion: RBT	CMSes send to MICO the RBT metadata (see above - actual media may or may not be available) for analysis. Analysis is then performed on metadata only.
Ingestion: UGC	Content provider is an end user ingesting a media and additional metadata: ● location (coordinates) ● user gender ● user age ● user occupation (if available) The metadata on the user is derived from the social login of the application.
Results	MICO replies with: ● metadata of the content item (for a list of metadata properties, see Ontology for Media Resource 1.0 from W3C), ● Sequence detection data (in the form of Media Fragments), ● Entity annotations (e.g. People, Organizations, Places, Products) recognized in the content item (and related to the Media Fragments), ● Similarity and recommendations. Helix Cloud/CMSes might send a context vocabulary.
	Content editor verifies the results and accepts them.

Datasets required and available

Several sample files are provided, they include:

- sample full-track music files from a radio music portal that offers mobile services to its listeners,

- sample video clips,
- ringtones, formats optimized for mobile phones (see next table), the duration of which is no longer than 30 seconds,
- images suitable for use on mobile phones as screen wallpapers.

An actual dataset of data used in live systems is provided as reference for development and testing:

Name	Description
ID	DS-02
Partner/Person	IO10/ David Riccitelli, Andrea Volpini
Description	Several value added contents used in live TLC platform for resale to end customers. Divided in music (full-track songs), videos (video clips), ringtones (short tracks, usually less than 30 seconds, suitable as ringtones), images (such as those used for screen background). Some metadata (such as file category) may be made available in related spreadsheets or CSV files.
Formats	music: MP3 files; videos: 3GP files; ringtones: WAV, MP3, MMF, AMR, AAC; images: PNG, JPG, GIF

Technology Enablers

TE#	TE application	Applies to
TE-207	Speech-Music Discrimination, to filter irrelevant segments and improve visualisation / navigation	US-01, US-02, US-03, US-04, US-05
TE-208	Music Annotation, which provides data for all types of recommendations	US-02, US-04, US-05
TE-209	Music Similarity, which provides data for all types of recommendations	US-01, US-02, US-04, US-05
TE-215	Extract keywords from comment fields that may be helpful for genre analysis	US-05

TE-225	Robust audio identification based on audio fingerprint extraction and matching; also needs access to a respective database	US-01
TE-402	Find music based on semantic descriptions via fulltext search	US-05
TE-406	Find music based on semantic descriptions via boolean matching	US-05
TE-407	Provide a music set similar to a music item	US-02, US-03
TE-408	Provide music, whereby metadata is similar to a certain pattern	US-02, US-03
TE-409	Provide pattern-building with magic properties, e.g. User magic:hasMusicPattern p1 => return music with pattern p2, whereby p1 similarTo p2	US-02, US-03
TE-501	User activity and context monitor	US-02, US-03, US-05
TE-506	Cross-modal content recommender; determines which content should be delivered to a specific user	US-01, US-02, US-03, US-04, US-05
TE-507	Item similarity calculator; determines the similarity of media items	US-01

InsideOut10 News Video Showcase [SC-02]

Description

The *news showcase* is focused on the production and consumption of videos created by large audiences as user-generated contents, both professional or not, and their publication on community portals or institutional news portal.

The case therefore encompasses both the UGC application *Shoof* where the video production is mainly by the general public using a mobile handset, and a news-gathering agency such as AP, that features a large archive of data from its associates.

The challenges are on different stages of the production-consumption workflow. Beginning with the production, MICO features will enable automated quality control of ingested files for any evident issue with audio and video (for example, missing audio, audio/video sync issues, low quality video, etc.). MICO will also provide sequence detection and thumbnails generation to enhance the accessibility of content and power preview and smooth navigation inside video contents. From the consumption side, assets enriched with MICO semantic analysis will support a new level of navigation, not limited to the hierarchical categories, but also using the semantic metadata which will include geographic information, time information, entity recognition (people, organizations, …) and so forth. Finally, based on general or specific user usage patterns, MICO will enable tailored recommendations.

