



## **SEVENTH FRAMEWORK PROGRAMME**

### **Networked Media**

*Specific Targeted Research Project*

# **SMART**

(FP7-287583)

## **Search engine for Multimed*i*A environment generated conten*T***

### **D6.3 Evaluation of the SMART Search Framework and Applications**

Due date of deliverable: 01-05-2014

Actual submission date: 03-06-2014

Start date of project: 01-11-2011

Duration: 36 months

## Summary of the document

<b>Code:</b>	<b>D6.3 Evaluation of the SMART Search Framework and Applications</b>
<b>Last modification:</b>	02/06/2014
<b>State:</b>	Final
<b>Participant Partner(s):</b>	S3Log, Prisa, Glasgow, SDR
<b>Author(s):</b>	Massimiliano Tarquini, Dyaa Albakour, Romain Deveaud, Tomás Garcia Fresno, Jorge Tome Hernando
<b>Fragment:</b>	No
<b>Audience:</b>	<input checked="" type="checkbox"/> public <input type="checkbox"/> restricted <input type="checkbox"/> internal
<b>Abstract:</b>	This deliverable describes the project's methodology towards evaluating the SMART search framework and its use cases (live news, security/surveillance, social search). It also includes a first round of evaluation results corresponding to the first evaluation cycle of the project.
<b>Keywords:</b>	N/A
<b>References:</b>	See within document for references.

## Table of Contents

1.1	Scope .....	7
1.2	Audience .....	7
1.3	Terminology .....	7
1.4	Structure.....	7
Methodology .....		9
1.5	Introduction .....	9
1.6	Quantitative Evaluation Approach: Annotation for Event Detection and Ranking .....	9
1.6.1	Requirement of annotation .....	10
1.6.2	Methodology .....	10
1.7	Evaluation of Venue Recommendations.....	11
1.8	Evaluation of Proof of concept Applications based on Workshops .....	14
1.8.1	SMARTCOP evaluation workshop.....	14
1.8.2	Live news Evaluation workshop.....	15
2	Event Annotation Infrastructure.....	16
2.1	Video formatting .....	16
2.2	Event annotation web interface.....	16
2.3	Pilot Study .....	19
2.3.1	Study of the interface.....	19
2.3.2	Study of the annotation effort.....	20
3	Evaluation Results of SMART Applications based on Workshops .....	22
3.1	Introduction .....	22
3.2	Security and Surveillance.....	22
3.2.1	Santander SMARTCOP evaluation workshop results .....	22
3.2.2	S3Log evaluation workshop results .....	24
3.3	Live News Evaluation workshop results.....	26
3.3.1	Santander LIVE NEWS evaluation workshop results .....	26
3.3.2	Livenews applications. El País evaluation workshop .....	27
4	Venue recommendation evaluation results .....	30
5	Conclusions .....	32
5.1	Key findings.....	32
5.2	Lessons learned.....	33
5.3	Future Evaluation steps .....	33
6	References .....	34
7	Appendix I – SMART COP Evaluation Questionnaire .....	35



8	Appendix I – SMART COP Evaluation Questionnaire .....	37
---	---	----

## Table of Figures

Figure 1: City selection step. The participants have also the opportunity to give information about their knowledge of city.....	11
Figure 2: Judgment interface, asking the user to judge if the venue “M&M’s World” corresponds to the type of activity he would do at 5pm while being in Farringdon, London. ....	13
Figure 3: Introduction page of the event annotation interface. Three examples are shown to the annotators to give them a sense of the intensity of events in the city of Santander. ....	17
Figure 4: Annotation interface, showing the video as well as the four action buttons.....	17
Figure 5: When the annotator clicks on the “Event starts” button, he can enter information about the event in the form displayed on the right-hand side of the interface. ....	18
Figure 6: When the annotator clicks on the “Event ends” button, the information about the event are recorded and shown to the annotator.....	19

## Table of Tables

Table 1: Main Elements of the SMART Quantitative Evaluation Approach .....	9
Table 2: Agenda of the Evaluation Workshops .....	14
Table 3 Number of Identified Events with Change Point Analysis in Ayuntamiento .....	20
Table 4: Projection of annotation effort .....	20

## Executive Summary

### 1.1 Scope

The main objectives of WP6 within the SMART project are to demonstrate the concepts and technologies defined in the other work packages and associated tasks. The demonstration activities are focused on providing specific solutions to the SMART use cases.

The objective of this deliverable is to present the results of the task 6.3 focusing on evaluation of the proof of concepts implemented in each of the tasks and presented in the associated deliverables D6.1 and D6.2.

As a result of this step, the target is to obtain a quantitative and qualitative multi-facet assessment of both the SMART open source platform and of the SMART applications. To this end, the project has specified a range of complementary evaluation modalities aiming at covering different evaluation aspects associated with the SMART platform and applications. These modalities will be used to carry out the evaluation on the basis of two cycles.

The present document is devoted to elaborating on the evaluation methodology, as well as to describing the results of the first evaluation cycle. A subsequent version/release of the deliverable (at the end of the project) will include the results of the second evaluation cycle.

### 1.2 Audience

This document is mainly addressed towards:

- **SMART project members:** Members of the SMART team will find in this deliverable valuable insights (including feedback potential users and stakeholders) towards improving the SMART platform and its application.
- **Open Source Community:** This deliverable will be of interest to open source contributors and users dealing with the SMART platform, since it can help them identify strengths and weaknesses of the open source platform and applications along with areas for potential contributions.
- **Researchers on Search Engines and IoT Systems:** IoT/Search researchers are likely to have an interest on the deliverable in their effort to understand the scope of the SMART systems and to position them to the overall IoT and search research landscape.

### 1.3 Terminology

- **A High-level event:** Something interesting happening in the real-world that a use-case would notify a user about. It has a location, a start time and an end time. For example, a concert or a demonstration in the town square.
- **A Low-level event:** a real-world occurrence that can/should be perceived in the metadata. For example, an increasing traffic in the town square can be detected but it is not necessary interesting.

### 1.4 Structure

The core of the document is split into three distinct parts.

**Section 2** provides details about methodology used to test the different pilots;

**Section 3** provides details about the infrastructure used for data annotation;

**Section 4, 5** are about the evaluation results and conclusions.



**Section 6** concludes the document.

**Section 7** is about references.



## Methodology

### 1.5 Introduction

The present deliverable comes after two previous ones produced by this work package D6.1 and D6.2, which illustrate the implemented prototype version of the SMART open source platform, as well as the SMART applications. Both deliverables are in their second cycle and are, respectively, devoted to reporting about the proof of concepts and the open source platform.

This chapter is devoted to describe methodologies and tools adopted to proceed with evaluation of the project outcomes reported in deliverable D6.1 and D6.2. To be able to evaluate the open source platform, we have adopted different approaches: a quantitative approach based upon annotation for event detection and ranking, and a qualitative approach especially for the evaluation of the prototype (proof-of-concept) applications. Note that the qualitative evaluation involved the engagement of stakeholders such as potential users of the SMART applications.

