



## Combined Use Cases: First Prototype

<b>Deliverable Nr Title:</b>	7.2.1 & 8.2.1
<b>Delivery Date:</b>	May 2015
<b>Author(s):</b>	Chris Lintott, Grant Miller and Alex Bowyer (University of Oxford - Zooniverse)  Marcello Colacino, Piero Savastano, David Riccitelli, and Andrea Volpini (InsideOut 10)
<b>Publication Level:</b>	Public Document

## Table of Contents

[Table of Contents](#)

[Executive Summary](#)

[A: Zooniverse First Prototype Use Case](#)

[Introduction](#)

[References](#)

[Snapshot Serengeti](#)

[Conclusions](#)

[B: InsideOut10 First Prototype Use Case Description](#)

[Introduction](#)

[Overview](#)

[User Stories / Requirements / Goals](#)

[MICO Prototype - Technology Stack Available for Testing](#)

[Functional requirements mapping to available MICO components](#)

[Conclusions](#)

## Document Context Information

<b>Project (Title/Number)</b>	MICO - "Media in Context" (610480)
<b>Work Package / Task</b>	Work Package 7 - Use Case: Crowd Sourcing Platform Work Package 8 - Use Case: Video Sharing Platform
<b>Responsible person and project partner</b>	Grant Miller (University of Oxford - Zooniverse) Andrea Volpini (InsideOut10)

## **Copyright**

This document contains material, which is the copyright of certain MICO consortium parties, and may not be reproduced or copied without permission. The commercial use of any information contained in this document may require a license from the proprietor of that information. Neither the MICO consortium as a whole, nor a certain party of the MICO consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and accepts no liability for loss or damage suffered by any person using this information.

Neither the European Commission, nor any person acting on behalf of the Commission, is responsible for any use which might be made of the information in this document.

The views expressed in this document are those of the authors and do not necessarily reflect the policies of the European Commission.

## Executive Summary

This document outlines the Use Case Prototype setups by the use case partners (Zooniverse and Insideout10) for the evaluation of the MICO platform.

### **Zooniverse [REF-ZOO] Use case**

The Zooniverse [REF-ZOO] is focused on utilizing technologies developed by MICO results within the context of citizen science platforms and the scientific research community, validating cross-media recommendation in this context.

By definition, Zooniverse projects have large amounts of media objects 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 by:

- 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

We have made good progress so far setting up the individual components (namely the user analytics collector and the experiments server) and the architecture that will allow us to implement and validate our chosen use cases. The next step will be integrating the recommendation engine with this infrastructure, so that we can run the Happy Volunteer experiment on the live Snapshot Serengeti site, and analysing the results of how that change to the interface.

### **Insideout10 First Prototype Use Case**

The aim of this document is to analyse context and requirements for integrating MICO functionalities within existing enterprise applications and to prepare for validating the effectiveness and business value of these technologies in two real-world scenarios:

- ❑ [A] a responsive news magazine website produced by Greenpeace Italy for its supporters
- ❑ [B] a UGC mobile video recording application for Android developed in Cairo, Egypt by Insideout Today (a sister company of InsideOut10)

Great progress has been made in refining all business requirements for the video news showcase, proper engagement among all stakeholders has been guaranteed for both scenarios and an end-to-end environment, traversing the different applications, has been setup for the validation of MICO.

## A: Zooniverse First Prototype Use Case

### Introduction

The Zooniverse **[REF-ZOO]** is focused on utilizing technologies developed by MICO results within the context of citizen science platforms and the scientific research community, validating cross-media recommendation in this context.

By definition, Zooniverse projects have large amounts of media objects 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 by:

- 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 - especially important in moving to larger datasets and real-time processing - and also to create a system that will stimulate higher levels of motivation and enjoyment amongst volunteers.

### References

Reference ID	Link
REF-INT	<a href="http://www.www2015.it/documents/proceedings/companion/p331.pdf">http://www.www2015.it/documents/proceedings/companion/p331.pdf</a>
REF-REQ	<a href="http://goo.gl/Vcjb8L">http://goo.gl/Vcjb8L</a>
REF-ZOO	<a href="http://www.zooniverse.org/">http://www.zooniverse.org/</a>
REF-SS	<a href="http://www.snapshotserengeti.org/">http://www.snapshotserengeti.org/</a>
REF-HAPPY	<a href="https://goo.gl/vRUaCr">https://goo.gl/vRUaCr</a>
REF-EXPERT	<a href="https://goo.gl/u1xFwA">https://goo.gl/u1xFwA</a>
REF-SPEEDY	<a href="https://goo.gl/LRpnTg">https://goo.gl/LRpnTg</a>
REF-INFORMED	<a href="https://goo.gl/Y7TZGr">https://goo.gl/Y7TZGr</a>
REF-GEORDI	<a href="https://github.com/zooniverse/geordi">https://github.com/zooniverse/geordi</a>
REF-ZOOEXP	<a href="https://github.com/zooniverse/ZooniverseExperimentServer">https://github.com/zooniverse/ZooniverseExperimentServer</a>

REF-FB

<https://facebook.github.io/planout/>

## Snapshot Serengeti

MICO and the Zooniverse have decided to focus on Snapshot Serengeti **[REF-SS]** from the list of various potential showcases outlined in the earlier requirements gathering stage.

The task on the project involves identifying various animals, and their behaviour, from camera trap images obtained in the Serengeti National Park, with the goal of helping researchers better understand how the numerous species interact with each other. Volunteers have to identify the animals from a list of 51 species, either directly or by using a set of filters to identify less familiar species, and the interface also lets them provide information on their numbers and activities.

A number of different potential use cases relating to Snapshot Serengeti were identified:

1. Happy Volunteer **[REF-HAPPY]** - where we experiment with showing users popular or interesting images to maintain their participation
2. Expert Volunteer **[REF-EXPERT]** - where we present experts with more difficult images
3. Speedy Volunteer **[REF-SPEEDY]** - where we avoid showing a user too many complex images back to back
4. Informed Volunteer **[REF-INFORMED]** - where we educate a user by showing them training content
5. Aware Volunteer - where volunteers are given visibility of the classifications assigned by others after performing a classification
6. Assessed Volunteer - where the volunteer is shown after classification how much + / - their classification differs from consensus
7. Stimulated Volunteer - where the volunteer is guaranteed to be shown a wider variety of animals than random selection would give.
8. Profiled Volunteer - where the user's performance over gold standard data is used to group them into a user type, each of which is treated differently.

## Happy Volunteer Use Case

In the Happy Volunteer use case we plan to deliver a higher frequency of images that a particular volunteer might find interesting. In order to do this we need to collect information on what type of images a particular volunteer may be more interested in, e.g. if they favourite and collect a lot of images of zebras, we would infer that they are interested in zebras and therefore we would deliver that volunteer images of zebras at a higher rate than the current random deliver set-up.

The goal of this experiment is to gauge whether this approach leads to the volunteer spending more time on the classification interface / performs more classifications, when compared to a control group as this would suggest they are having a more enjoyable time as a direct result of our experimental interventions in subject selection.

## Infrastructure Readiness & Experimental Approach

Zooniverse have successfully used our analytics collector, Geordi [see 2.2.1 below] to track volunteer events, save the information in an SQL database, for all visitors to Snapshot Serengeti since February 2015. This will provide the data to determine user preferences in the Happy User Experiment. Zooniverse have also implemented the experimental framework in the Snapshot Serengeti web application. We have identified a set of ~20,000 previously unclassified images from the Serengeti project which can be used for the first experiments. They are drawn from across the eight seasons of data which have so far been classified on the site, and thus are representative of the full range of conditions in which we expect the experiment to proceed.

The volunteers will be split into separate cohorts. Those who are in the experiment, and therefore receiving a higher frequency of images than their user profiles might suggest; and those who are in the control group receiving images in a random fashion as normal. We will then analyse the general behaviour of these two cohorts to see if the experiment is having any effect on volunteer behaviour on the site.

