



WeKnowIt

Emerging, Collective Intelligence for Personal,
Organisational and Social Use

FP7-215453

D7.5.1

Consumer and emergency response use case first evaluation report

Dissemination level	Confidential
Contractual date of delivery	Month 21, 31-12-2009
Actual date of delivery	02-03-2010
Workpackage	WP7, Case studies
Task	T7.1.3 and T7.2.3
Type	Report
Approval Status	Accepted
Version	19
Number of pages	100
Filename	D7_5_1_2010-03-01_v19.docx
Abstract:	<p>This deliverable describes the evaluation results for the first prototypes of the Consumer Travel Scenario and Emergency Response demonstrators.</p> <p>The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.</p>



co-funded by the European Union

History

Version	Date	Reason	Revised by
0.0	02/02/2010	TOC	TID
0.1	03/02/2010	CSG- Travel preparation evaluation	Yahoo
0.4	08/02/2010	CSG – Mobile guidance evaluation	TID
0.6	09/02/2010	ER – Emergency response evaluation	USFD
0.7	10/02/2010	1 st Integration	TID
0.8	10/02/2010	Corrections	Yahoo
0.9	11/02/2010	2 nd Integration – Executive Summary	TID
1.0	11/02/2010	Fixed figure and table caption lengths	Yahoo
1.1	11/02/2010	Figure captions modified	Yahoo
1.2	15/02/2010	Added captions, fixed comments, added appendix	USFD
1.3	15/02/2010	3 rd Integration	TID
1.4	23/02/2010	Corrections	TID
1.5	23/02/2010	Corrections	Yahoo
1.6	23/02/2010	Corrections	USFD
1.8	24/02/2010	4 th Integration	TID
1.9	02/03/2010	Language check, visual improvements	USFD

Author list

Organization	Name	Contact Information	Reason
TID	Manuel Escriche	mev@tid.es	Author
YAHOO	Lluís Garcia Pueyo	lluis@yahoo-inc.com	Author
USFD	Simon Tucker	s.tucker@dcs.shef.ac.uk	Author
KIT	Andreas Sonnenbichler	Andreas.sonnenbichler@kit.edu	Internal Reviewer

Executive Summary

This document reports the results of the evaluations of the first-iteration prototypes developed within the WeKnowIt project. Two demonstrators have been developed – one in the domain of Emergency Response and one in the domain of a group of travellers visiting a new location. Accordingly the evaluation results for each demonstrator are reported separately.

Two groups evaluated the Emergency Response prototype implemented by USFD. Six Emergency Response experts evaluated the access interface to the Emergency Response demonstrator and a smaller group of experts evaluated the upload interface to the demonstrator. Additionally, 12 citizens evaluated both the upload and access interfaces.

The two components of the Consumer Social Group demonstrator were evaluated separately by the corresponding project partner. 15 end users and 21 project partners evaluated the travel preparation module under the supervision of Yahoo!. Four end users and 22 project partners evaluated the mobile guidance module.

The evaluations were split between lab exercises and field trials which were arranged by the respective project partners. Users were given tasks typical of the usage of the respective demonstrators. Following exposure to the demonstrators, the user response to the prototypes was evaluated through questionnaires or discussed in consensus meetings. These materials were analysed in order to identify recommendations and are presented in this report.

The evaluations addressed several dimensions: usability, complexity, efficiency, responsiveness, satisfaction etc. The results of quantitative analyses are additionally presented within this report. Feedback was also collected in a non-structured format in order to get explicit recommendations on how the demonstrators can be improved.

The primary focus of the evaluations was uncovering issues with the current prototypes and identifying areas of the prototypes that could be improved so that project partners can take advantage of these findings when producing the second version of the prototype demonstrators.

Table of Contents

1. Introduction	8
1.1. Goal and scope	8
1.2. Structure of the Document	8
2. Emergency Response Evaluation.....	9
2.1. Demonstrator Evaluations	9
2.1.1. Evaluation Methodology	9
2.1.2. Evaluation Data Collection.....	11
2.2. Expert Evaluations.....	11
2.2.1. Paper Evaluations	11
2.2.2. Access Interface Evaluation.....	12
2.2.3. Upload Interface Evaluation	20
2.3. Citizen Evaluations	21
2.3.1. Paper Evaluation.....	21
2.3.2. Access Evaluation	22
2.3.3. Upload Interface Evaluation	29
2.4. Conclusion.....	30
2.4.1. Recommendations	31
3. Consumer Social Group Evaluation.....	34
3.1. Travel Preparation	34
3.1.1. Experimental Setup.....	35
3.1.2. Usability and Satisfaction	38
3.1.3. Quantitative Analysis.....	41
3.1.4. Qualitative Analysis.....	47
3.1.5. Open Question Feedback	51
3.1.6. Conclusions.....	53
3.2. Mobile Guidance.....	56
3.2.1. Evaluation.....	56
3.2.2. Evaluators	57
3.2.3. Evaluation Process	58
3.2.4. Evaluation Outcome	59
3.2.5. Conclusions.....	67

4. References.....	70
5. Appendix 1: CSG – Travel Preparation -Questionnaires.....	71
6. Appendix 2: CSG – Mobile Guidance – Questionnaire	78
7. Appendix 3: ER – Evaluation Material	80