The quantitative evaluation approach is based upon the following schema/table:

Quantitative evaluation approach	
Research Question	How accurate is the response to high-level events
Annotation methodology	Which is the adopted annotation methodology? Which tools have been used to annotate?
Annotation volume	The volume of the annotated data
Scoring	How the scoring tools works?

**Table 1: Main Elements of the SMART Quantitative Evaluation Approach**

Differently from the open source framework, the methodology adopted to annotate the test application uses a qualitative approach and is intended to evaluate the deliverables reported in D6.1 by the stakeholders' point of view. To achieve those objectives we have adopted a method based upon several evaluation workshops, each one of each engaged stakeholders notably potential users of the evaluated applications.

Following paragraphs provide more information, for both the quantitative and the qualitative evaluation methodology. In particular, they describe our approach to event annotation (including the tools used), along with the structure of the evaluation workshops.

### 1.6 Quantitative Evaluation Approach: Annotation for Event Detection and Ranking

Since the SMART infrastructure has been deployed in Santander, we aim to further investigate the effectiveness of our models for the detection and retrieval of high-level events developed in [SMART-D5.3]. These models were only evaluated empirically on video analysis datasets collected in indoor environments or using public social media posts geotagged within the city. The availability of the infrastructure deployed in Santander opens up opportunities to further evaluate these models in real-world scenarios for our designated use cases.

In order to do this, we aim to build a test collection that consists of real-world events identified manually or from external resources to serve as ground truth data suitable for evaluation. The collected real-world events together with the meta-data gathered from the various SMART perceptual components

(e.g. crowd video analysis, audio classification and weather data) and the associated social media posts (e.g. geotagged tweets and FourSquare check-ins) will be a valuable novel resource that can be used for evaluating event retrieval models within and beyond the scope of the SMART project. For example, such test collection is necessary for a learning-to-rank based event retrieval model.

In this section, we describe our requirement and methodology for annotating real-world events using the infrastructure deployed in Santander.

### 1.6.1 Requirement of annotation

We aim to build a test collection, similar to those created for document retrieval tasks within the information retrieval community (e.g. the TREC evaluation campaigns). In particular, for a given textual query at a given time and a location, we should know the start and end times of relevant high-level events for that query. As per standard good practices in the IR community, we target at least 50 such queries for labelling. Therefore, we should obtain a textual description of each relevant event. To generate realistic user queries, an event and an approximate time should be provided to a user and ask the users how they would search for such an event (to simulate the query formulation habits of real users). Also, to support graded relevance, we can label events with a degree of importance (e.g. high density events vs. low density events).

We could reverse-engineer known high-level events into the test collection. For instance, events may be identified (a) from the agenda of the city<sup>1</sup> or (b) by the results of existing models or their annotation. This task can be crowdsourced using Amazon's AMT service. Moreover, we will value the types of queries the use-case partners foresee for their specific use cases.

Finally, the zero query situation should be considered, where we know high-level events are going on in a given location at a given time.

### 1.6.2 Methodology

Our methodology for annotation aims to meet the requirement identified above. Using the SMART infrastructure deployed in Santander, we have collected over a thousand hours of video and audio recordings in two locations (the main Square in Santander, and public open market) over a period of approximately two months between October and December 2013 [SMART-D3.5]. In addition to the recordings, we have collected metadata from the audio and video analysis (e.g. crowd density, audio classifications, and other) and environmental observations (temperature, light, etc.), in addition to social media posts in Twitter and visit counts in FourSquare.

Since it is very expensive to examine the entire recordings manually and annotate them with real-worlds event, we employ a pooling approach [Spark-Jones1975] commonly used in IR campaigns, such as TREC<sup>2</sup>, to generate a ground truth for relevance. With pooling, retrieval systems can be evaluated without having to judge all documents in the collection. Instead, a pool is constructed by putting together top n retrieval results from a set of N systems. Humans can then judge every document in the pool and documents outside the pool are considered irrelevant.

Similarly, we aim to identify candidate segments of videos (these are the documents in our case) where we have some belief that an actual event may be appearing in the video (i.e. the event is taking place in the corresponding time and location). To generate these candidate segments (the pool), we employ two different approaches:

- We employ our change component of the event ranking model developed in D5.3 [SMART-D5.3], on the various feeds. It will identify points of time where an observation has changed unusually in a location (unusual change in crowd density from the video analysis or unusual traffic heard from the audio analysis). We consider video segments corresponding to these points of

---

<sup>1</sup> In which case, the agenda feed cannot be a feature considered by the search engine.

<sup>2</sup> <http://trec.nist.gov>

time to be potential candidates of videos that require annotation.

- We also consider segments of videos corresponding to planned events in the city agenda.

The generated candidate segments of videos are then annotated with human assessors that should examine the videos and describe events that they may observe. The annotation should meet the requirements identified above. Therefore, the annotators should indicate whether an event has happened when an event starts exactly. In addition the annotators can describe the events with keywords which are considered to be the queries for finding these events. Each candidate video segment can be annotated by more than one human assessor and their judgments can be aggregated using an appropriate voting mechanism.

For this purpose, we developed a web-based annotation interface and performed a pilot study to refine this interface and make ready for large-scale annotation. We describe this interface in Section 5.

## 1.7 Evaluation of Venue Recommendations

As a part of our evaluation effort for assessing the performances of the recommendation models we implemented, we reported in [SMART-D5.3.2] the results of a user study that involved 100 participants. The technical details and experimental methodology are already mentioned in [SMART-D5.3.2]. Here we will only focus on the web interface that we set up (which is based on the venue recommendation demo), and on the process of judging the recommendations.

The Venue Recommendation model we implemented infers the interests and preferences of the users from the pages they *Like* on Facebook. Consequently, we first had to recruit participants that owned a Facebook account, with ideally a few pages liked in their profile. As already emphasised in previous deliverables, we do not store the personal information of the participants, and require their Likes through the official authentication plugin provided to developers by Facebook<sup>3</sup>.

### Are you interested in these venues?

We will ask you to judge several suggestions of activities in London, Amsterdam or San Francisco given a location in the city and a time of the day. Please rate **how likely you are** to go to these venues (~ 5 to 10 minutes).

We require assessors to log in with their Facebook account in order to infer their interests and suggest personalized venues. Please bear in mind while judging suggestions that they may not necessarily be touristic venues.

As the suggestions will come from different algorithms, **not all of them would be relevant** but this is normal.

Please log in with your Facebook account (blue link above), then select a city for which you want to judge suggestions:

London

How much familiar are you with London?

- ☐ I have never been in this city.  
☒ I already visited this city.  
☐ I lived in this city.

Begin

**Figure 1: City selection step. The participants have also the opportunity to give information about their knowledge of city.**

After their login, we asked participants to choose one city between London, Amsterdam, and San Francisco, the three cities we consider in our evaluation (see Figure 1). They were also required to detail their level of knowledge of the city they selected: this allowed us to further identify patterns of venue preferences towards certain types of users (residents vs. tourists).