## Preliminary Findings

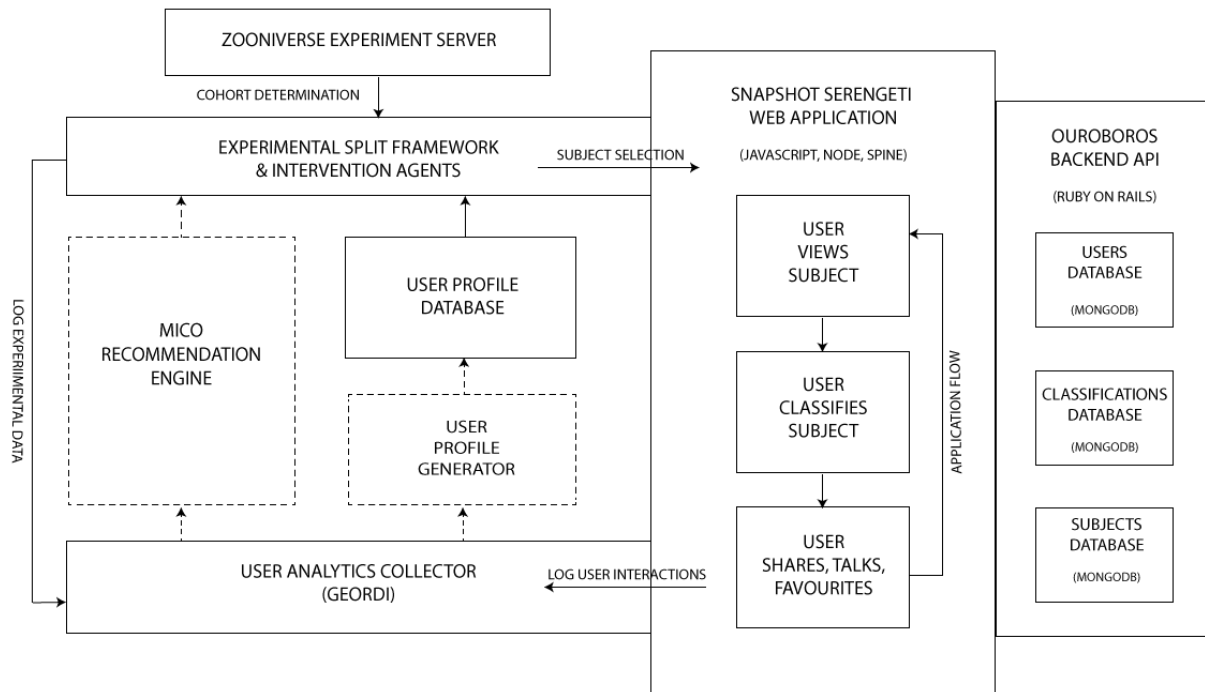
Prior to the completion of the recommendation engine, a test of the Happy Volunteer experiment was carried out with known subjects, using custom scripts and algorithms in its place to generate user profiles from analytics data and select next subjects for users. The findings so far show, contrary to expectations, that insertion of “liked” animals has a detrimental effect on the number of subjects classified and on the length of the user’s session. Analysis suggests that the reason for this may be that the experimental cohort received fewer “empty” subjects in their experimental data set (given that the subset of inserted images were definitely NOT blank), and that an ideal of around 80% blank subjects is optimal to enable longer sessions and greater numbers of classifications. This has the following implications:

- Future experiments must very carefully control not just the interestingness of the animals a user is presented with, but the ratio of “empty” images to “contains animal(s)” images.
- It may not be possible to recommend a single subject to a user. The recommender will need to recommend sets of subjects rather than treating them singularly.
- The work to identify “empty” images is now even more important; originally it was thought that “empty” images need to be filtered out, however latest findings suggest

that in fact we must know which images are likely to be empty so that we can better control the makeup of the user's subject set and not deter them from participation. As of yet, the reasons for the importance of empty images is not known. It may be that a disparity between background emptiness and unusual animal images heightens the psychological reward of finding and classifying an animal image.

## Architecture

Using the Happy Volunteer use case as a starting point, we designed an architecture that allows us to collect various pieces of volunteer information (regarding their behaviour on the site) and set up experiments to validate the use cases. The two main components we have built for this are the User Analytics Collector (a.k.a Geordi) and the Zooniverse Experiments Server. In the future, the components will serve as the interface between the Zooniverse and the MICO architecture, specifically using recommendation components from WP5 and extraction components from WP2.





## User Analytics Collector (Geordi)

### [REF-GEORDI]

Geordi is an analytics database server, which receives user interaction events from the web application via its REST API, which is based upon the Loopback/Node.js platform, which uses MySQL for storage. We wanted something quite simple for the implementation, however an initial prototype built using JSON storage and the FortuneJS/Node library was found to be unable to scale and underdeveloped. Loopback is a proven open source platform with a company backing and supporting it, and MySQL is well known to be able to handle the number of records and queries we need to deal with.

Geordi records the time, userID, subjectID, event type, experimental cohort and other related information in a MySQL database every time the user triggers a key event such as hitting a favorite button or identifying an animal. As well as user interaction events, it can also be used to log errors and to track experimental results.

Different events can be used as indicators of positive or negative interest in a subject:

STRONG POSITIVE	MILD POSITIVE	NEUTRAL	MILD NEGATIVE	STRONG NEGATIVE
FAVORITE	TALK	IDENTIFY	TUTORIAL	LEAVE
PIN	MAP	NOTHING	FILTERS	LOGOUT
TWEET	VIEW	FIRE, YOUNG		
FACEBOOK	FAVORITE THEN UNFAVORITE	PLAY / FRAMES		
(COLLECT)	(POST)	LOGIN		
	(COMMENT)	CLASSIFY		

Most of these events are fairly self explanatory. Some additional details:

- “Leave” occurs when the user closes the tab or browser, or otherwise leaves the site
- “Filters” are a set of events when the user uses filters to try and narrow down a species selection, for example “has horns” or “is brown in colour”
- “Frames” is when the user clicks to view individual images (each subject consists of three images taken in short succession)

- “View” occurs when the user is first presented with a subject.
- “Identify” occurs when the user identifies a species within a subject (this could happen more than once per subject)
- “Classify” occurs when the user is finished classifying this subject
- “Map” means they viewed the map to see where the images were taken
- “Young” means they marked that young animals are present
- “(Collect)” (not yet implemented) means the user has added the subject to a collection

## Zooniverse Experiments Server

### [REF-ZOOEXP]

The experiment server is a Ruby-based REST API running on Facebook's PlanOut architecture [REF-FB], to be used by the experiment framework within the Snapshot Serengeti web app. We chose PlanOut because it offers a simple, lightweight way to assign users to cohorts and track experimental details, it also allows us to maintain several different experiments running in parallel if we want to do that.

The experiment server keeps track of a number of different Zooniverse Experiments of which the Happy Volunteer Experiment is one. Its primary purpose is to assign each user to a cohort - either the control group or the "interesting animals" group. For the Happy Volunteer Experiment, it also keeps track of each user's participation in the experiment, including which subjects they will be shown next, and which they have already seen.

## User Stories / Requirements / Goals

Many user stories were created and detailed in the original Compendium Use Case Requirements Analysis Deliverable [REF-REQ]. Only a subset of these are actively being pursued. The status of each user story is detailed in the following table. US-60 is a new user story not originally documented in the Compendium.

User stories	Goal	Status
<b>US-27</b>	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’	On hold, awaiting future WP2 development.
<b>US-28</b>	As a Zooniverse admin, I would like to be able to detect when a scientist should be prompted to look at a subject,	On hold, awaiting future

	based on annotations and information from 'Talk' comments	WP2 development.
<b>US-29</b>	As a Zooniverse admin, I would like to identify volunteer types	On hold, awaiting future WP2 development.
<b>US-31</b>	As a Zooniverse admin I would like to be able to automatically detect Snapshot Serengeti images with no classifiable animals in them	<b>ACTIVE</b>
<b>US-32</b>	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	Will still be useful; no current plans
<b>US-33</b>	As a Zooniverse admin I would like to be able to perform automatic animal species pre-classification in Snapshot Serengeti (48 species)	<b>ACTIVE</b>
<b>US-34</b>	As a Zooniverse admin I would like to be able to perform automatic animal attribute pre-classification in Snapshot Serengeti	<b>ACTIVE</b>
<b>US-35</b>	As a Zooniverse admin I would like to be able to perform automatic animal number detection in Snapshot Serengeti	<b>ACTIVE</b>
<b>US-49, US-50</b>	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 - and whether I should interrupt a volunteer with text, an image, or a video	Being worked on within Zooniverse team and as part of a separate research project. [REF-INT]
<b>US-51, US-52, US-53</b>	As a Zooniverse admin, I'd like to know when I should educate a volunteer, and whether I should educate that volunteer with text, an image, or a video, and which piece of education I should give to that volunteer	Will still be useful; no current plans
<b>US-54</b>	As a Zooniverse admin, I'd like to know when a volunteer has made an interesting comment on a subject	On hold, awaiting future

		WP2 development.
<b>US-55</b>	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	On hold, awaiting future WP2 development.
<b>US-56</b>	As a Zooniverse admin, I'd like to know when Zoonibot should give an explanation	On hold, awaiting future WP2 development.
<b>US-57</b>	As a Zooniverse admin, I'd like to know what Zoonibot should say to a volunteer	On hold, awaiting future WP2 development.
<b>US-58</b>	As a Zooniverse admin, I'd like to be able to group subjects (i.e. images, videos or audio files) by similarity	On hold, awaiting future WP2 development. May also build upon the work of the recommendation engine and feature extractors.
<b>US-59</b>	As a Zooniverse admin, I'd like to be able to recommend different projects to volunteers based on their previous experiences	Will still be useful; no current plans
<b>US-30, US-36, US-37, US-38, US-39, US-40, US-41, US-42, US-43, US-44, US-45, US-46,</b>	User stories relating to other Zooniverse projects: Galaxy Zoo, Plankton Portal, Worm Watch Lab, Crisis Response, Asteroid Zoo, Whale FM	These user stories have all been relegated to "Future ideas" - for the immediate future we are focussing on Snapshot Serengeti and general

<b>US-47, US-48</b>		Zooniverse use cases.
<b>US-60</b>	As a Zooniverse admin, I'd like to be given recommended next subjects for a Snapshot Serengeti user, based upon their previous behaviour on the site	<b>ACTIVE</b>

## MICO Prototype - Technology Stack Readiness for Testing

Here is a list of all related technology extractors, along with their priority. “Expected” indicates an essential component for testing of current active user stories (which are marked in **bold**), which is expected to be available for testing soon. “Ready” indicates such a component which is already known to be ready for testing. “Plan” means that the TE is not ready yet, but it is expected that this will be added to short term plans, if not already planned - and that when the specified user stories are implemented, this component will be required. This also includes items where work is in progress. “Optional” means the component has not yet been built but even when the specified user story is implemented, this component may be optional. “N/A” means there is no immediate or expected need to build this, and it is not available.