List of Figures

Figure 1: ER- Mean Score by Condition.....	13
Figure 2: ER - Average time by condition (seconds)	14
Figure 3: ER - I found it easy to understand what was going on in the city	15
Figure 4: ER - I felt I understood what was going on quickly	15
Figure 5: ER - I found it easy to process all the information available to me.....	16
Figure 6: ER - Interface rankings for ER experts	17
Figure 7: ER - Interface rankings for efficiency	17
Figure 8: ER - Interface rankings for Usability	18
Figure 9: ER – Average score by condition	23
Figure 10: ER – Average time taken by condition (seconds)	23
Figure 11: ER - I found it easy to understand what was going on in the city	24
Figure 12: ER - I felt I understood what was going on quickly	25
Figure 13: ER - I found it easy to process all the information available to me.....	25
Figure 14: ER - Overall interface rankings (lower is better)	26
Figure 15: ER – Interface rankings by efficiency (lower is better).....	26
Figure 16: ER – Interface rankings by usability (lower is better)	27
Figure 17: CSG – Web questionnaire	36
Figure 18: CSG - Task 1 step during the evaluation.....	37
Figure 19: CSG - SUS Average for both groups. Normalized values per question	39
Figure 20: CSG - Tool click zones.....	42
Figure 21: CSG - Number of favorites marked across time for <i>External Users</i> group	44
Figure 22: CSG - Number of favorites timeline for the <i>Consortium Members</i> group	44

Figure 23: CSG - Number of favorites across time for *Consortium Members* group 45

Figure 24: CSG - Number of favorites across time for *External Users* group 45

Figure 25: CSG - Number of favorites across time for *Consortium Members* 46

Figure 26: CSG - Number of favorites across time for the *External Users* group 46

Figure 27: CSG - Answers to specific questions about the interface for *Consortium Members* 48

Figure 28: CSG - Answers to specific questions about the interface for *External Users* 48

Figure 29: CSG - Answers to specific questions about the interface for *Consortium Members* 49

Figure 30 : CSG - Answers to specific questions about the interface for *External Users* 50

Figure 31: CSG - Answers to "additional sources needed" question for *Consortium Members* 50

Figure 32: CSG - Mobile guidance: uses cases 56

Figure 33: CSG - Search for POIs and Image gallery 57

Figure 34: CSG – Mobile Guidance deployment..... 57

Figure 35: Summary Evaluation Outcome 68

Figure 36: Mobile upload screen 80

Figure 37: Mobile tagging screen 80

Figure 38: Mobile advertising screen 81

Figure 39: Desktop Homepage 81

Figure 40: Desktop incident view 82

Figure 41: Incident view 82

Figure 42: Incident list 83

Figure 43: Upload; File select 83

Figure 44: Upload; Title and Description 83

Figure 45: Upload; Tag and Localise..... 84

Figure 46: Upload; Complete..... 84

List of Tables

Table 1: SUS Average results per question.....	40
Table 2: Number of favorites marked by user for <i>Consortium Members</i> .	41
Table 3: Number of favorites marked by user for <i>External Users</i>	41
Table 4: Average number of clicks on different parts of the tool for the <i>Consortium Members</i> group	43
Table 5: Average number of clicks on different parts of the tool for the <i>External users</i> group	43
Table 6: First query time for the <i>Consortium Members</i> group	46
Table 7: First query time for the External Users group	47
Table 8: Mean and median user satisfaction/experience for Task 2.....	51
Table 9: Tool improvement suggestions based on aggregated outcome of answers to open questions.....	53
Table 10: CSG – Improvements for the CSG Web Prototype	55

1. Introduction

This deliverable describes the evaluations of the first prototypes implemented both for the Emergency Response Case Study and the Consumer Social Group Case Study. This is done accordingly to section B.1.3.6 – Work package descriptions of WeKnowIt Annex I – “Description of Work”, which describes objectives, task structure and committed deliverables.

This document reports the first outcome from Tasks T7.1.3 and T7.2.3, in which users trials and evaluation are performed for the Emergency Response and Consumer Social Group case studies.

1.1. Goal and scope

The objective of these evaluations is to both ensure that the prototypes that have been developed are functional with regard to their respective use cases and additionally to identify new directions for research and development. Since the demonstrators are designed to address different user needs, each required their own methodology. Since the Consumer Social Group prototype consists of two distinct phases, separate methodologies were used to address each phase.

In this document USFD report the evaluation of the Emergency Response demonstrator [4] from the perspectives of both a group of Emergency Response experts and Citizens. Yahoo! Report on the evaluation of the Consumer Social Group prototype [5] from the perspective of the travel preparation phase. Finally TID report on the evaluation of the CSG prototype from the perspective of the mobile guidance phase.

1.2. Structure of the Document

This document is arranged as follows: after this short Introduction, Section 2 describes the evaluation of the first-iteration prototype implemented for the Emergency Response Case Study; Section 3 describes the evaluation for the Consumer Social Group Case Study in both the preparation and guidance phases.

2. Emergency Response Evaluation

2.1. Demonstrator Evaluations

For the ER first prototype demonstrator three rounds of evaluation were carried out. Two interfaces were evaluated in this process – the access interface, used by citizens and ER experts to explore incidents and the upload interface used to provide information to WeKnowIt. Firstly, before the prototype was finalized, a paper walkthrough was performed with the two target user groups (The ER experts and citizens). Secondly, the desktop access interface was evaluated by a group of ER experts and the upload interface was evaluated by two ER experts, and thirdly, a group of citizens evaluated the upload and access interfaces.