After clicking on the “Begin” button, the participants were presented with a map showing their current

<sup>3</sup> <https://developers.facebook.com/docs/facebook-login/>

location and the location of the current venue to judge, a small text describing the context (time, location, judging instruction), a bloc describing the venue, and a form allowing the user to enter his judgments (not likely/likely/highly likely to visit the venue). An example of such a page is displayed in Figure 2. The participants cannot go to the next venue to judge unless they have selected one of the three available judgments. Clicking on the “Next suggestion” button (not shown on Figure 2) brings the user to another similar page, with another venue to judge. A progress bar (not shown on Figure 2) allows the participants to check at any time the number of remaining venues to judge.

We record each judgment as soon as it has been validated (when the participant goes to the next venue); hence if a participant drops the judgment task in the middle, we still have a few judgments from him. Once all venues have been judged, the participants were asked to enter optional comments in a text box. We detail this qualitative feedback in Section 4 Venue recommendation evaluation results



**Situation:** you are in **Farringdon, London**, and the time is **5pm**. Your current location is marked on the map above by a green arrow.


Given this situation, **how likely would you visit the venue described below (marked by a red bubble on the map) in the next hour?**

- ☐ Not likely  
☐ Likely  
☒ Highly likely

**M&M's World**

<http://mmsworld.com>

Categories: Candy Store.



**Figure 2: Judgment interface, asking the user to judge if the venue “M&M’s World” corresponds to the type of activity he would do at 5pm while being in Farringdon, London.**

## 1.8 Evaluation of Proof of concept Applications based on Workshops

Towards evaluating the proof of concept applications a set of workshops were organized. The main idea was to engage industry stakeholders with relevance to the proof-of-concept Live News and Security/Surveillance (COP) applications, along with , operators and citizens. Despite difference in the audience, evaluation workshops were all organized around the following agenda:

10'	Welcome and Purpose of the Workshop
10'	Introduction of the participants
10'	Introduction to SMART
15'	Demonstration of SMART Applications <ul style="list-style-type: none"><li>• Live News</li><li>• Security &amp; Surveillance</li></ul>
15'	Use of the Application (optional)
15'	Feedback and General (Free) Discussion
15'	Questionnaire Form Filling
10'	Getting Involved and Next Steps
10'	Wrap Up and Conclusion

**Table 2: Agenda of the Evaluation Workshops**

As evident from the agenda, the goal was to keep the workshops short, focused and productive.

The first half of each workshop was devoted to identify the skills of the participants and to present the SMART project and the applications. It was followed by a free discussion about the proof of concept application to be evaluated.

Evaluation questionnaires was also created and administered to the participants. While the philosophy of the questionnaires was similar for each one of the workshops, there were several variations to the questionnaires based on the target audience and the type of proof-of-concept being evaluated. The aim of each questionnaire was to evaluate each proof of concept in a qualitative way. The intent is to evaluate how the demo application fits with their respective domains (live news, security), which are the most relevant functionalities, are the applications ready for the market. We also asked to participant to suggest extra features and potential uses of the applications with the intent to drive the last few months of development and lay the foundation for future development of the products.

In order to achieve the result we proposed the evaluation questionnaire to an audience made up of different categories: citizens and operators, business developers, technicians, potential customers. One of the key aspects was to show the application to the users but not giving to them too many technical details. The main goal with this workshop was to collect the features and requirements suggested by them but from their perspective (as users or managers of the city services in the case of citizens or operators, or as security experts in the case of policemen or security companies).

### 1.8.1 SMARTCOP evaluation workshop

The questionnaire used in the scope of SMARTCOP evaluation workshops is made of three sections:

- First section is about contact details and general information about contact activities;



- Second section collects qualitative data proposing questions with a structured answer;
- Third section is devoted to free comments and conclusions.

The questionnaire is depicted in Appendix 1.

### 1.8.2 Live news Evaluation workshop

Following the same structure than the questionnaire used in the SmartCOP workshop the questions for the final users and operators were the same. Only the fifth question was adapted in order to know their preferences about the different functionalities shown during the presentations.

**Identify functionalities of the application most relevant to you and your organization (SECURITY EXAMPLE):**

Functionality	Tick
News of Interest from Social Networks	
Scheduled events in the city	
Information about POI (Museums, tourism, restaurants, ...)	
Other Specify: _____	

The questionnaire is depicted in Appendix 2.

## **2 Event Annotation Infrastructure**

In Section 2.4, we described our requirements and the devised methodology for event annotation. In this section, we describe the infrastructure we have developed to perform such annotation.

### **2.1 Video formatting**

The recorded videos of the two locations in Santander exist as hourly segments in very high resolutions (3MP) in the MKV format, where a video file may exceed 1.5 GB. Therefore, they do not necessarily correspond to each identified candidate segment for annotation and they are not suitable to be presented and used in our annotation interface. Hence, those videos needed to be processed and formatted in order to be used in our interface. In particular, the video formatting involves:

- Transferring the videos into lower reasonable resolutions (1024x768) and a compressed format (mp4) that can be played in standard web browser like Chrome.
- Obtaining on demand a certain chunk of a video.
- If necessary, merging two chunks of videos that correspond to a segment of time which overlaps two videos.

### **2.2 Event annotation web interface**

In this section, we detail the web interface that we set up to support the event annotation task described in Section 1.6. The interface first introduces the task to the annotators and gives three examples of events with varying intensity (see Figure 3). The goal of this introduction is to familiarize the annotators with the process, and to give them some context about the city of Santander and the usual intensity of the events. As we can see from Figure 3, the annotator can then choose from 6 different videos that we selected for the sake of a first pilot study. The complete interface will not give this choice to the annotator, and will randomly present a video from our generated pool.

Clicking on one the videos brings the annotator to the annotation interface, which comprises two main components: the video player and the action buttons (see Figure 4). These action buttons are: “Fast forward”, “Event starts”, “Event ends”, and “There is no event in this video”. The first one allows speeding up the video playback (in case there are no clear event for example) by holding down the button. For an easier browsing, annotators can also click on the time cursor of the video player to jump to a specific part of the video. The last button is a convenient way for the annotator to state that there are no events in the given video.

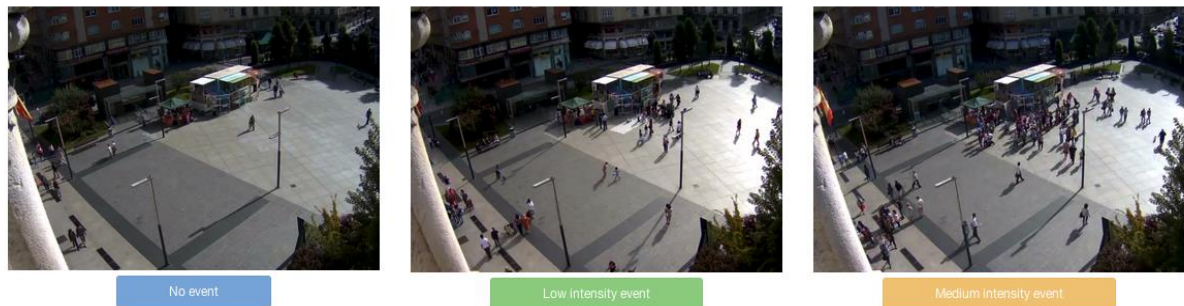
The “Event starts” and “Event ends” buttons represent the core of the annotation. When the annotator notices an event in the video, he can click on the “Event starts” button to indicate the starting time of the event. Once he does this, a form appears on the right-hand side of the interface, allowing him to enter information about the event (see Figure 5). The video continues to play while the annotator fills the form.