ID	Name	User Story	Description	Readiness
<b>TE-201</b>	Feature extraction	US-27, <b>US-31</b> , <b>US-35</b> , US-58	Low-level feature extraction for RoI detection	Plan
<b>TE-202</b>	Automatic detection of empty images	<b>US-31</b> , <b>US-33</b> , <b>US-35</b>	Automatic detection of image with no classifiable animals in it (Semi-)automatic animal detection (Semi-)automatic animal number detection	<b>Expected</b>
<b>TE-203</b>	Pre-classification of animals	<b>US-33</b> , <b>US-34</b>	Automatic animal species pre-classification Automatic animal attribute pre-classification	Plan
<b>TE-210</b>	Image series detection	US-32	Automatic image series detection for the case of timestamping malfunctions	N/A

<b>TE-211</b>	Machine learning on features	US-27, US-58	Image appeal and similarity are classified by means of machine learning or feature space distance based on extracted low level features	Plan
<b>TE-213</b>	Sentiment analysis	US-27, US-28, US-29, US-54, US-55, US-58	Determine the polarity of forum entries, e.g. positive, negative, or neutral	Plan
<b>TE-215</b>	Features from text	US-58	Derive features from text fields to be used in cross-media classification	Plan
<b>TE-216</b>	Text cleaner	US-27, US-28, US-55	Text cleaner to remove markup and standardize citation, punctuation, etc. Necessary as a preprocessing step for e.g. the phrase-structure parser	Plan
<b>TE-217</b>	Phrase structure parser	US-27, US-28, US-55	Phrase structure parser	Plan
<b>TE-218</b>	Interactive wrapper generator	US-27, US-28, US-55	Interactive wrapper generator	Plan
<b>TE-219</b>	Model semantic features	US-54	Graph operation toolkit to model semantic features	Plan
<b>TE-220</b>	Keyword extraction	US-27, US-28, US-29, US-55, US-56, US-59	Extract keywords related to e.g. species or activities	Plan
<b>TE-401</b>	Spatial media fragment	<b>US-35</b> , US-52, US-57	Support spatial media fragment e.g for counting the number of animals on query-time, support User with image snippets (e.g. that shows a specific animal for training)	Optional

<b>TE-403</b>	Regional query functions	<b>US-32, US-35, US-52</b>	Support regional query functions to identify and aggregate regional fragments e.g. return a lion right beside a gazelle, Support image metadata retrieval	Optional
<b>TE-404</b>	Metadata browsing	US-53	Allow users browsing the database for images, that shows specific scenes (e.g. a group of animals) and support them with useful metadata (e.g. what are the characteristics of this subject)	N/A
<b>TE-405</b>	GUI	US-53	Support user with graphical user interface  (Note: A good API is more useful to Zooniverse, therefore this is low priority.)	N/A
<b>TE-407</b>	User Trainer	US-53	Train the users by showing them images and metadata with a high similarity	Plan
<b>TE-411</b>	Pivot vocabularies	<b>US-33</b>	Support for pivot vocabularies (diverse datasets for animal classification)	Optional
<b>TE-501</b>	User behaviour monitor	US-29 , US-49 , US-50 , US-51 , US-52 , US-53 , US-54 , <b>US-60</b>	User activity and context monitor; collect user, usage and context information.  This (a.k.a. Geordi) has been developed in-house at Zooniverse.	<b>Ready</b>
<b>TE-502</b>	Project similarity	US-59	Project similarity calculator	Plan

<b>TE-503</b>	User similarity	US-29	User Similarity calculator; determine the similarity of volunteers based on their activities on the projects	Plan
<b>TE-504</b>	Volunteer analysis	US-29, US-49, US-50, US-51, US-52, US-53, US-56, US-57	Volunteer-type analysis; determine the characteristics of a volunteer based on their activities on the projects	Optional
<b>TE-505</b>	Subject analysis	US-57, US-58	Subject-type analysis; determine the characteristics of the subjects (images, audio, video files) based on their content	Plan
<b>TE-506</b>	Content recommender (WP5)	US-49, US-50, US-51, US-52, US-53, US-55, US-56, US-57, <b>US-60</b>	Cross-modal content recommender; determines which content should be delivered to a specific volunteer	<b>Expected</b>
<b>TE-507</b>	Item similarity	US-29, US-58	Item similarity calculator; determines the similarity of media items	Plan

## Functional requirements mapping to available MICO components

Looking now only at ACTIVE user stories and Expected/Ready TE's, we can map these to functional requirements as follows. These are the only TE's that will be addressed in the test plan at this stage.

Component	Feature Description	TE
Animal Detection	<ul style="list-style-type: none"> <li>Ability to be given a suggested animal type (e.g. horned animal) for a subject.</li> <li>Ability to be given an estimate of the number of animals present in an image.</li> </ul>	<b>TE-202</b>



	<ul style="list-style-type: none"> <li>• Ability to be given a determination/likelihood of an image being “empty”.</li> </ul>	
User Behaviour Capture	<ul style="list-style-type: none"> <li>• Ability to know which subjects a user has viewed, classified, shared, favourited, etc.</li> </ul>	<b>TE-501</b>
Next Subject Recommendation	<ul style="list-style-type: none"> <li>• Given data on user behaviour, ability to be provided with a set of recommended subjects that a user is likely to hold a favourable disposition towards</li> </ul>	<b>TE-506</b>

The other “Planned” TEs will be built and tested in a later stage of the project.

## Conclusions

We have made good progress so far setting up the individual components (namely the user analytics collector and the experiments server) and the architecture that will allow us to implement and validate our chosen use cases. The next step will be integrating the recommendation engine with this infrastructure, so that we can run the Happy Volunteer experiment on the live Snapshot Serengeti site, and analysing the results of how that change to the interface. We still have work to do to connect the relevant MICO components to this architecture, however we have been able to perform experiments and gain new understandings of the impact of subject selection even before the completion of the recommendation engine.

In addition to that, we will look towards technical design and experimental design for the experiments that make use of the other extractors, particularly those around empty image detection, animal pre-classification, and those that result from the use of the semantic/sentiment analysis components.

## B: InsideOut10 First Prototype Use Case Description

### Introduction

InsideOut10 is an Italian startup and consulting firm with an in-depth experience on web publishing and media delivery platforms. This document is written in the context of the activities of WP8 of the MICO project and investigates on the challenges in the creation of a compelling user experience for news and content rich websites.

The document is intended for anyone interested on the future of news and innovation including but not limited to journalists, front-end developers, technologists, startups and marketing experts.

The pervasiveness of online communication, the massive shift towards mobile devices and the ever-changing landscape of news production and consumption patterns requires a lean approach in the creation of news outlets.

The massive amount of content being produced inside and outside the newsroom needs to be properly organised and curated to meet the evolving demands of the readers. The current status quo where content is locked in different platform (silos) shall evolve to make content offerings seamlessly accessible across different channels. The cross-media analysis, querying and recommendation functionalities provided by MICO can play a crucial role for both readers and content creators. Integrating existing publishing workflows and applications with MICO technology models extend readers dwell time with repurposing matching content and reduce the complexity of content management operations by 'unveiling' the hidden semantics of raw multimedia content. In other words MICO can help reducing the time spent by online editors for bolstering their media contents by creating a context, detecting quality issues for online videos and supporting the interlinking between different media assets whether in textual or visual form.

User-generated content ("UGC"), also known as 'citizen journalism' or 'participatory media', is an emerging content form that is gradually entering the newsrooms due to developing technologies like mobile audio-video streaming and recording that are now widely available to a great number of people. As content grows in volume and billions of videos are published on the Internet, understanding these contents, providing helpful information for each video and organising it in meaningful ways is an enormous challenge.

MICO can help to automatically derive useful information from UGC contents and support developers, journalists and news publishers learn more and do more with UGC when creating innovative news products and services.

## Overview

The aim of this document is to analyse context and requirements for integrating MICO functionalities within existing enterprise applications and to prepare for validating the effectiveness and business value of these technologies in two real-world scenarios:

- ❑ [A] a responsive news magazine website produced by Greenpeace Italy for its supporters
- ❑ [B] a UGC mobile video recording application for Android developed in Cairo, Egypt by Insideout Today (a sister company of InsideOut10)

The challenges addressed by the Video News Showcase in MICO WP8 are shared by content creators, journalists and news publishers worldwide and can be summarised as follows:

- Independent news agencies struggle to find their audience on the web
- The shift to mobile requires a “mobile first approach”<sup>1</sup>
- Video is key for sustaining organic growth and advertising revenues for media outlets
- Creating premium and attention-grabbing news contents is expensive and extremely time-consuming for editors
- Metrics are changing as advertisers seek engagement and time rather than just clicks and impressions<sup>2</sup>
- Next generation UGC enter the newsroom but they are hard to manage for startups in the news sector, broadcasters and independent content providers<sup>3</sup>
- UGC is 50% more trusted than other media by Millennials<sup>4</sup>

For both scenarios we intend to implement in WP8 a semantic workflow within our existing technology modules to add MICO and improve content creation, content management and content delivery. The diagram below shows the various steps involved in the implementation of both scenarios from content acquisition to delivery.