As outlined in D7.3.1 the primary goals of the evaluations were twofold:

- To assess the functionality of the ER first prototype demonstrator from a high level. Thus, to answer the question “Are we doing the right thing?”
- To determine if the user interface and workflows that the interface enables is working correctly. Thus, to answer the question “Are we doing the thing right?”

As this is a first prototype of the ER demonstrator, the evaluation methodology was to focus on lab studies. Such studies are appropriate for assessing the practicality of the goals and implementation of the technology and represent a straightforward method of collecting initial feedback in order to ensure that future field trials will be of maximum benefit. Given the two aims above, the evaluation focused on both whether target users were able to make use of the application to carry out the essential tasks and whether their performance in these tasks was better than the current status quo.

2.1.1. Evaluation Methodology

Paper Evaluations

The paper evaluations were carried out using the system mock-ups described in D7.3.1. To assess the interfaces, paper copies of each screen were printed in colour and to approximate sizes for each interface. These ‘screens’ were then presented to participants in a serial fashion and the user was asked to talk about how they would approach the interface. Specifically users were asked:

1. What actions they thought were appropriate for each screen in order to carry out the task
2. Whether there were any concepts present in the interfaces which they did not understand
3. What actions they would take to continue the task

Participants were encouraged to think-aloud and to discuss what their opinions were about each screen and the concepts that were present at each stage. The screens presented to the users are given in the appendix.

Application Evaluations

To assess the functionality of the first prototype demonstrator, users were asked to carry out the tasks they would typically be involved in and assess the interface on the basis of these tasks. Thus, for the citizens, they were asked to upload an image and to use the access interface to determine information related to an incident they had previously witnessed in the city.

The ER experts were asked to primarily evaluate the access interface, although a secondary evaluation assessed the functionality of the upload interface. The task for the ER experts was to gather information about and distinguish between three events which were co-occurring in the city.

In both the access cases a further goal of the evaluation was to assess how the Collective Intelligence implemented by WeKnowIt impacted on the ability of citizens and ER experts to make sense of the information provided to them. Therefore, evaluations were carried out with three different interfaces in order to measure this impact.

The first interface reflected the type of data currently received during emergency incidents – information gathered from incoming calls from the general public. Thus, in this condition participants were given access to the comments alone. These comments were provided on pieces of paper, which the participants were free to use as they wished to learn about the incident. In the second case, participants were given the raw data provided to the WKI system in the form of a simple web interface which allowed them to see the images and comments in a serial manner. Finally participants were given access to the same information that had been uploaded to the WeKnowIt interface.

In both cases participants were instructed to process information quickly and were informed that they need not see every piece of information before making their decision but could use each interface as they wished. Prior to starting each condition, a short tutorial was given on each interface and the facilitator was on hand during the experiments to ensure that the participant knew how each interface functioned. Participants used a different interface for each dataset to ensure that their answers were not biased.

The evaluations also examined the use of the upload portion of the application. Whilst the upload process is designed to function on a mobile phone, these initial tests were carried out on a desktop version of the interface. The primary reason for this is that it was desirable to assess the upload process from a high level before identifying specific usability issues present in the upload interface. Thus, the focus of the assessment was

how people tag and comment on images rather than whether there were issues with a specific mobile operating system.

2.1.2. Evaluation Data Collection

The data collection phase was carried out by a single person who took multiple images of the synthetic incidents with a camera phone. These images were then manually localised and comments were also generated manually using public descriptions of incidents as guidelines. The data followed the schema described in D7.3.1 of synthetic images of three different instances of the same type of incident co-occurring in the city of Sheffield. The three incidents were fires, floods and tree fells. In addition to these three incidents a further set of data was collected relating to explosion incidents that was used for demonstration and training purposes only.

The data was collected as described with two alterations. Firstly, the medium level fire was described explicitly as being a toaster fire as a larger fire was considered as highly severe by ER experts due to its location. Secondly, the location of the severe tree fell was moved to a major road since this would have a much higher impact on the severity level from a local authority perspective. The collection procedure was as described except that in order to ensure a consistency to the comments, these were generated by a single person.

The data set used in the evaluations was assessed by ER experts as being representative of the type of information received during an emergency. Primarily the data was designed to reflect the types of uncertainty seen in data. Thus, the comments were ambiguous and often referred to colloquial locations. A description of the specific data collected is given in the appendix.

2.2. Expert Evaluations

Three evaluations were carried out with ER experts. Prior to the completion of the first prototype demonstrator a paper-based interface evaluation was carried out. This evaluation set out to determine if the concepts and designs proposed for the first prototype demonstrator were meaningful to both user groups. Following the implementation of the first prototype, the access interface was evaluated by a group of ER experts and a smaller group also assessed the upload interface.

2.2.1. Paper Evaluations

Upload Interface

The ER experts were asked to consider themselves as being a Forward Liaison Officer who had been sent to the scene of an emergency to gather information for the main ER team. They were asked to simulate the

process of uploading an image to the main ER system. The simulation began with the first screen of the upload process (the file selection screen).

The primary problems with the upload process were that the interfaces were considered to be too cluttered and the ER users often found it difficult to correctly identify the next step when using the interfaces. Considering each screen in turn:

Mobile Home Page

The terminology used on this screen was unclear and the correct action for uploading an image was difficult to determine. One user correctly selected "Information Upload" whereas another clicked on "Resource Management".

Mobile Upload Page

The ER users were unable to determine the meaning of some of the icons present in the upload screen and were additionally confused as to which button they should press to perform the upload operation.