## Annotation of events happening within a smart city.

We will ask you to annotate videos by **identifying events** that are happening. These annotations will form a ground truth from which predictive models can be learned and **applied to unlabelled data**.

You will see a video from a place in Santander (Spain), and will be asked to identify the different periods when events occurred (if any). Below are three examples of the states of this place: no event, an event is building (people are gathering, low intensity), and an event is occurring (with medium intensity). Once you are ready, choose between one of the videos proposed below to start annotating.



Please choose between one of the following available 10 minutes videos:

Video 1

Video 2

Video 3

Video 4

Video 5

Video 6

**Figure 3: Introduction page of the event annotation interface. Three examples are shown to the annotators to give them a sense of the intensity of events in the city of Santander.**



Fast forward

Event starts

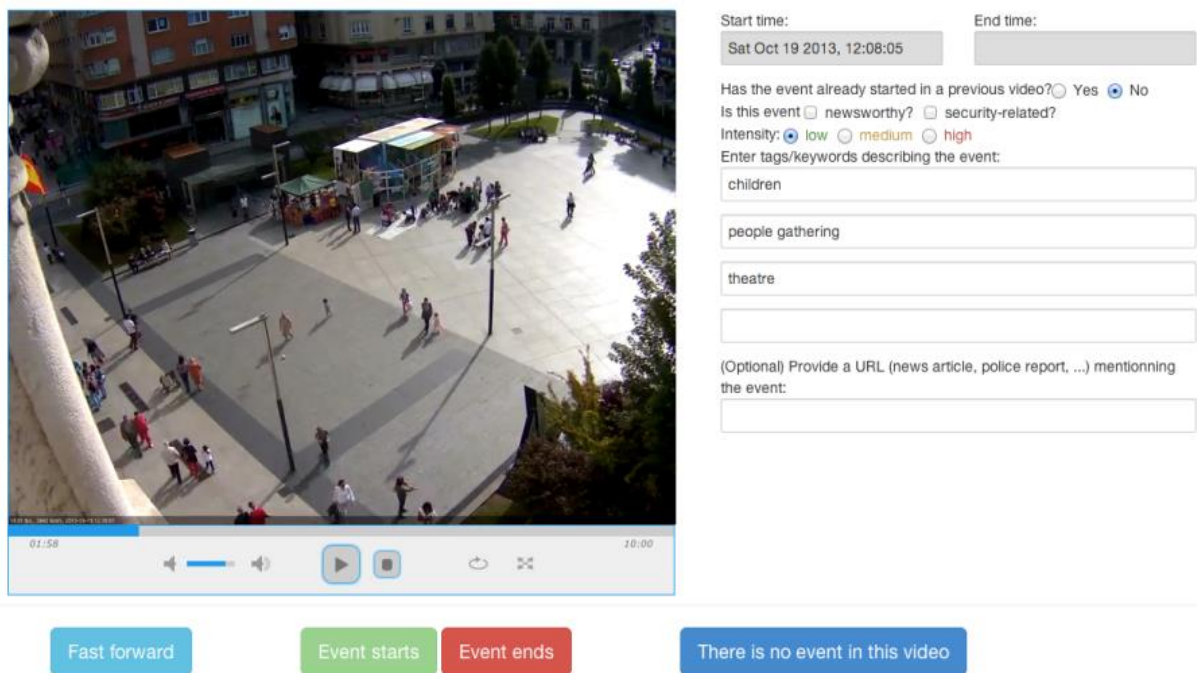
Event ends

There is no event in this video

**Figure 4: Annotation interface, showing the video as well as the four action buttons.**

The available fields of the form are:

- “Has the event already started in a previous video?” [Yes or No]: Some events can be very long and can span over multiple videos. This information allows us to identify the different videos that relate to a single event.
- “Is this event [newsworthy?, security-related?]”: A newsworthy event is an event that could have been reported in a local newspaper, such as a protest or a big market for example. A security-related event is an event that should have required a police intervention.
- “Intensity” [low, medium, or high]: How the annotator perceives the intensity of the event. This is a rather subjective criterion, but since we plan to recruit several annotators we could conduct inter-annotator agreement studies.
- “Enter tags/keywords describing the event”: The annotator can enter an unlimited amount of tags, which are single words or terms that allow understanding what is currently happening in the video.
- “(Optional) Provide a URL (news article, police report ...) mentioning the event”: If the event was significant enough to be reported on the Web, the annotator can give an evidence.



The screenshot shows a video player on the left displaying a street scene with people. On the right is an annotation form with the following fields:

- Start time: Sat Oct 19 2013, 12:08:05
- End time: (empty)
- Has the event already started in a previous video? ☐ Yes ☒ No
- Is this event ☐ newsworthy? ☐ security-related?
- Intensity: ☒ low ☐ medium ☐ high
- Enter tags/keywords describing the event:
  - children
  - people gathering
  - theatre
- (Optional) Provide a URL (news article, police report, ...) mentioning the event: (empty)

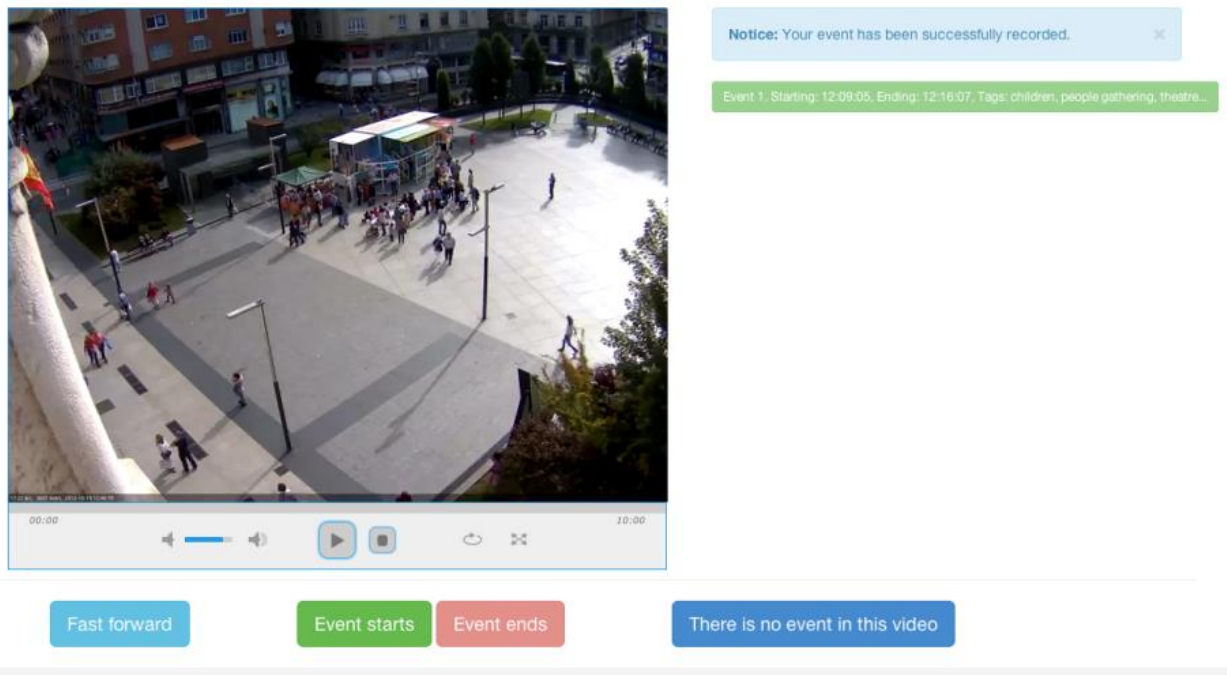
Below the video player are four buttons: "Fast forward", "Event starts", "Event ends", and "There is no event in this video".