---

<sup>1</sup> “[Only Digital Media Sees Growth in daily consumption](#)” by eMarketer

<sup>2</sup> “[What is the Attention Web?](#)” by Chartbeat

<sup>3</sup> “[CNN's Futurist Sees Shift On UGC, wearable](#)” by Netnewscheck

<sup>4</sup> “[IPSOS Media Research](#)” by Crowdtap

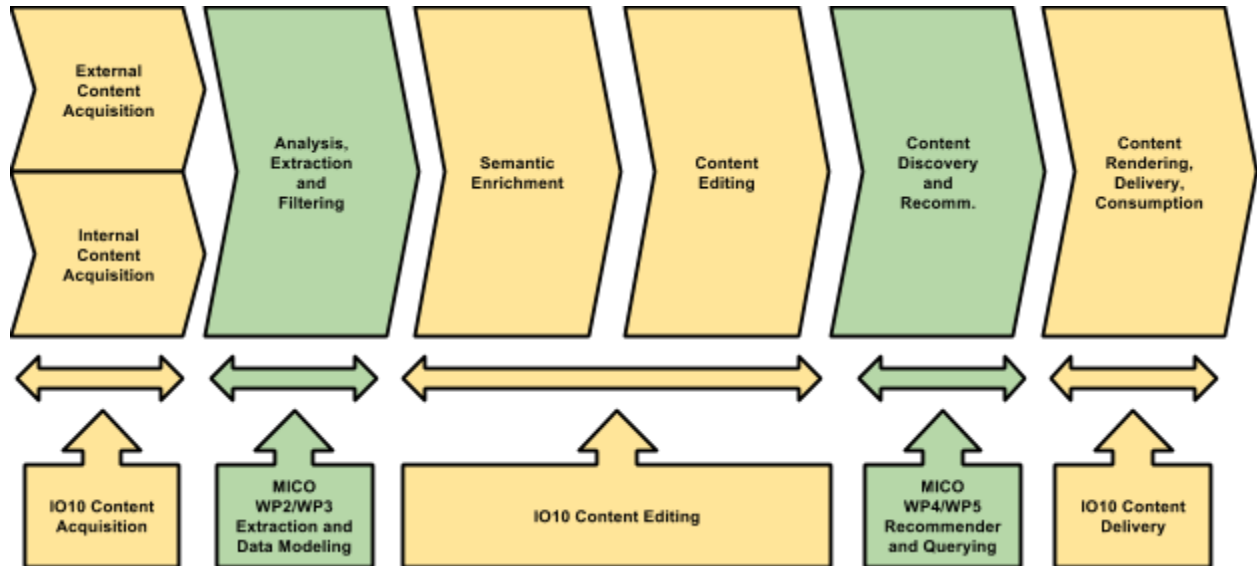


Figure 1: Video News Use Case Semantic Workflow

This paragraph continues with the description of the stakeholders and their problem statements for both Greenpeace Italy and Shoof the UGC video recording application. The listed stakeholders are directly or indirectly participating in the validation of MICO. The paragraph also introduces all software modules involved in the setup of WP8, their value proposition, the existing list of features and the expected integration with MICO.

## Stakeholders

Stakeholders behind each scenario are key to drive the development of MICO and to properly evaluate its technologies against concrete business needs.

### Greenpeace Italy

Name	Description	Problem statement and objectives
Supporters	<p>Already profiled by GP in various ways they are eager to protect and conserve the environment and to promote peace. They expect to be engaged by GP with compelling contents and stimulating activities.</p> <p>They represent the most strategic asset of the organisation (as main income source).</p>	<p>Inspired by GP mission they want to learn more about the organisation and its activities.</p> <p>They expect a 'premium' content experience as paying supporters of GP.</p>

Prospects	<p>A diversified audience with some degree of interest along particularly environmentally conscious lines. Can be divided in:</p> <ul style="list-style-type: none"> <li>- <i>Behavioral Greens</i> that think and act green</li> <li>- <i>Think Greens</i> that think green but not necessarily act green</li> <li>- <i>Potential Greens</i> that remain ambivalent but have the potential to convert</li> </ul>	<p>Interested in GP and/or a in specific topic proposed by the organisation.</p> <p>They might eventually turn into active supporters.</p>
Internal	<p>The main internal stakeholders are the <i>Communication</i> and <i>Fundraising Department</i>.</p> <p>The magazine is a strategic asset for the <i>Fundraising Department</i> that considers it as retention / supporter relationship management tool. The <i>Digital Unit</i> (working in the Communication department) sees the new digital version as a great opportunity to test innovation on digital content publishing offering a more engaging content discovery experience to the users; for the Communication Department which still consider the magazine as a de facto house organ.</p>	<p>Struggle to increase:</p> <ul style="list-style-type: none"> <li>- the number of donors willing to increase their donation quota above the average result</li> <li>- the engagement of readers who decide to make a new one-off donation</li> <li>- the loyalty of donors (reducing the attrition rate within 12, 18 and 24 month below the average results)</li> <li>- the number of prospects gathered through the organic traffic (no advertising costs)</li> </ul>
Civic society and open data community	<p>Active citizens, civic hackers, institutions, government, organisations and academic willing to solve the environmental challenges and limit the impact on future generations.</p>	<p>They want to be able to:</p> <ul style="list-style-type: none"> <li>- “think” themselves using any data provided by Greenpeace</li> <li>- Making data accessible and explore the assumptions of the organisation<sup>5</sup></li> </ul>

<sup>5</sup> [Explorable Explanations by Bret Victor on encouraging active reading](#)

## Shoof

Name	Description	Problem statement and objectives
Journalist	<p>They work in independent news organisations and are very active on Social Networks.</p> <p>Their need is to spot the event in the first place and create a compelling news article in a limited amount of time.</p>	<p>Main issues are:</p> <ul style="list-style-type: none"> <li>- quickly scanning fresh new contents that might be relevant to a specific event</li> <li>- creating “stories” by using fragments from multiple videos</li> <li>- crafting news using UGC contents</li> <li>- getting help when composing their articles (help = less time spent in searching for contextual information and support in properly interlinking existing web resources on the same topic)</li> </ul>
Citizen (content producers and content consumers)	<p>The target lives in Cairo. It is very active on Social Networks and willing to contribute in terms of content creation. Some are already involved in the news making process as field journalists (content producers).</p> <p>Others are constantly seeking fresh updates and social engagement opportunities (video consumer) but not necessarily are willing to create their own content.</p>	<p>Creators look for:</p> <ul style="list-style-type: none"> <li>- visibility both on and off-line</li> <li>- positive impact within their peers</li> <li>- new engagement opportunities</li> </ul> <p>Consumer are interested in:</p> <ul style="list-style-type: none"> <li>- short form entertainment</li> <li>- compelling local content</li> <li>- personalised and fun user experience</li> <li>- being always first in sharing valuable content</li> </ul>
Internal	Represented by the management of Insideout Today.	Struggle to:

	Startup-minded people deeply engaged with social media and willing to create an innovative app for local communities.	<ul style="list-style-type: none"> <li>- develop an innovative application with a small team of resources</li> <li>- organise the editorial team behind the application (the problem of receiving too many contents and not being able to properly filter is significant)</li> <li>- implement a compelling business model with the help of telcos and advertisers</li> <li>- respect the law and the privacy of the users</li> </ul>
Hosting partner or Telco	Large businesses with millions of clients and an offering of communication services over multiple devices.	<p>They look for:</p> <ul style="list-style-type: none"> <li>- innovative business opportunities to grow their profits while re-using their network</li> <li>- a new advertising inventory</li> <li>- applications that fully comply with their existing terms as well as national regulations</li> </ul>

## Components

### HelixWare

HelixWare is an Online Video Hosting Platform designed for telcos, internet service providers, enterprises, news and media publishers and developers.

HelixWare runs in the cloud: with an easy-to-use video management UI users can upload their videos and deliver a best in-class multi-screen video experience. The platform has a full set of APIs to quickly integrate video within existing publishing workflow. HelixWare also features a WordPress plugin for integrating online videos with the world famous open source CMS.

### Features provided by HelixWare

ID	Description
1	VOD Encoding
2	Multi-screen live streaming
3	Video analytics
4	Near live video channels and playlists
5	Player customisation
6	Developer APIs
7	Security, geo blocking, IP blocking and access control
8	A WordPress plugin for media upload, player customisation, video seo and video embedding

HelixWare existing client base include large organisations like A1 Telekom Austria, TotalErg and FastTelco that use it in production environments with many users. HelixWare is also been used by startups like Insideout Today as backbone video service for creating Shoof (the video recording application described below).

### WordLift

WordLift is an extension of WordPress to help writing, organising, tagging and sharing content online. It is designed for bloggers, journalists and content creators to inspire and make writing more fun.

WordLift adds semantic annotations and combines information publicly available as linked open data to support the editorial workflow by suggesting relevant information, images and links.