Mobile Tagging Page

The ER experts were clear as to the purpose of the tagging screen and what each piece of information referred to. They were also clear as to what tags they would apply to the images. There was some confusion as to how to continue on from this screen. One ER expert thought that this would be the conclusion of the upload process and so didn't realise that they had to continue to the next screen.

Mobile Advertising Page

The mobile advertising page was clear to the ER experts although the necessity of the screen was questioned.

Access Interface

The Access interface was more straightforward and simpler to use. The ER experts were able to use the paper interface to explore information relating to a single event. Again the experts noted problems with some of the icons used in the interface and were unsure as to what some of the icons meant. In addition, the experts had an issue with the use of the term "event". The ER experts had a different view of the term event and generally used the term "incident" to refer to emergencies occurring in the city.

2.2.2. Access Interface Evaluation

The access interface evaluation was carried out as part of an Emergency Response workshop organised by the Sheffield City Council. The workshop involved 6 participants from Emergency Response organisations across the UK. The average amount of ER experience held by the participants

was 17 years and the average amount of computer experience was 15 years.

The participants generally make use of paper based technologies when organising and dealing with incidents, although some use the AIMS ATLAS¹ system for visualising emergency information.

Overall Scores and Timings

To measure the performance (as opposed to the efficiency) of the participants, a simple scoring system was used. The answers given by the participants were compared to the information used to generate the evaluation data on a one-to-one basis and given a score of 1 if the participant gave the same answer and 0 otherwise. The scoring was made for each incident, for the location, severity and type of incident. Thus each answer was scored out of 9 (9 meaning that the answer was completely correct and 0 meaning that the answer was completely incorrect). The scoring was carried out by a single person without reference to the condition to prevent biasing.

The figure below shows the mean scores per condition for the access evaluations.

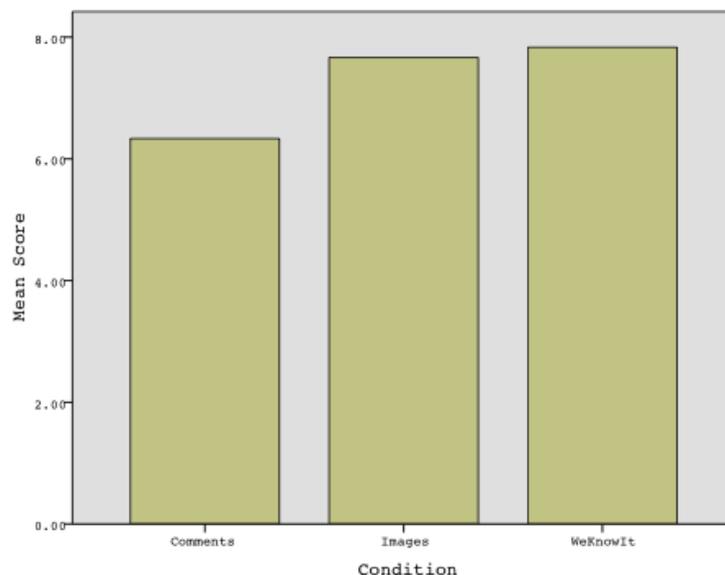


Figure 1: ER- Mean Score by Condition

As the figure shows, the mean scores were generally high. As expected the score arising from the comments condition alone were lowest and the scores achieved in the WeKnowIt condition were marginally higher than those for the images (7.8 for WeKnowIt and 7.7 for Images). Given that the information present in the two interfaces is the same, this result is not surprising. The scores for the comments are lower; the near equality of

¹ <http://www.atlasops.com/>

the scores for the images and WeKnowIt conditions is reflective of the power that the image has for this user group.

The average time taken for each condition is shown below.

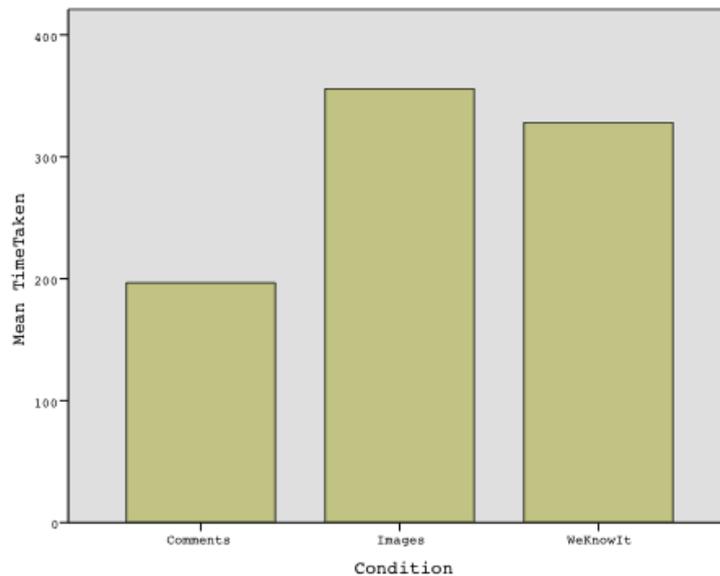


Figure 2: ER - Average time by condition (seconds)

Users are fastest with comments since there is a small amount of data to deal with. In the images condition users may spend a lot of time serially accessing the images, in an effort to piece together -pictures of the same incident. Users are marginally quicker using the WeKnowIt interface (320 seconds as opposed to 350 seconds). Given that this was the first exposure the experts had to the interface, this is a good indication that the interface is assisting the users in gaining an understanding of the incidents without being too cumbersome to use.