**Figure 5: When the annotator clicks on the “Event starts” button, he can enter information about the event in the form displayed on the right-hand side of the interface.**

We can also notice from Figure 5 that the start time of the event is automatically inferred when the annotator click on the “Event starts” button.

When the annotator identifies the end of the event, he can click on the “Event ends” button. This button cannot be clicked until an event has been previously started; likely, an event cannot be started unless the previous one has been ended before. Ending an event automatically sends the event information to the web application server and stores them. Just like for the start time of the event, the end time is also inferred when the button is clicked. A notice, warning the annotator that the event information has been submitted, is displayed (see Figure 6). The list of all annotated events is also displayed on the right-

hand side of the interface.



**Figure 6: When the annotator clicks on the “Event ends” button, the information about the event are recorded and shown to the annotator.**

With regards to the requirements identified in Section 2.4.1, we have developed an annotation interface that would allow us to obtain judgments on a large scale for videos that correspond to candidate events in our generated pool. The interface meets the requirements of the annotation where we can obtain keywords describing the events and estimate the intensity of the event.

## 2.3 Pilot Study

In this section, we describe our pilot study to refine the annotation interface and estimate the annotation effort needed for the recorded data in hand. We first describe the study conducted on the interface and then describe the estimation of the annotation effort

### 2.3.1 Study of the interface

The pilot study involved performing annotation of six sample videos by 5 assessors. We recorded their annotations and asked for qualitative feedback. The comments were generally centred on the usability and the design of the interface, and were very valuable.

All the participants agreed that the “Fast forward” button was either confusing or not needed; while it is still present in the interface we are considering removing it. One participant asked us if we could increase the size of the video player or make it available in full screen, but we think it would harm the usability of the interface. One other reviewer highlighted the user-friendliness, which was encouraging. He said however that auto-completing tags could improve it. This is an interesting suggestion, but we cannot implement it. Auto-completion requires a predefined list of tags, and the entire purpose of this interface is to gather these tags.

Another very useful suggestion, which we will implement for the final version of the interface, is the possibility to edit an event after it has been recorded. Indeed, it would be useful for the annotators to be

able to add/remove/modify information after the end of an event. We will also propose several URL fields, just like we did for the tags.

Finally, some participants conducted a small stress test of the interface, highlighting some bugs and misbehaviours that can crash the application.

### 2.3.2 Study of the annotation effort

As described in Section 2.4, the annotation involves creating a pool of segment videos that can represent candidate of events. We generate this pool by employing our change component of the event ranking model developed in D5.3 [SMART-D5.3], on the various feeds. It will identify points of time where an observation has changed unusually in a location (unusual change in crowd density from the video analysis or unusual traffic heard from the audio analysis). Applying the change component on each sensor observation feed corresponds to taking the top retrieved documents from a number of retrieval systems in pooling approaches [Karen-Spark1975].

We made a pilot study where we estimated the annotation effort needed for the recordings within 3 days (5pm on October 16, 2013 to 5pm on October 19, 2013). We applied our change component of the ranking model on four different feeds of processed A/V sensor observations. Namely, we used (i) crowd analysis from video, (ii) crowd analysis from Audio, (iii) traffic analysis from audio and (iv) music analysis from audio. We used the experimental setup described in [SMART-D5.3] for the change component estimated by employing change point analysis with Grubb's test. In particular, given a location  $l_i$  and at each point of time, e.g. on minute intervals, we maintain a moving window of size  $k$  points, e.g.  $k$  minutes, over the previous observations. We apply the Grubb's test to each moving window to determine if the sensor observations  $r_j$  the last point is an outlier that stands out with respect

to previous observations. With Grubb's test,  $r_j$  is an outlier if  $v = (r_j - \bar{x}_{j,k}) / \sigma > z$ ,  $v = (r_j - \bar{x}_{j,k}) / \sigma > z$  where  $\bar{x}_{j,k}$  is the mean observations in the window  $(t_{j-k}, t_j)$ ,  $\sigma$  is the standard deviation in the window  $(t_{j-k}, t_j)$ , and  $z$  is a fixed threshold. We considered one location in the city (ayuntamiento) and report the number identified segments from each sensor observation feed in Table 1.

**Table 3 Number of Identified Events with Change Point Analysis in Ayuntamiento**

	Video Crowd Analysis	Audio Crowd Analysis	Audio Music Analysis	Audio Traffic Analysis
Number of Segments	8	3	2	3

To estimate the total effort required to annotate 2 weeks of recording, we project these figures for both considered locations and on the assumption that we will judge each video segment using at least three annotators. We also assume based on the pilot study of the interface that on average a 10 minute video can be judged within 2-3 minutes (approx. 30% of video length). The projection is reported in Table 2.

**Table 4: Projection of annotation effort**

	Locations	Duration	# Candidate segments	Video Lengths	Annotation effort (3 annotators)
<b>Estimated</b>	1	3 days	16	160 minutes	
<b>Projected</b>	2	15 days	160	1600 minutes	(1600 * 30%)*3 1440 minutes = 24h.





This projection shows that we need 24 hours (3 full working days) of annotations time to conduct the annotation of a pool generated from 4 different observation feeds in a period of 2 weeks. This is a conservative estimation, since we did not consider agenda events and not the entire period of recordings in D3.5 (around 2 months), but it shows with a reasonable effort, we would be able to meet the minimum requirement of annotation.

### 3 Evaluation Results of SMART Applications based on Workshops

#### 3.1 Introduction

This section provides a summary of the two proofs of concept evaluation workshops. In order to evaluate the software, it have been held a total of 35 participants to the workshops devoted to evaluate the security use case and 19 participants to the workshops to evaluate the live news use case.

There were a total of 54 participants, 54 of whom completed the evaluation form.

The next sections report about each workshops results. Workshops are grouped by domain: security and surveillance and live news.