WordLift analyses articles using Named Entity Recognition (NER). Entities may belong to different vocabulary sets including but not limited to DBpedia, GeoNames and Freebase. WordLift also provides UIs for creating and curating custom vocabularies.

While annotating contents editors can identify the basic '*who, what, when and where*' of an article and structure information around it by creating new entities in their custom vocabularies.

Named entities are stored in the local WordPress database as well as in an optimized triple store in the cloud running Apache Marmotta. Annotation and entities are accessible via a Web



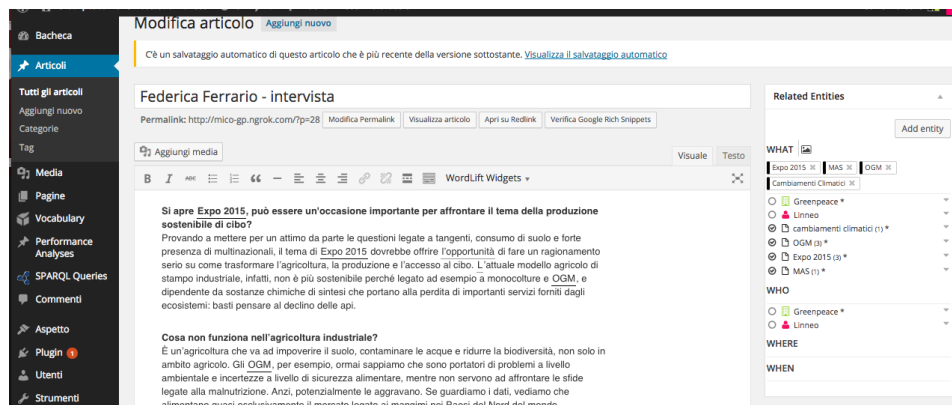
Page and also using RDF, N3 and JSON-LD formats. The triplestore in the cloud can be queried via SPARQL.

In the content creation workflow WordLift brings to content authors (producers):

- support for self-organising contents using publicly (or privately) available knowledge graphs (linked open data);
- support for creating news content with fact-based information that are contextually relevant to the article being written;
- valued relevant and free to use photos and illustrations from the Commons community ranging from maps to astronomical imagery to photographs, artworks and more;
- insightful visualisations to engage the reader;
- new means to drive business growth with meaningful navigation systems and innovating content discovery path;
- content tagging for better SEO

WordLift brings to readers (content consumers):

- multiple means of searching and accessing editorial contents around a specific event or topic otherwise spread in separate content silos;
- increased accessibility for readers with limited domain understanding;
- an intuitive overview of the all content being written on the site around a specific topic or graph of topics;
- meaningful content discovery paths.



*Figure 2 - WordLift provides content annotation, tagging and enrichment integrated in the Wordpress content publishing workflow.*

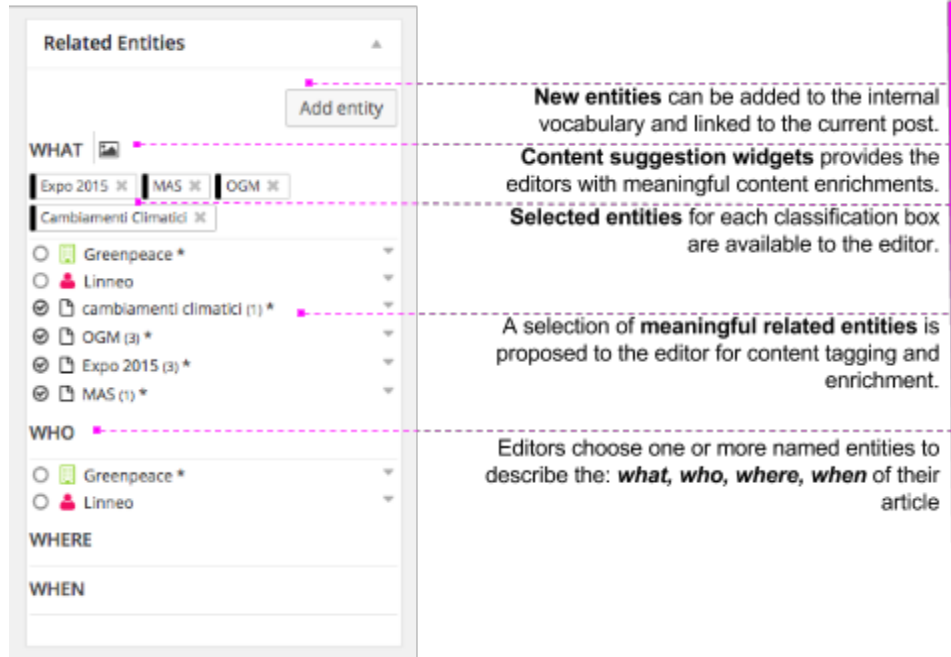


Figure 3 - WordLift Edit Post Widget, "top down" mode on: there are no text annotations selected. Entity tiles are referred to the whole post content.

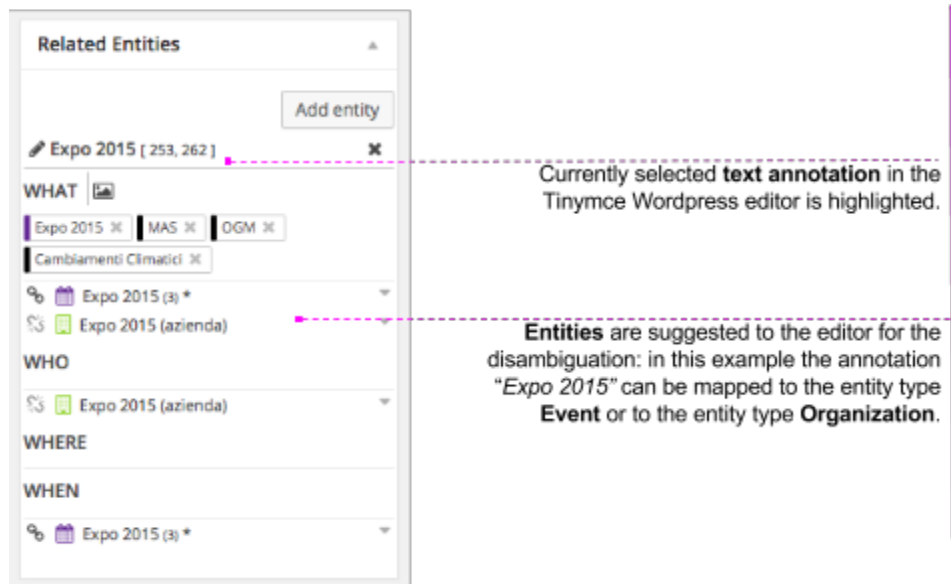


Figure 4 - WordLift Edit Post Widget, "bottom up" mode on: the text annotation "Expo 2015" is selected in the text editor. Entity tiles are referred to the current annotation

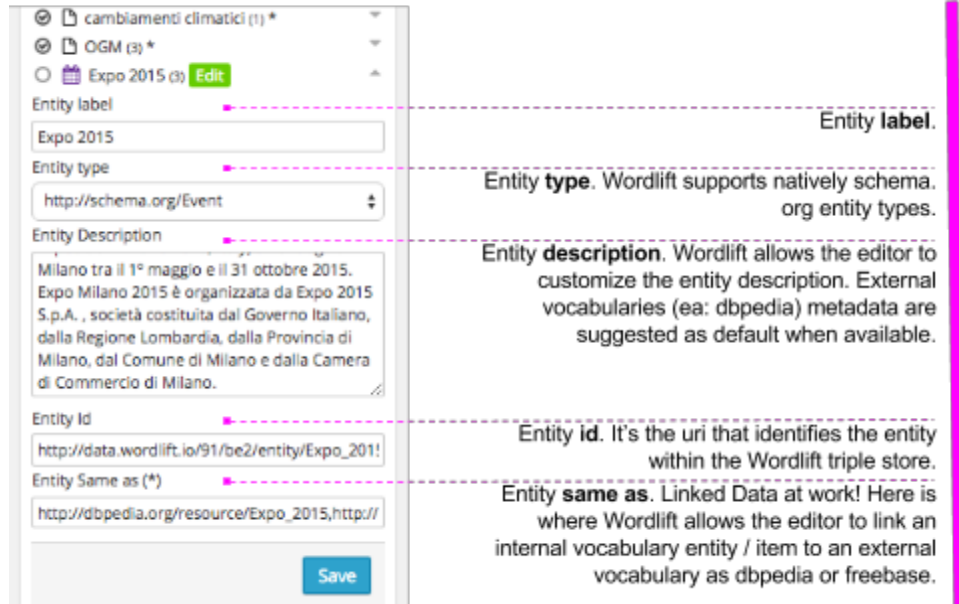


Figure 5 - Entity tile: Entity tile: how WordLift allows entity metadata contextual editing working as a linked open data engine