User stories

ID	PP	Description
US-06	IO10	As a user, I want to upload tagged videos and have tags available for all further processing steps; tags can be geographic or low-level (possibly embedded in the file), or high-level tags provided by the user.
US-07	IO10	As a user, I want to view thumbnail sequences of video segments.
US-08	IO10	As a user, I want to get info about similar video items and segments (item similarity).
US-09	IO10	As a user, I want to get info about perceptually similar video items and segments (item similarity).
US-10	IO10	As a user, I'd like to receive content recommendations from similar profiles (based on gender, age, occupation and location).

US-11	IO10	As a user, I'd like to receive content recommendations based on most relevant news events in my area.
US-12	IO10	As a user, I want to identify persons in videos, and receive related information about them.
US-13	IO10	As a user, I want to identify products in videos, and receive related information about them.
US-14	IO10	As a user, I want to identify broadcaster logos in videos, and receive related information about them.
US-15	IO10	As a user, I want to identify brands in videos, and receive related information about them.
US-16	IO10	As a user, I'd like to take a picture or upload an image and be recommended with videos related to that picture or image.
US-17	IO10	As a user, I'd like to walk around, access content using a geographical map and be informed if there are videos covering the area (ie a neighborhood), monuments or any other relevant point of interest.
US-18	IO10	As a user, I'd like to search content items using automatic speech recognition.
US-19	IO10	As an admin, I want to see, validate and edit segmentation of videos.
US-20	IO10	As an admin, I want to annotate news video segments and extract a topic for each segment.
US-21	IO10	As an admin, I want to filter and rank videos based on A/V quality aspects.
US-22	IO10	As an admin, I want to get a quick assessment of whether audio material has been edited before the upload.
US-23	IO10	As an admin, I want to pre-filter nudity content.
US-24	IO10	As an admin, I want to pre-filter content using search based on low-level asset information as audio/video technical parameters, format, duration, quality, resolution, color depth, and so forth.

US-25	IO10	As an admin, I want the ingested contents to be automatically validated for copyright infringement.
US-26	IO10	As an admin, I want to pre-filter content using search on high-level asset information such as metadata and tags either recognized by the system of provided by the user.
US-60	IO10	As an admin, I want to see a summary of articles and be able to reuse them.

Datasets required and available

Content may include the following:

- content generated by the users using their mobile handsets, this content may include embedded metadata such as the geographical coordinates, or other low-level information,
- UGC content posted on YouTube and accessible by means of news feeds (for more details see the following tables),
- video content and related metadata available from institutional web sites (metadata may include the following: Title, Summary, Source, Aspect Ratio, Date, People, Categories/Tags).

Name	Description
ID	DS-12
Partner/Person	IO10/ David Riccitelli, Andrea Volpini
Description	User generated content coming from “Shoof”
Formats	videos in MP4 or 3GP format; images: PNG, JPG, GIF

Name	Description
ID	DS-13
Partner/Person	IO10/ David Riccitelli, Andrea Volpini
Description	AP news video content (http://www.aparchive.com/) along with the following optional metadata:

	• Title • Summary • Story No • Source • Aspect Ratio • Date • People • Subscription • Shortlist extract • Categories/Tags
Formats	videos in MP4 or 3GP format; images: PNG, JPG, GIF • Resolutions: HD, SD • Statuses: Not Digitized, Digitized • Colors: Color, Black & White • Aspect Ratios: 16:9, 4:3

Name	Description
ID	DS-14
Partner/Person	IO10/ David Riccitelli, Andrea Volpini
Description	YouTube video content (http://www.youtube.com/) for video segment identification and content exploration
Formats	Formats: H.264, WebM formats. Metadata: available using YouTube Data API v3 (https://developers.google.com/youtube/v3/) Resolutions: 240p, 360p, 480p, 720p, 1080p

Technology Enablers

TE#	TE application	related US
TE-202	Object detection: Product/brand detection and recognition; logo recognition	US-13, US-14, US-15
TE-204	Face / person recognition	US-12
TE-205	A/V Media Quality	US-21