#### 3.2 Security and Surveillance

##### 3.2.1 Santander SMARTCOP evaluation workshop results

<b>Execution date:</b>	15 <sup>th</sup> April 2014
<b>Organization:</b>	Santander police department
<b>Number of participants:</b>	20
<b>Total of questionnaires:</b>	20
<b>Domain:</b>	Security and surveillance
<b>Application:</b>	SMARTCOP

<b>Have you ever heard about Smart Cities or Smart Cities Projects?</b>		
Yes	18	90%
No	1	5%
Not Sure	1	5%
<b>Do you use services similar to the ones presented in the workshops?</b>		
Already Using	6	30%
Considering or evaluating to use	6	30%
Not Using at All	8	40%
<b>Grade the potential usefulness of the demonstrated applications:</b>		
(1) Not at all useful or Irrelevant	1	5%
(2) Useful Under Certain Conditions	2	10%
(3) Somewhat Useful	1	5%
(4) Useful	11	55%
(5) Very useful	5	25%
<b>In case you selected «3» above, please briefly indicate the conditions:</b>		
To improve the quality at work		0%

To improve traffic security		0%
It should be taken into account some legal restrictions about privacy		0%
It is a very interesting project but firstly it requires technical, human and material resources		0%
<b>Identify functionalities of the application most relevant to you and your organization</b>		
Area Surveillance	11	26%
Event Detection	9	21%
Crime Prevention	12	28%
Crowd Panic Management	9	21%
Other - Traffic surveillance	2	5%
<b>Grade (label) from 1-5 the usefulness of the main features of the demo application(s):</b>		
Display and Processing of Data from Social Media	50	24%
Display and Processing of Data from Sensors and Sensor Networks	60	29%
Data Processing / Analytics	51	24%
Data Visualization / Presentation	48	23%
Other Specify: _____		

<b>Describe extra features or functionalities that you would like the presented demo to have</b>
Combine sensors with the traffic lights in the city to optimize traffic

<b>Describe potential uses of the demonstrated applications from your perspective</b>
<p>Recording of potential criminal activities and illegal activities</p> <p>Automatic crash detection base on video and audio</p> <p>Obtaining data on suspicious people across multiple data sources (using of credit cards, public services, accessing to hospitals or public bodies, etc) and showing on a map their physical location.</p> <p>Alternative routes to drivers based on information from sensors and social networks</p> <p>Create a system to inform the patrol cars about different security situations that could be occurring in the city in real time and geolocated on a map.</p> <p>Alert system for locating suspicious vehicles and people through video recording</p> <p>Large Event monitoring</p> <p>Monitoring of critical targets</p> <p>Support decision systems for crisis management</p>

<b>Any other comments</b>
Using sensors to monitor and oversee certain ill people or dangerous criminals (rapists, paedophiles, ...)

### 3.2.2 S3Log evaluation workshop results

Participants to the evaluation workshop in Rome where from different business areas (operators, technicians, business development). We have also invited system integrators and possible customers. One system integrator and one possible customer participate to the workshop.

<b>Execution date:</b>	
<b>Organization:</b>	S3Log / Vitrociset
<b>Number of participants:</b>	15
<b>Total of questionnaires:</b>	15
<b>Domain:</b>	Security and surveillance
<b>Application:</b>	SMARTCOP

<b>Have you ever heard about Smart Cities or Smart Cities Projects?</b>		
Yes	13	87%
No	2	13%
Not Sure	0	0%
<b>Do you use services similar to the ones presented in the workshops?</b>		
Already Using	5	33%
Considering or evaluating to use	6	40%
Not Using at All	4	27%
<b>Grade the potential usefulness of the demonstrated applications:</b>		
(1) Not at all useful or Irrelevant	0	0%
(2) Useful Under Certain Conditions	1	7%
(3) Somewhat Useful	2	13%
(4) Useful	10	67%
(5) Very useful	2	13%
<b>In case you selected «3» above, please briefly indicate the conditions:</b>		
To improve the quality at work		0%
To improve traffic security		0%
It should be taken into account some legal restrictions about privacy		0%
It is a very interesting project but firstly it requires technical, human and material resources		0%
<b>Identify functionalities of the application most relevant to you and your organization</b>		
Area Surveillance	9	26%
Event Detection	10	29%
Crime Prevention	6	18%



Crowd Panic Management	8	24%
Other - Traffic surveillance	1	3%
<b>Grade (label) from 1-5 the usefulness of the main features of the demo application(s):</b>		
Display and Processing of Data from Social Media	55	27%
Display and Processing of Data from Sensors and Sensor Networks	50	24%
Data Processing / Analytics	52	25%
Data Visualization / Presentation	48	23%
Other Specify: _____		

**Describe extra features or functionalities that you would like the presented demo to have**

Real-time presentation  
Discrimination between positive events and events concerning security  
Integration with actuators  
Analytics enriched with users activities  
Increase the capability of low level acoustic sensing  
Monitoring for security within and nearby stadiums

**Describe potential uses of the demonstrated applications from your perspective**

Recognition of specific areas by sound (i.e. ships sirens means we are near to a port)  
Large Event monitoring  
Monitoring of critical targets  
Support decision systems for crisis management  
Crowd analysis with the aim to prevent dangerous situations within the stadium area  
Recognition of specific areas by sound (i.e. ships sirens means we are near to a port)  
Large Event monitoring  
Monitoring of critical targets  
Support decision systems for crisis management

**Any other comments**

Include a slide about potential users and applications  
Login page use a local map instead of world map to locate edge nodes  
Future development could address distance/locations of edge nodes vs. reliability of smart assessment  
Social network information might be useful to discriminate positive/dangerous events  
Threshold should be trained using real data

### 3.3 Live News Evaluation workshop results

#### 3.3.1 Santander LIVE NEWS evaluation workshop results

<b>Execution date:</b>	16 <sup>th</sup> April 2014
<b>Organization:</b>	Neighbourhoods representatives in Santander
<b>Number of participants:</b>	30
<b>Total of questionnaires:</b>	19
<b>Domain:</b>	News and services
<b>Application:</b>	LiveNews

Have you ever heard about Smart Cities or Smart Cities Projects?		
Yes	18	90%
No	1	5%
Not Sure	1	5%
Do you use services similar to the ones presented in the workshops?		
Already Using	10	52%
Considering or evaluating to use	6	31%
Not Using at All	3	15%
Grade the potential usefulness of the demonstrated applications:		
(1) Not at all useful or Irrelevant	0	0%
(2) Useful Under Certain Conditions	2	11%
(3) Somewhat Useful	0	0%
(4) Useful	9	47%
(5) Very useful	8	42%
In case you selected «3» above, please briefly indicate the conditions:		
In some cases it would require police presence		
Information from social networks could introduce a lot of false alarms and news		
Identify functionalities of the application most relevant to you and your organization		
News from social networks	14	26%
Scheduled events in the city	16	30%
Points of interest in the city (restaurants, museums, monuments, ...)	15	28%
Others - Public transport information	2	4%
Others - Cultural activities	1	2%
Others - Information about neighbourhoods	1	2%
Others- Weather forecast	1	2%
Others - Traffic information	1	2%
Others - Security	2	4%

Grade (label) from 1-5 the usefulness of the main features of the demo application(s):		
Display and Processing of Data from Social Media	49	26%
Display and Processing of Data from Sensors and Sensor Networks	56	29%
Data Processing / Analytics	42	22%
Data Visualization / Presentation	43	23%
Other Specify: _____		

Describe extra features or functionalities that you would like the presented demo to have
Integrate mobile phones as data providers