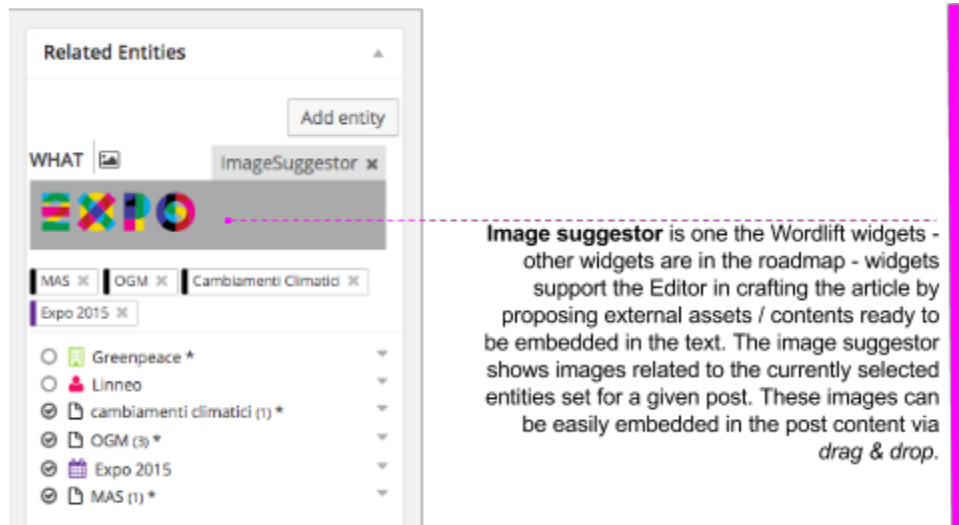


Figure 6 - Entity tile: how WordLift supports content publishing workflow

### Features provided by WordLift

ID	Description
1	Dynamic Semantic Publishing
2	Creation and management of internal vocabularies using linked open data standards
3	Easy to use UI for semantic content annotation, tagging and enrichment
4	Triplestore linked open data publishing and querying

5	Schema.org markup support
6	Integrated suggestions for contextual images
7	Content discovery and interlinking
8	Seo friendly 'entity' pages with faceted content browser

WordLift version 3 is currently in closed-beta and it is being tested by a selected numbers of organisations such as Greenpeace Italy.

### WordLift Rendering Engine

The WordLift Rendering Engine is a WordLift submodule - not yet included in the 3.0 beta WordLift release - usable both on Wordpress backend and frontends in order to enhance the UI by adding dedicated content blocks called containers.

The WordLift Rendering Engine moves from the “Container Model” IA approach as defined by Konstantin Weiss<sup>6</sup> and recently tested by relevant content publisher as The Guardian. It also takes inspiration from W3C web component<sup>7</sup> / polymer based web development methodology. The WL Rendering engine allows to manage the page context by combining independent modules known as containers. Each container brings to the user a combination of information and interaction patterns. Contextual information such as the description of a named entity from DBpedia can be placed into an independent, encapsulated container (we could have a geographical map if the entity is a place or an image/logo if the entity is an organisation). Containers are stacked together to compose web pages.

A **container** is a full-width **stackable**, **reusable**, **self-consistent** set of contents. Containers can be used for content federation: each container can be displayed on any site, in any stack. The source can come from a different site than the displayed stack's site. A container is defined by:

- an **origin** - a public uri that identifies the container;
- a **skin** - it's the container template;
- a **structure** - container data and interaction possibilities, in a machine-readable way.

The WordLift Rendering Engine is built on an AngularJs application made up the following components:

<sup>6</sup> [Why Architecting Information with Containers by Konstantin Weiss](#)

<sup>7</sup> [W3C: Introduction to Web Components](#)

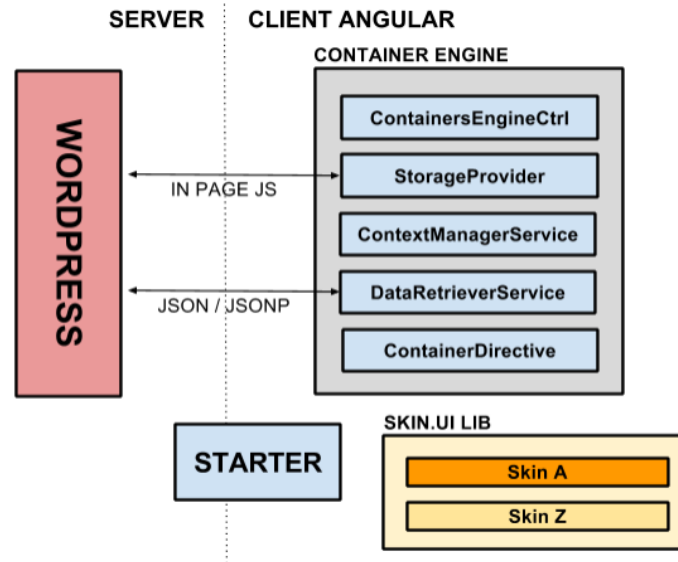


Figure 7 - Entity tile: WordLift rendering engine architecture

- **ContainersEngineCtrl** - Manages the page stack. The page stack is loaded on bootstrap and updated on context changes. Allows interactions between containers and the ContextManagerService.
- **StorageProvider** - Its main responsibility is to inject wordlift.containers local storage within angular app during the app bootstrap;
- **ContextManagerService** - Allows to add and store new properties for the current user; to add and store current user interactions on rendered contents; to track user interactions; to rewrite container origins considering both container listeners and the current page context
- **DataRetrieverService** - Retrieves data both from local storage or remote origins (JSONP is supported to ensure cross-domain communication) and prepare them for rendering; performs client side caching;
- **ContainerDirective** - wraps and communicates with the ui skin components.

#### Features provided by WordLift Rendering Engine

ID	Description
1	Agnostic support for container based UI in Wordpress
2	Extensible skin libraries
3	Page context management
4	Fully integrated with WordLift and Helixware Plugin
5	JSONP support for cross-domain communication useful for content federation

6	Client side native caching
7	Easy to use listener configurations

The WordLift Rendering Engine is an early stage prototype. On the same idea we're designing the front-end widgets for WordLift v3.

### Shoof

Shoof allows users to record videos on the go, share it with friends and publish it over blogs and social networks.

First, Shoof works in conjunction with HelixWare and pulls all recorded videos into the cloud making them accessible across multiple screens (each video is ingested and pre-processed by HelixWare before being available for online preview).

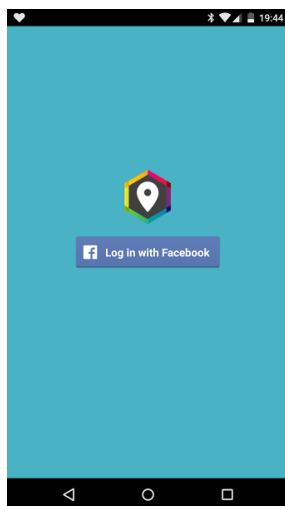
Second, each video belongs to the place where it has been recorded. All videos from the same neighborhood help creating the unique identity of that area of the city.

Third - where we expect Shoof to really shine - the App shall create "Stories" made of the best 8 seconds of all videos belonging to the same "album" (and album can be created around a neighborhood, an hashtag or a user).

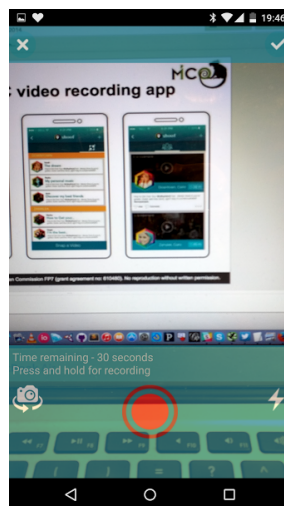
That is Shoof. Capturing the pulsing moments of a city and sharing them in a cohesive way over the cloud with a tight integration with users blogs and social networks.

The Beta of Shoof is designed and developed (with love) in Cairo by Insideout Today using the native Android application development SDK.

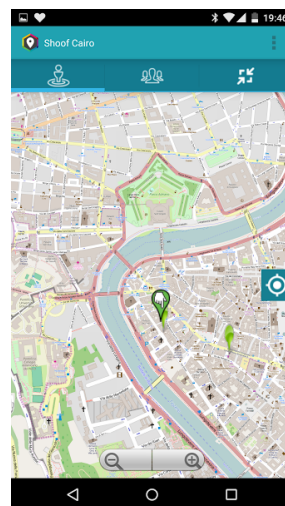
The user experience is straightforward.



**Social Login**



**Video Recording**



**Confirm Geo**



**Preview & Share**

### Features provided by Shoof

ID	Description
1	A “Vine” style video recording
2	Geo-localisation of each recorded video
3	Mapping of each video to a specific neighborhood and area of the city of Cairo
4	Video upload to HelixWare
5	Video post-processing via HelixWare (currently a rotation is applied to the video after recording)
6	User tagging for each video
7	User description of each video
8	User commenting for each video
9	Social Sharing
10	Local playout (on the user handsets) and multi-screen playout via HelixWare (including video-embedding on WordPress blogs running the HelixWare plugin)

Shoof v1 is a prototype currently made available in closed-beta to a selected number of content creators.

## Architecture

### Greenpeace

The Greenpeace showcase has different roles for editors and users, whereas editors are authorized people from within the organization that publish tailored content to the web site. They are empowered with the HelixWare Cloud platform and HelixWare Plugin for WordPress in order to be able to ingest audiovisual contents to the web site.

These contents are transcoded by HelixWare and made available for a variety of devices as well sent to MICO for enhancement such entity detection, video segmentation, audio/video quality, speech-to-text transcription and so forth.