TE-206	Temporal video segmentation, for easier navigation, segment annotation, keyframe extraction	US-19, US-07, US-20
TE-207	Speech-Music Discrimination	US-18
TE-211	Segment-based visual similarity / matching	US-08, US-09
TE-214	Speech-to-text transcription	US-18
TE-221	Text summarization	US-60
TE-224	Audio cutting detection	US-22
TE-225	Robust audio identification based on audio fingerprint extraction and matching; also needs access to a respective database	US-25
TE-226	Nudity detection	US-23
TE-227	Media container tag extraction	US-06
TE-401	Support spatial-temporal fragmentation selection and aggregation functions	US-19, US-20
TE-402	Support Fulltext Search	US-18, US-24, US-26
TE-407	Find recommendations for similar videos based on a stored or new video/image (not based on low-level information)	US-10, US-11, US-16
TE-410	Present thumbnails for video sequences	US-07
TE-413	Support Geographic properties and functions (e.g. nearby)	US-17
TE-503	User Similarity calculator; determine the similarity of users based on their activities on the sites	US-10

TE-506	Cross-modal content recommender; determines which content should be delivered to a specific user	US-08, US-09, US-10, US-11
TE-501	User activity and context monitor	US-08, US-09
TE-507	Item similarity calculator; determines the similarity of media items	US-08, US-09

Cross-Cutting user stories and technology enablers

Responsible Partner / Author: FHG / Patrick Aichroth

Apart from the user stories and derived use cases, representing mostly functional requirements, **non-functional requirements** will also play an important role. Aspects with respect to usability, testability, performance, scalability, security, privacy, portability, interoperability, maintainability, modifiability, integrability, extensibility will be checked in the further process.

For instance, this will include security aspects including authentication of metadata and content, or access control requirements when dealing with licensed content that requires such handling in the InsideOut10 showcases. It will also take into consideration the privacy aspects when dealing with user and usage-related information. Both security and privacy aspects will be dealt with in a “by-design” approach, tackling them also during system design, and all following project phases.

Beyond that, some cross-cutting user stories (CUS) were identified, which are listed with the respective technology enablers in the following:

Knowledge Enhancement

ID	PP	Description
CUS-1	UP	As a user, I want further information about the content that i see/hear in my multimedia content: Concepts that are contained in multimedia items can be connected to the Linked Open Data in order to retrieve further information
CUS-2	UP	As a user, I want further information for core concepts of given comments: By disambiguating core vocabulary of a comment or post, these can be connected to the Linked Open Data for further information

CUS-4	SRFG / UP	As a user, I want to search and browse for Multimedia Items and Fragments using textual description (fulltext search) as well as structured information. I want to use spatial as well as temporal predicates and functions (after, leftBeside, firstShot, etc.).
CUS-5	SRFG	As a user, I want to get media content that is similar to a given media item.
CUS-7	IO10	As an admin, I want the ingested contents to be classified using user-generated vocabularies (custom schemas) or existing linked data classification / vocabularies (i.e. IPTC)

Visualisation

ID	PP	Description
CUS-3	UP	As a user, if there is further information (see CUS-1 and CUS-2) about concepts of multimedia content or comments, I want to browse and consume them in a graphical user interface
CUS-6	SRFG	As a user, I want to have a proper query UI. It should support me in writing queries (that include features from CUS-4 and CUS-5). It also should provide me a proper result visualization that may be browsable (as a kind of query by example).

Technology Enablers

TE#	Specifics of TE application	Applies to
TE-211	image similarity	CUS-5
TE-301	Disambiguation from surface forms to Semantic Web URIs	CUS-1, CUS-2, CUS-3
TE-302	Linked Data endpoint discovery; when looking up a semantic concept, the query can be spread to different available endpoints in order to achieve enriched results	CUS-1, CUS-2, CUS-3, CUS-7
TE-401	Support spatio-temporal queries	CUS-4

TE-402	Support Fulltext Search	CUS-4
TE-403	Retrieval of structured data	CUS-4, CUS-7
TE-404	RDF visualisation and information enhancement; this visualises semantic concepts in combination with their corresponding information in a clickable modern tile design	CUS-3
TE-404	Query Builder UI Result Representation Result Format Browsable Result Sets	CUS-6
TE-406	Support boolean retrieval	CUS-4
TE-407	Similarity Search via RDF Graph similarity for graph snippets (that describe a media resource)	CUS-5, CUS-7
TE-408	Support confidence values for search results (fuzzy search)	CUS-5
TE-411	Search is agnostic about metadata vocabularies	CUS-5