Describe potential uses of the demonstrated applications from your perspective
Show works and road cuts in different parts of the city
Traffic and parking management
Quick communication mechanism with the municipality about traffic, illumination, maintenance
Waste management
Children surveillance in public parks

Any other comments
Extend these technologies to suburbs. Currently are focused in the downtown
It is very important to disseminate these kinds of technologies among the citizens

### 3.3.2 Livenews applications. El País evaluation workshop

<b>Execution date:</b>	30 <sup>th</sup> May 2014
<b>Organization:</b>	Journalist coming from El País and Cinco Días news papers
<b>Number of participants:</b>	10
<b>Total of questionnaires:</b>	8
<b>Domain:</b>	News and services
<b>Application:</b>	Livenews Search View & Livenews City View

1. Have you ever heard about Smart Cities or Smart Cities Projects?		
Yes	7	100%
No	0	0%
Not Sure	0	0%
2. Do you use services similar to the ones presented in the workshops?		
Already Using	3	43%
Considering or evaluating to use	1	14%
Not Using at All	3	43%

3. Grade the potential usefulness of the demonstrated applications:		
(1) Not at all useful or Irrelevant	0	0%
(2) Useful Under Certain Conditions	2	29%
(3) Somewhat Useful	0	0%
(4) Useful	4	57%
(5) Very useful	1	14%
In case you selected «3» above, please briefly indicate the conditions:		
More sensors and evenly distributed		
More data streams		
More control over the noise coming from the social networks		
4. Identify functionalities of the application most relevant to you and your organization		
News from social networks	3.00	
Scheduled events in the city	4.00	
Points of interest in the city (restaurants, museums, monuments, ...)	3.17	
Others	0.00	
5. Grade (label) from 1-5 the usefulness of the main features of the demo application(s):		
Display and Processing of Data from Social Media	2.33	
Display and Processing of Data from Sensors and Sensor Networks	3.50	
Data Processing / Analytics	3.33	
Data Visualization / Presentation	3.17	
Other Specify: _____	0.29	

Describe extra features or functionalities that you would like the presented demo to have
<p>Being able to launch queries combining words, this is key in social networks</p> <p>Personalization and user profile</p> <p>Advertising indexes</p> <p>Predictions of data of commercial use</p> <p>Add more data sources (flickr, instagram, ...) in order to enrich the visual aspect</p> <p>Add more cameras and sensors (all of them)</p> <p>Improve the visualization of the City View demo</p> <p>Include specific information about the characteristics of the different places</p> <p>Include consum tendencies</p>

Describe potential uses of the demonstrated applications from your perspective
<p>Commercialize information of interest for companies and advertisers</p> <p>Information about events for companies of insurance, security, ...</p> <p>To calculate the number of attendants to an event/protestands on a demonstration</p> <p>Detect news on the streets</p>

Touristic information (II)

Follow up of demonstrations (protests) on the streets or sports events

Advertisement on the streets

Touristic information

Measuring the impact of advertising on the streets

Discovering paths for running or biking

External advertising

## 4 Venue recommendation evaluation results

The quantitative results of our user study have already been extensively reported in [SMART-D5.3.2], showing the effectiveness of our approach compared to strong baselines in a variety of settings. However, since we also asked participants for (optional) comments at the end of the study, we report in this section some of the qualitative feedback we received. From all the participants that participated in the user study, 15% entered some comments at the end of the judgments. In this section, we propose to detail and analyse some of them (we did not change orthographic/grammatical errors).

*"Most suggestions were for restaurants at 4pm. I never eat that early hence the negative picks."*

*"Why the time is always 9pm? If the time was different, some suggestions/recommendations would be more relevant. e.g. no way I would visit a museum at 9pm (it is closed anyway!), but I would at 3 or 4pm."*

*"Given the time is 9pm in the scenario a lot of places seem odd, as they would be either closed or unlikely venues at that time."*

*"There were a lot of suggestions for nightclubs for it being 2pm!"*

These comments directly target the (rather low) quality of the recommendations, and especially the time dimension. However, as described in [SMART-D5.3.2], the venues that are presented to the participants come from 5 different algorithms, including a random one. The fact that very irrelevant venues have been recommended is then normal.

*"I am vegetarian - I have that info on Facebook - The suggestions were meat based in terms of restaurants. It was 5pm - a bit too early for the pub, and too late for coffee. The office suggestions were a waste of time :("*

*"Almost every third place was an art gallery / similar place for some reason. I have never been to / liked art galleries. Also other places (like Starbucks) were shown more than once. The "situation" never seemed to change in any substantial way in terms of distance / time of day. Would also be easier to do this survey if some information about the place I was being asked if I would visit was shown on the survey page rather than me having to click a link to visit that places website, which often doesn't make it immediately obvious what type of place it is. Also, took much longer than 5 minutes - about 10-15 for me. That's fine but would be better if the survey said this at the start."*

These were the only two comments that mentioned suggestions that did not fit the interests of the users. Also, the second one shows that the task may be difficult to comprehend at first, as the location and the time of the user do not change when judging several recommendations (we did this to avoid biases, details are given in [SMART-D5.3.2]).

*"It's better to define whether it's summer or winter because given that the situation time is 4 pm, in winter it's already dark while in summer it's still completely sunny. It would affect my decision."*

This was the only comment that mentioned seasonality. We assumed in this study that the recommendations were made at the exact time of the user, on the same day, but we never actually stated this, which apparently caused some confusion.

Finally, we saw some very positive comments, which are self-explanatory:

*"Looks like a good application, I think it could be useful for an future visit !"*

*"Looks good - I see where you're going with it and it would be a great recommendation tool linked with*



*foursquare. One note: 08am should be just 8am”  
“this was pretty cool!”*

## 5 **Conclusions**

This report has illustrated the project's evaluation methodology and has also presented some initial evaluation results regarding the SMART applications.

We have devised a methodology for collecting annotations of high level events that may have occurred in a city. In particular, we have developed a web interface (web application) that allows annotators to describe events that occur in a physical environment (city square and market in our case) upon examining videos collected in the past. In particular, in our methodology we generate a pool of video segments from the various metadata feeds that are likely to be good candidates for an occurrence of an event (e.g. increased crowd in the scene, scheduled events). We conducted a pilot study to validate the interface and estimate the effort needed for performing a large scale annotation. The pilot study has resulted in useful feedback that helped in refining the web annotation interface. In the next version of the deliverable, we will report the results of the large annotation collection.

The most of the suggestions received (those that relate to the objectives of smart) will be used to drive the final cycle of development. There are also indications that will be taken into account for future development of the products.

### 5.1 **Key findings**

With regard to SMARTCOP,

- All respondents gave good ratings to the Poc's
- The participants appreciated the ability to integrate data from social networks with data from physical sensors;
- The feature of using social network to integrate data to refine the security scenario was much appreciated;
- Both workshops devoted to safety have highlighted the need to integrate the application with an actuator system;
- Video recording is mandatory especially for security related use case;

We also noticed that there is a growing interest in fusing data from sensors and social network. The interest of operators and stakeholders during the workshops has demonstrated that there is already a lot of rumours in the foreground, about the possibility of using social network in such a way.