The enhancement results of MICO's analysis are used to increase the overall user experience, with better findability and content access. A key feature is the MICO recommendation engine which receives the web site data (content views and user profiling) which is then used to suggest users relevant contents that might be of personal interest.

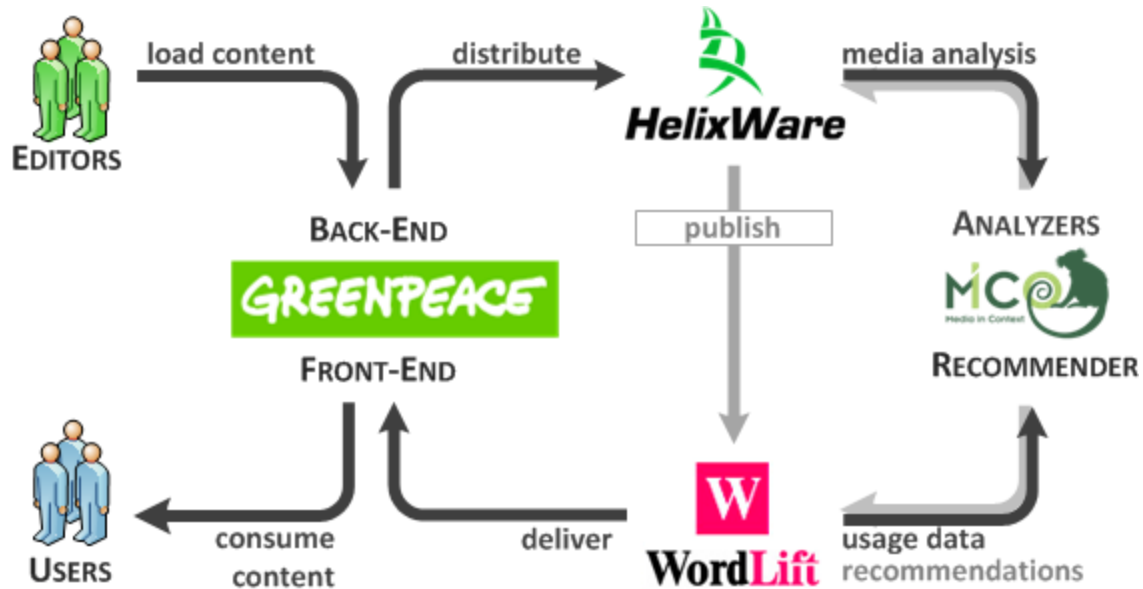


Figure 8 - The architecture of the Greenpeace showcase

## Shoof

The Shoof showcase involves several components that work together to fulfill the desired user experience. Following is the list of components:

- the Shoof mobile application, is an application for smartphones that allows live recording and consumption of audio/video contents;
- the Shoof backend, which holds users' data and preferences,
- the HelixWare platform, which stores online videos and performs automatic transcoding and streaming to multiple devices;
- the MICO platform:
  - the analyzers subsystem, which provides media analysis;
  - the recommender subsystem, which provides the recommendation engine;
- the WordLift plugin and backend, which turn existing CMS such as WordPress into semantic CMS.



Following is a high level flow diagram:

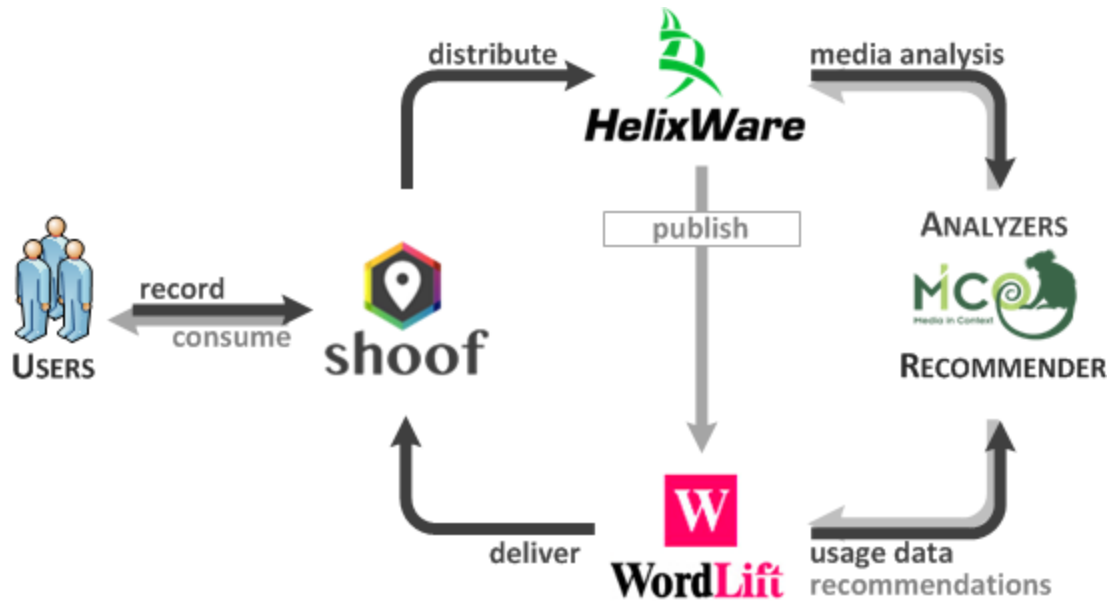


Figure 8 - The architecture of the Shoof showcase

Starting from the left side, Shoofers (Shoof users) record live audio/video from their surroundings using the Shoof mobile application. The recorded content is sent to the HelixWare platform, which stores it on the cloud and performs automatic transcoding of the original recording into several bitrates and formats in order to be able to stream to a variety of devices and at different bandwidths.

HelixWare sends a task request to MICO to perform media analyses such as audio/video quality analysis, nudity detection, copyright detection, video segmentation, speech-to-text transcription, entity recognition, and so forth. The results are stored by HelixWare in its local datastore. Contents are progressively published to WordLift along with the analysis results.

The Shoof mobile application delivers the published audio/video assets by using WordLift as back-end. Every access is logged by WordLift along with user data; the information gathered is then transmitted to MICO recommender to generate recommendations that will further enhance the content delivered to Shoofers.

### Staging and Production environment

A staging environment was configured for “Greenpeace Magazine” use case. All contents in staging - posts, images, videos - are duplicated from production environment in order to keep the environment as much as possible consistent: our goal is to use this staging environment to gradually introduce architectures components and validate the MICO integrations once available.

**The staging environment includes at the moment working instances of both WordLift and HelixWare.** Semantic annotation / tagging was performed on available contents and a first static json feed for the WP5 prototype was generated. WordLift content discovery widgets were integrated with Greenpeace Magazine frontend.

Testing strategy will be a 2-steps iterative strategy where the first step is mandatory for the second one:

1. **Technical validation in staging** - MICO integrations (TEs and WP5) will be firstly tested on the staging environment in order to check all expected functionalities and requirements compliance;
2. **Performance evaluation in production** - MICO outputs will be pre-processed on staging and submitted in production for a first evaluation round. Evaluation process will be focused on A/B testing and kpi monitoring.

The testing environment for Shoof consists of an application server running the open source framework Ruby on Rails. On this application server the middleware intercepts the requests from the clients and embeds the business logic of the application. The middleware also interfaces HelixWare for video processing and streaming. A web front-end running WordLift will also be made available in the coming weeks providing quick web access to the videos.

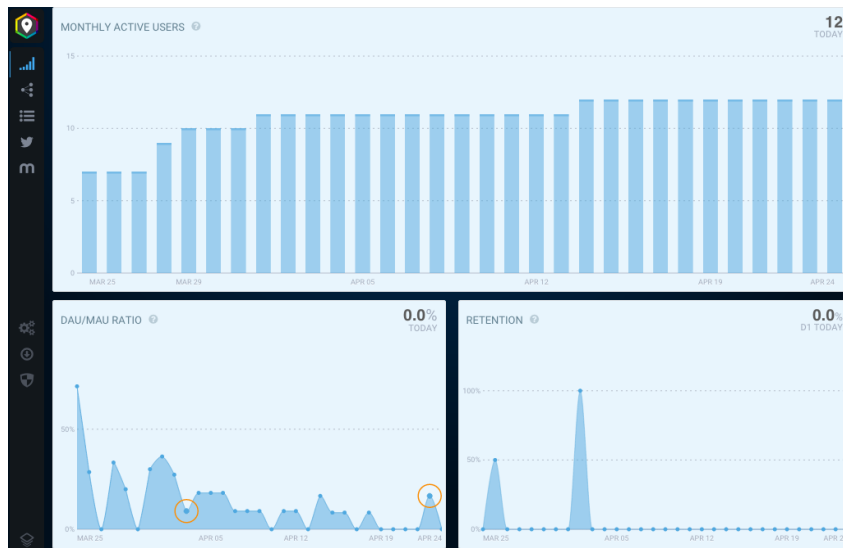


Figure 9: crashlytics dashboard used for the closed beta of Shoof

For the mobile application deployment, for the app analytics and to gather system and user feedback from the close beta the online service crashlytics (crashlytics.com) is been used.

## Integration with MICO WP5

The MICO recommendation engine - WP5 - plays a strategic role in our architecture. Our challenge is to offer cross-media interesting / contextual meaningful content suggestions both to editors (better supporting their editing workflow) and to readers (increasing their engagement). WP5 is where we expect to start the validation of MICO.