Overall these results indicate that there is an interaction between the accuracy of the interpretation, the time taken to reach this interpretation and the amount of information available. In the comments case, users were quick to interpret the data but this interpretation was relatively poor. When given access to images, users took longer but the quality of their responses improved – with the WeKnowIt interface (considering that this is the first exposure that users have had) users were marginally quicker still and gave better answers overall.

Post-Condition Questionnaire Results

After each condition, the users were asked to fill out a brief questionnaire which assessed their feelings about the interface they had just used. The results of these questionnaires are presented below (the questionnaire itself is given in the appendix). In the figures below the Likert scores have

been converted into numbers, where 1 is “Agree Strongly” and 5 is “Disagree Strongly” – thus a lower score indicates more agreement.

I found it easy to understand what was going on in the city

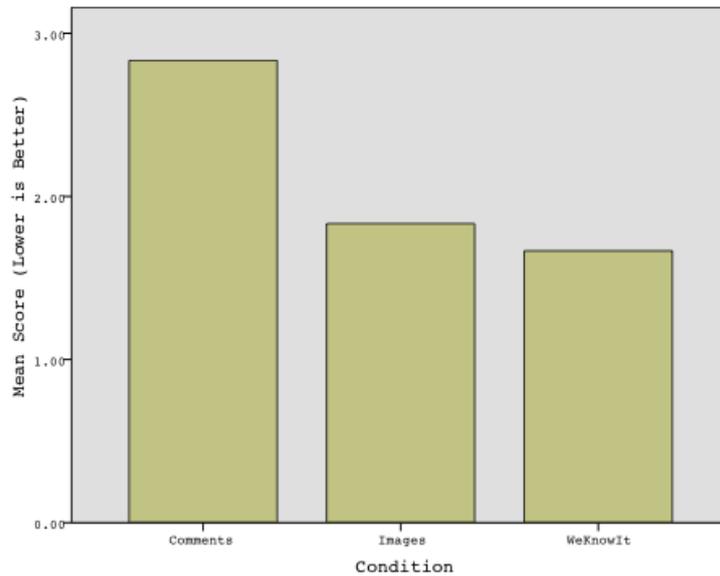


Figure 3: ER - I found it easy to understand what was going on in the city

Thus, users felt that they could understand what was going on better in the WeKnowIt and the Images conditions, although the improvement in understanding in the WeKnowIt condition is marginal. This result is reflective of the scores data presented above.

I felt I understood what was going on quickly

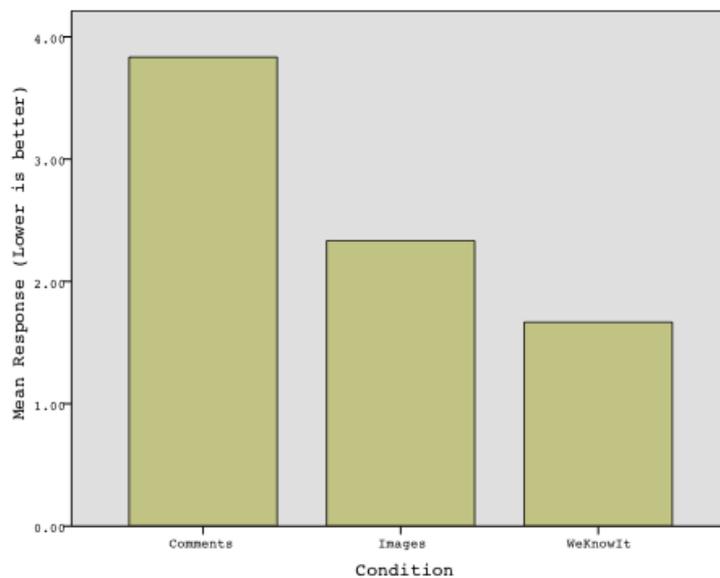


Figure 4: ER - I felt I understood what was going on quickly

Here users clearly found that the WeKnowIt interface enabled them to get a rapid understanding of the incidents occurring in the city. Though this conflicts with the numerical data, this could be an indication that users did not feel that they had built an accurate understanding of the incidents using just the comments.

I found it easy to process all the information available to me

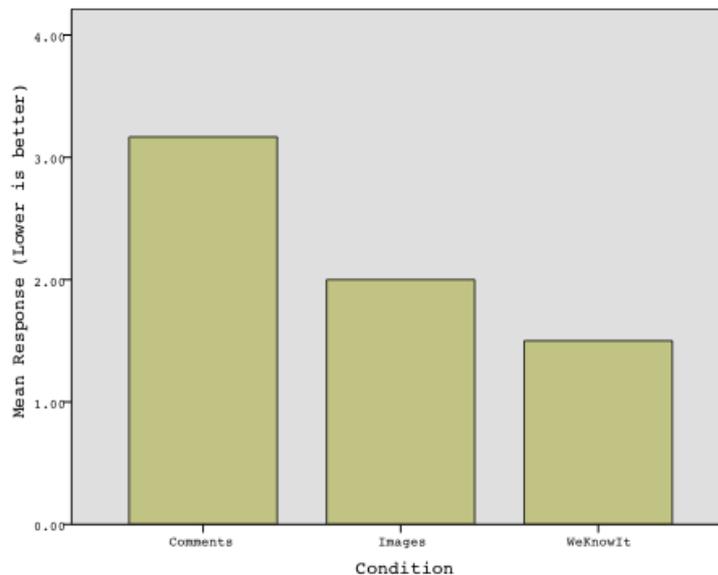


Figure 5: ER - I found it easy to process all the information available to me

Again, users found the WeKnowIt interface enabled them to process the information more easily than with the other two interfaces.