In the case of the LiveNews application, one of the most recurrent comments was the importance of privacy. Citizens were very interested in these technologies as they deal with public data. But they tend to associate these kinds of projects to something like the Big Brother and they are very worried about the use of their private data, especially when we were speaking about video and audio recordings. Hence, it is very important that these kinds of technologies always include safe mechanisms to warranty data security and privacy; otherwise we could find a strong opposition from users to use the applications.

In the side of the journalist the comments was mainly positive. There were a lot of interest about the process of processing and merging the information coming from the social networks. Also people has seen a lot of commercial opportunities for a system like this, many of the related with measuring the performance of the advertising on the streets.



With regards to the venue recommendation application, our user study shows that users (citizens) are generally interested in this kind of applications. However, the user study was limited in the sense that it was conducted in a laboratory setting where the participants were presented with artificial contexts. Therefore, we aim to develop and promote a mobile version of this application [SMART-D6.1] to be able to test its performance with real users on the move. For example, we can conduct A/B testing to compare the effectiveness of recommendations generated by different algorithms. We would also obtain quantitative measures of success from the number of downloads on the Play Store and other statistics describing various recorded user interactions.

## **5.2 Lessons learned**

It is important in these workshops leave people speak and express their opinions. In some cases those opinions are not related directly with the goal of the workshop, but afterwards the information will be extracted from their questionnaires and comments. If we try to impose or direct the workshop from the technical point of view we could find that they are not going to understand anything and probably they are not going to collaborate with you.

For the future evaluation step, the workshop materials could be improved; we received some verbal comments about the need to provide more clear information in the SMARTCOP Case Studies.

## **5.3 Future Evaluation steps**

The initial impression of the SMART project by stakeholders has been very favourable. Applications have been perceived to be interesting on how they address data from multiple sensors, especially on how data from social network is fused with data from physical sensors.

Many specific comments have been addressed by the user. Such comments will be used to improve the software.

For the future evaluation step, planned for the end of the project, workshops will be repeated taking into account suggestions operators made.

Where possible, it will also be carried out a comparative analysis of the results obtained during the two evaluation sessions.

The consortium will also propose evaluation questions to the open source community (questions 5-6 to Those That will download the software).

## **6 References**

[Sparck-Jones1975] K. Sparck Jones and C. van Rijsbergen. *Report on the need for and provision of an ideal" informationretrieval test collection*. British Library Research and Development Report 5266, Computer Laboratory, University of Cambridge, 1975.

[SMART-D6.1.1] SMART FP7 consortium. Deliverable D6.2.1, "Integrated Open Source Framework", 2014.

[SMART-D6.2.1] SMART FP7 consortium. Deliverable D6.1.1, "Integrated Applications of the SMART Multimedia Search Engine", 2012.

## 7 **Appendix I – SMART COP Evaluation Questionnaire**

### **Please fill out contact details**

Name (optional): \_\_\_\_\_

Title (optional): \_\_\_\_\_

Your Organization: \_\_\_\_\_

Business Sector: \_\_\_\_\_

E-mail: \_\_\_\_\_

Phone Number (optional): \_\_\_\_\_

Address (optional): \_\_\_\_\_

Websites (optional): \_\_\_\_\_

### **General information**

Provide a short overview of the activities of your organization

#### **1. Have you ever heard about Smart Cities or Smart Cities Projects?**

1) Yes 2) No 3) Not Sure

#### **2. Do you use services similar to the ones presented in the workshops?**

1) Already Using 2) Considering or evaluating to use 3) Not Using at All

#### **3. Grade the potential usefulness of the demonstrated applications:**

	Not at all useful or Irrelevant	Useful Under Certain Conditions	Somewhat Useful	Useful	Very Useful
Usefulness	1	2	3	4	5

#### **4. In case you selected «3» above, please briefly indicate the conditions:**

#### **5. Identify functionalities of the application most relevant to you and your organization (SECURITY EXAMPLE): - THIS HAS TO BE ADAPTED PER APPLICATION**

Functionality	Tick
Area Surveillance	
Event Detection	
Crime Prevention	
Crowd Panic Management	

Other Specify: _____	
----------------------	--

6. Grade (label) from 1-5 the usefulness of the main features of the demo application(s):

Functionality	Grade
Display and Processing of Data from Social Media	
Display and Processing of Data from Sensors and Sensor Networks	
Data Processing / Analytics	
Data Visualization / Presentation	
Other Specify: _____	

7. Describe extra features or functionalities that you would like the presented demo to have:

---



---



---

8. Describe potential uses of the demonstrated applications from your respective:

---



---



---

9. Any other comments

---



---



---



---

## 8 Appendix I – SMART COP Evaluation Questionnaire

Por favor cumplimente el siguiente cuestionario:

Localidad:

Ocupación:

Nombre y apellidos (opcional):

Correo electrónico (opcional):

Teléfono de contacto (opcional):

### Información general

Describa brevemente las actividades de su organización


10. ¿Ha oído alguna vez hablar sobre las ciudades inteligentes (*Smart Cities*) o proyectos relacionados con el tema?

- |                    |     |
|--------------------|-----|
| 1. Si              | [ ] |
| 2. No              | [ ] |
| 3. No estoy seguro | [ ] |

Marque con una X lo que proceda.

11. ¿Usa o ha usado alguna vez servicios similares a los que se han presentado en esta sesión?

- |                         |     |
|-------------------------|-----|
| 1. Si                   | [ ] |
| 2. No                   | [ ] |
| 3. No, pero me gustaría | [ ] |

Marque con una X lo que proceda.

12. Califique de 1 a 5 el grado de utilidad que le daría a las aplicaciones/tecnologías que se le han mostrado:

	Totalmente inútiles o irrelevantes	Útiles bajo ciertas circunstancias	medianamente útiles	Útil	Muy útiles
Utilidad	1 [ ]	2 [ ]	3 [ ]	4 [ ]	5 [ ]

Marque con una X lo que proceda.

En el caso de haber seleccionado la opción «2», indique brevemente cuales serían esas condiciones:


13. Puntúe de 1 a 5 la relevancia que, desde su punto de vista, tienen las siguientes funcionalidades:

Funcionalidad	Puntuación
Noticias de interés desde las redes sociales	
Eventos planificados en la ciudad	
Información sobre puntos de interés (museos, turismo, restaurantes)	
Otros (especificar): _____	

14. Puntúe de 1 a 5 la utilidad que, desde su punto de vista, tienen las siguientes funcionalidades:

Funcionalidad	Puntuación
Procesado y visualización de datos provenientes de redes sociales	
Procesado y visualización de datos provenientes de sensores	
Análisis y procesamiento de datos	
Presentación y visualización de los datos	
Otros (especificar): _____	

**15. Describa que otras características/funcionalidades adicionales le gustaría que fueran incluidas en las aplicaciones presentadas**


**16. Describa potenciales usos de las aplicaciones que acaba de ver**


**17. Comentarios adicionales**


¡Muchísimas gracias por su colaboración!

Puede encontrar más información del proyecto en nuestra web [www.smartfp7.eu](http://www.smartfp7.eu), o en nuestra cuenta de Twitter [@smartfp7](https://twitter.com/smartfp7).