In the last months, we defined the full requirements list for WP5 integration<sup>8</sup> and started to support WP5 team providing test data useful for the first WP5 prototype setup. We defined different use cases that can be split in:

1. **content item based recommendation** use cases, where users interactions are not required and content similarity depends on content item properties;
2. **user interaction based recommendation** use cases, where content similarity depends on user interactions on the contents.

Follows a quick overview of required use cases outputs:

ID	Type	Description	Required by
1.1	Content Item Based Use Case	<b>Ranked list of similar contents</b> where similarity depends on content items related entities. A custom similarity function has to be used in order to give more prominence, in similarity calculation, to those entities marked as content item <i>about</i> .	Greenpeace Magazine
2.1	User Interaction Based Use Case	<b>Ranked list of similar users</b> where user similarity depends here both on users interactions and contents related entities: if user $u_1$ likes <sup>9</sup> an item $i_1$ about entity $e_1$ then $u_1$ is similar to other users who likes contents - item $i_1$ included - about entity $e_1$ .	Shoof
2.2	User Interaction Based Use Case	<b>Ranked list of recommended contents depending on current user interests</b> defined through its previous interactions on contents: if user $u_1$ likes an item $i_1$ about entity $e_1$ then he could be interested in other items about entity $e_1$ .	Greenpeace Magazine, Shoof

<sup>8</sup> [WP8: VIDEO-87 on Jira](#)

<sup>9</sup> [UserLikes in schema.org](#)

2.3	User Interaction Based Use Case	<b>Ranked list of recommended contents liked by similar users and / or friends:</b> If user $u_1$ is friend of / is connected with user $u_2$ and user $u_2$ <a href="#">likes</a> an item $i_1$ about entity $e_1$ then also $u_1$ could be interested in item $i_1$ and in other items related to entity $e_1$ .	Shoof
-----	---------------------------------	---	-------

The WP5 prototype supports at the moment the use case 1.1. We provided production data - a set of articles coming from the new “Greenpeace News” magazine first issue launched in april - tagged / classified with WordLift on our staging environment. It’s a tiny but meaningful dataset made up of 12 items about environmental related different issues. Each item - identified by it’s uri - is defined by its related entities and about collections. Each entity its identified by it’s uri. Follows a sample resource:

```
[
  {
    "id": "http://data.redlink.io/91/be2/post/MAS_-_Parole",
    "about": [
      "http://data.redlink.io/91/be2/entity/MAS",
      "http://data.redlink.io/91/be2/entity/OGM",
      "http://data.redlink.io/91/be2/entity/cambiamenti_climatici"
    ],
    "entities": [
      "http://data.redlink.io/91/be2/entity/MAS",
      "http://data.redlink.io/91/be2/entity/OGM",
      "http://data.redlink.io/91/be2/entity/Greenpeace",
      "http://data.redlink.io/91/be2/entity/cambiamenti_climatici"
    ]
  },
  [...]
]
```

This dataset just includes textual contents at the moment. As soon as MICO extractors are available we should be able to include video assets in our feed in order to improve this use case. First WP5 prototype outputs are encouraging and represent a good start that need to be validated in the next months.

We also started to collect user interactions - pageviews - for registered / profiled users on “Greenpeace News” magazine. This data will be available soon for other user interaction based prototyping.

## Integration with MICO WP6

The main challenge when integrating with MICO WP6 is to allow for a variety of technical enablers which may return a vast amount of data using different formats and vocabularies. The result values may be returned all at once or progressively according to the completion time for each task. Also the actual results need to be displayed to users in a way that they can be interpreted and be accessed and used in a meaningful manner.

Within this context, HelixWare has been refactored to support **enterprise integration patterns**, decoupling the general code core of the platform and the tasks it can initiate towards external systems. As part of this refactoring the following key features have been implemented:

1. application events for incoming media and transcodings, which can be used by several other components to be notified when a video is available;
2. support for Apache Camel, in order to define flexible workflows according to the *Enterprise Integration Patterns*;
3. support for AMQP via the RabbitMQ component to reliably send and receive message across distributed systems.

This structure will allow HelixWare to support the MICO analyzers tasks by delivering the required data and progressively processing the results. Each pluggable component will be able to understand the results from MICO and display it in the HelixWare UI either in the HelixWare Cloud backend or via the HelixWare WordPress plugin.

Following is a high level diagram of the HelixWare integration with MICO (the diagram will be further detailed as soon as the specifications for the integration will be formalized in WP6):

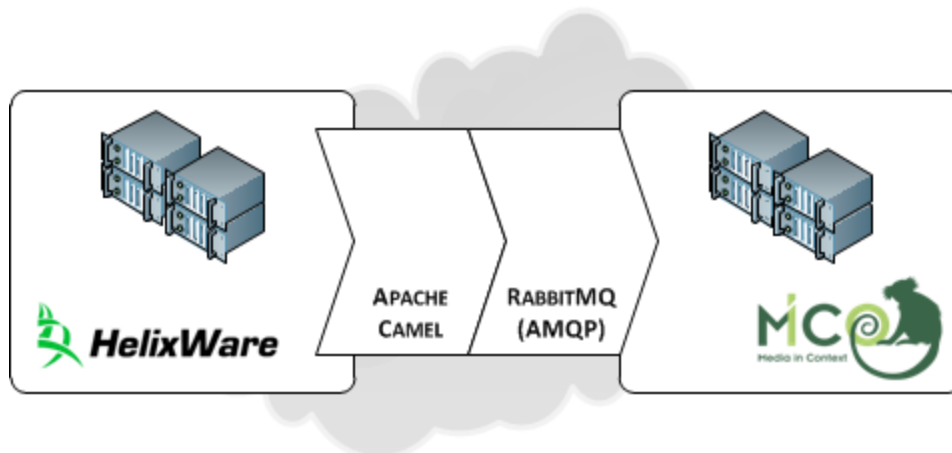


Figure 10 - HelixWare integration with MICO

## User Stories / Requirements / Goals

User stories	Goal
US-06, US-21, US-22, US-23, US-24, US-25, US-26	Get ready for UGC based news making process
US-07, US-08, US-09, US-12, US-18	Turning visitors into readers
US-07, US-08, US-09, US-19	Exploit media in context
US-10, US-11, US-20	Offer customized contents
US-12, US-18, US-60, US-20	Recommend cross-issue content discovery path

## MICO Prototype - Technology Stack Available for Testing

ID	Name	Description	Availability
/TE-205	A/V Error Detection and Quality Assessment	A/V Error Detection and Quality Assessment, especially for camera read errors, to remove artefacts, or for quality ranking / filtering purposes	<b>Must</b>
TE-206	Temporal Video Segmentation	Temporal video segmentation, for easier navigation, segment annotation, key frame extraction	<b>Must</b>
TE-204	Face Detection & Recognition	Face Detection & Recognition for cross media entity annotation	<b>Should</b>
TE-214	Speech-to-Text	Do automatic speech recognition on video material to produce a time-stamped transcription. The extractor scope is restricted to content that is free from background music and / or excessive noise.	<b>Should</b>

<b>TE-213</b>	Sentiment analysis	Determine the polarity of forum entries, e.g. positive, negative, or neutral	<b>Could</b>
<b>WP5</b>	Cross-media recommendation engine	Offer cross-media, meaningful and contextual content suggestions.	<b>Must</b>

## Functional requirements mapping to available MICO components

Product	Feature Description	TE	ShowCase
<b>HelixWare</b>	UGC content filtering / moderation (technical quality assessment, technical error detection, nudity detection, copyright infringement detection).	<b>TE-205</b>	<b>Shoof</b>
<b>HelixWare</b>	Advanced suggestions for video editing.	<b>TE-206 TE-214 TE-204</b>	<b>Greenpeace Shoof</b>
<b>HelixWare</b>	Enhanced video scrolling with previews and "chapter" bookmarks (using video segmentation)	<b>TE-206</b>	<b>Greenpeace Shoof</b>
<b>WordLift</b>	Semi-Automated video annotation and entity linking in WordPress media gallery	<b>TE-206 TE-214 TE-204</b>	<b>Greenpeace</b>
<b>HelixWare</b>	Smart Playlists - content similarity based automated playlist building (select all meaningful videos around a given place about a specific issue detecting overlapping).	<b>TE-206 TE-214 TE-204</b>	<b>Shoof</b>
<b>HelixWare</b>	Recommendation for Wordpress - content suggestion based on user profile or content item similarity	<b>WP5</b>	<b>Shoof Greenpeace</b>
<b>HelixWare, Wordlift</b>	Boosted Dynamic Interlinking - cross-media contextual content suggestions (suggest meaningful articles depending on current video segment relevant entities; suggest meaningful videos / video segments	<b>TE-206 TE-214 TE-204</b>	<b>Shoof Greenpeace</b>

	depending on the current paragraph relevant entities).		
<b>Wordlift</b>	Sentiment analysis for Wordpress - comments moderation sentiment analysis based support.	<b>TE-213</b>	<b>Shoof, Greenpeace</b>

## Conclusions

Great progress has been made in refining all business requirements for the video news showcase, proper engagement among all stakeholders has been guaranteed for both scenarios and an end-to-end environment, traversing the different applications, has been setup for the validation of MICO.