Overall these responses were generally positive towards WeKnowIt. The ER experts felt that the application enabled them to quickly and easily gather an understanding of what was going on in the city and allowed them to process all the information that was available to them.

Interface Rankings

In addition to the subjective ratings given above, the experts were asked to rank the interfaces in order of preference on three scales: Overall, In terms of efficiency and in terms of ease of use. The rankings are presented below.

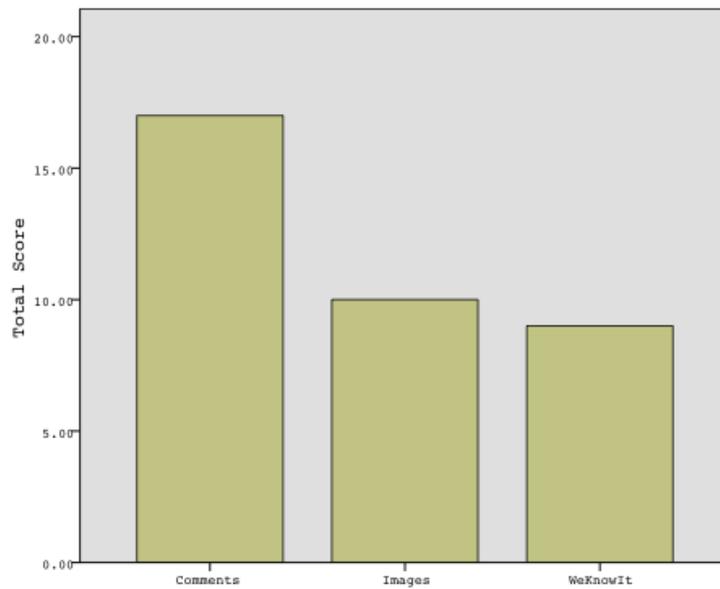


Figure 6: ER - Interface rankings for ER experts

Overall the experts score the WeKnowIt interface slightly better than the Images interface. That the rating is only marginal is representative of the significant added value afforded through the presentation of images and also indicative that there is room for improvement within the application.

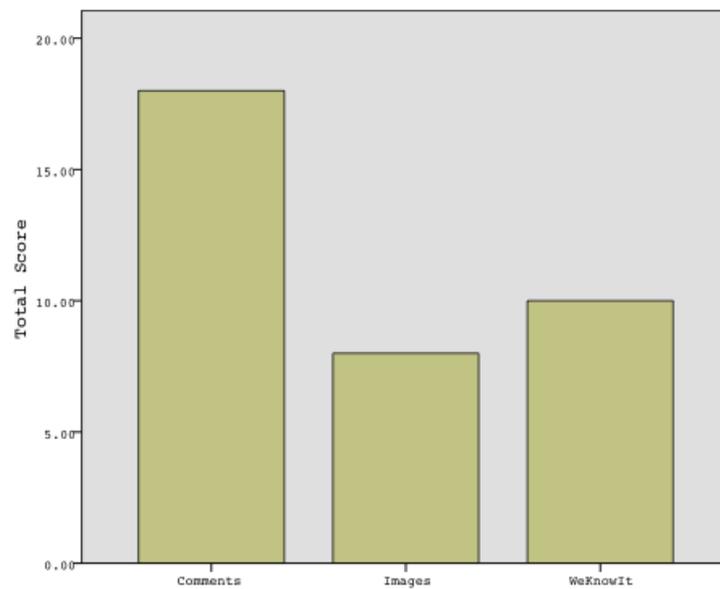


Figure 7: ER - Interface rankings for efficiency

Here the experts rated the images interface as being more efficient than WKI, both being more efficient than the comments interface. This is surprising, considering the positive comments and the scores result given above. Again, however this could be indicative of the value placed in images by the ER experts.

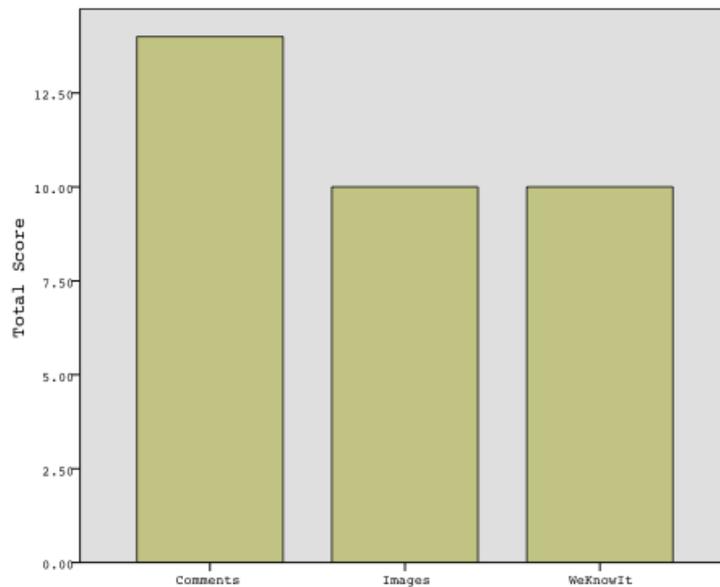


Figure 8: ER - Interface rankings for Usability

The WeKnowIt interface is rated as being as easy to use as the image interface. Considering the simplicity of the images interface and the complexity of the WeKnowIt interface this is a positive result and indicates that the usability of the application has been pitched correctly.

Standard Questionnaire Results

As with the CSG evaluation a standard usability questionnaire was used to assess the reaction of the ER experts to the WeKnowIt interface. The results of this questionnaire are presented below.

Question	Mean Score
I think that I would like to use this system frequently	2.0
I found the system straightforward and simple	2.0
I found the system easy to use	2.0
I would need the support of a technical person to be able to use this system	4.0
I found the various functions that the system performed were well integrated	2.3
I thought that there was too much inconsistency in the system	3.5
I think I could learn to use this system very quickly	1.3
I found the system cumbersome to use	4.2
I felt confident using the system	2.2
I needed to learn a lot of things before I could get going with this system	3.9

Overall these results are positive and, considering the relatively small amount of training the experts had with the interface these results

indicate that the experts responded favourably to the interface and that it was easy to use. The full responses to this questionnaire are given in the Appendix.

Focused Questionnaire Results

In addition to the generic usability questionnaire, a further set of questions were used to probe some specific properties of the WeKnowIt interface.

Question	Mean Score
I think that WeKnowIt added value to my understanding of the incident	1.6
I found the interface easy to use	2.0
I thought the information was presented clearly	2.3
I could use the interface to focus on information I was interested in	1.8
The interface helped me deal with information overload	2.3

Again the response from the ER experts is positive. The relatively low response to questions 2 and 3, however, indicates that the next iterations of the ER demonstrator (which will present more information to the user than this early version) should focus on clearly presenting relevant information, and add in some elements to deal with the information overload problem.

Comments

Finally we asked for some free text responses to a small number of questions.

What do you think was the most useful part of the WeKnowIt interface?

The respondents identified the ability to focus on particular areas of the city as being beneficial: *"being able to select the area I wanted to focus on"*. The value of images to the ER experts is evident here as 5 out of the 6 participants highlighted an aspect of having access to images as being beneficial.

How do you think the WeKnowIt interface would affect your decisions?

Three respondents highlighted the use of the application in improving the initial assessment of incidents and the ability to rapidly understand a given situation: *"I think it would aid my decision making by being able to grasp understanding of an issue more quickly"*. Again, the experts cited the use of images as aiding their understanding: *"[WeKnowIt] gives you a visual picture of the event. Better informed."*

What do you think was missing from the WeKnowIt interface?

One respondent identified that the ability to analyse the images in more detail would be useful and another felt that the ability to interact with time would be a useful addition. Overall, participants were largely positive about the application.

What three pieces of information would you like to see but couldn't?

Two of the respondents indicated that they would also like to see weather conditions as it would allow them to predict the evolution of certain incidents better. There were also comments about how the system would be integrated with the other systems that are used by the respective Emergency Response units.

2.2.3. Upload Interface Evaluation

Whilst the access interface is of more relevance to ER experts, the upload portion of the interface was also evaluated by two participants. The experts were asked to use the upload interface to add an image to WKI and were then asked some questions about this process.

Timings

On average the experts took 155 seconds to upload an image. Note that the majority of this time was taken up with localising the image on the map – scrolling and zooming the map to find an appropriate place for the map marker.

Questionnaire Results

The table below gives the results given for the questionnaire.

Question	Mean Score
It was clear from the interface how I should upload an image	1.5
It was simple to tag an image	1.5
I understood why I would tag an image	2.0
I could understand what I was able to do with the mobile interface	2.0
I think I could use this interface in a time-critical situation	1.5

The results were generally positive although the responses to the question about *why* an image should be tagged were rated slightly below the other questions (apart from the affordances of the mobile interface).

Comments

The questionnaire also enabled the participants to leave free text responses to what properties of the upload process they liked and what properties they disliked. Both cited the upload process as being simple and straightforward. However, one said that they had had difficulty localising the image using the map as it was too small and the other

wasn't sure how to add tags to the images as there was not specific button to achieve this.

2.3. Citizen Evaluations

The evaluation with citizens followed the same schema as that for the ER experts. Twelve citizens participated in the evaluations with a mean age of 36 and average of 16 years of computer experience. The majority of the participants had a mobile phone (83%) but only three owned a web-enabled smart-phone. The majority of the citizens used internet services to discover information about incidents that had occurred in the city.

The citizen evaluations followed the same procedure and used the same dataset as the ER expert evaluations. In the objective assessments, however, the citizens were given a brief piece of information about one of the incidents and asked to use the interfaces to identify exactly what happened. The piece of information was deliberately vague – e.g. “You saw a traffic jam on Mappin Street” and thus the citizens had to identify the relevant information from the irrelevant information. In addition to this, some of the questionnaire questions were altered to reflect the different perspective that the citizens would bring to the interface.

2.3.1. Paper Evaluation

The paper interfaces were evaluated with three representative citizens. As with the ER evaluations the citizens were asked

Upload Interface

The Citizens were asked to imagine that they were at the scene of an incident and wished to upload an image to WKI to assist in planning for the incident. Furthermore, they used the access interface to recover information about the incident.

As with the ER evaluations, the primary problems with the upload process was that the interface was confusing and it was difficult to identify the actions required to proceed through the upload process. In addition, the notion of tagging was confusing to the users and it was unclear to them what terms they would tag an image with. Considering each screen in turn:

Mobile Home Page

Again, the terminology present on this screen was confusing and similar selection problems were seen with regard to starting the upload process.

Mobile Upload Page

The citizens were unsure which button would actually perform the upload operation and they were additionally confused as to the difference

between the green arrow and the green arrow with the star (see Annex 3 for the corresponding screen shots).

Mobile Tagging Page

The citizen users were unable to define the precise meaning of priority in the context of the upload process as a whole. As with the ER users, the citizens were also unsure what to tag the images with, given that the image was being associated directly with a location and an event.

Mobile Advertising Page

The purpose of this screen was reasonably clear although the notion of rating of information was unclear to the citizens since it was their personal information that was being uploaded.

Access Interface

As with the ER users, the citizens found the access interface more straightforward to use and could easily use the interface to locate information relevant to their needs. Again, however, there was some confusion over the meaning of some of the icons and the relevance of some of the terminology.

2.3.2. Access Evaluation

As stated above the access evaluation was slightly modified for the citizen evaluations – the citizens were asked to gather information about the medium severity incident after being given a small piece of information about it.

Overall Scores and Timings

To score the results of the experiment, the score sheets were compared to the correct answers, and points were deducted if the respondent included any of the unrelated events or did not provide any additional information to that which they were given. Citizens were also asked to rate the severity of the incident and points were deducted if this part of the score sheet was incorrect. For the citizens, since they were only asked to evaluate the status of a single incident the resulting score is out of 3 rather than 9.

The figure below shows the average score achieved with each interface for the citizens.

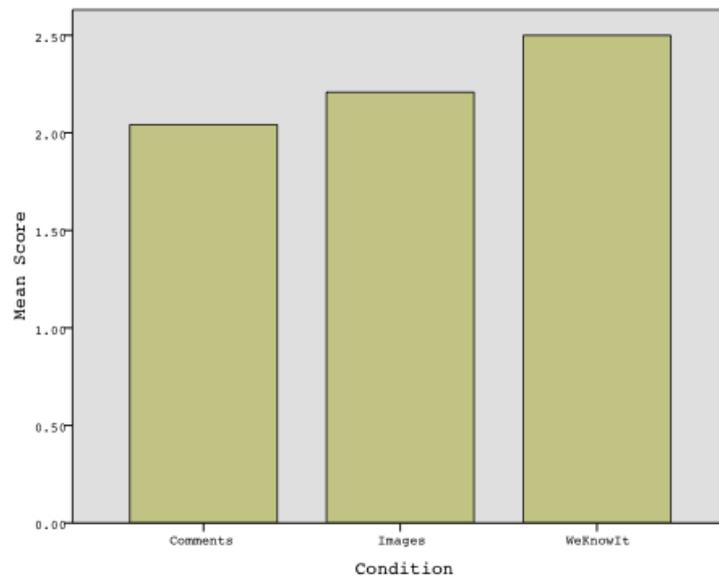


Figure 9: ER – Average score by condition

As the figure shows, the difference in performance between the conditions was less pronounced than that for the ER experts, though this is largely due to the simpler task that the citizen was asked to perform. Again, however, it can be seen that the WeKnowIt application receives the highest score overall and that the citizens found that the images added some value to their interpretation of the incident. The overall time taken organised by condition is shown below.

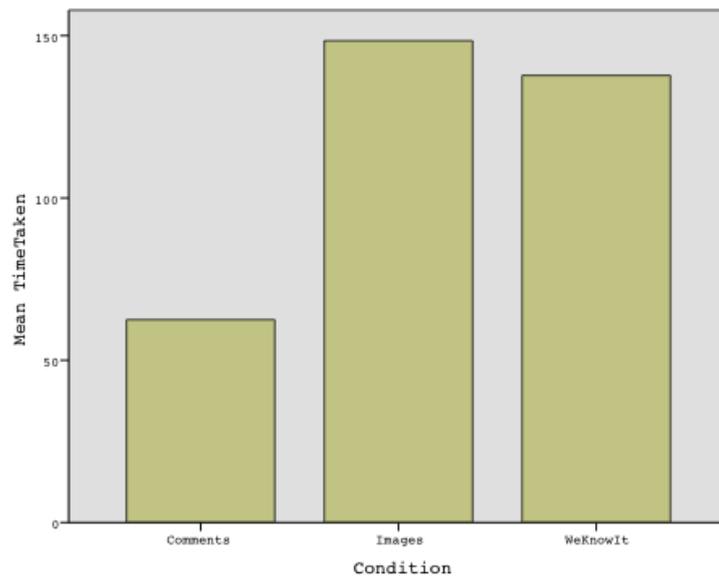


Figure 10: ER – Average time taken by condition (seconds)

Again this figure indicates that users of the Comments interface were able to process information fastest of all but that users of the WeKnowIt interface were marginally quicker than those using the Images condition.

Again, considering the relative complexity of the WeKnowIt interface, this can be seen as a positive result.

Thus, the results from the citizens indicate a similar trend to those from the ER experts. Users are able to reach better conclusions faster using the WeKnowIt interface than if they had access to the raw information.

Post-Condition Questionnaire Results

The same questions were posed to the citizens after each condition. As above, the Likert scores were converted to numerical values (lower values indicate more agreement).

I found it easy to understand what was going on in the city

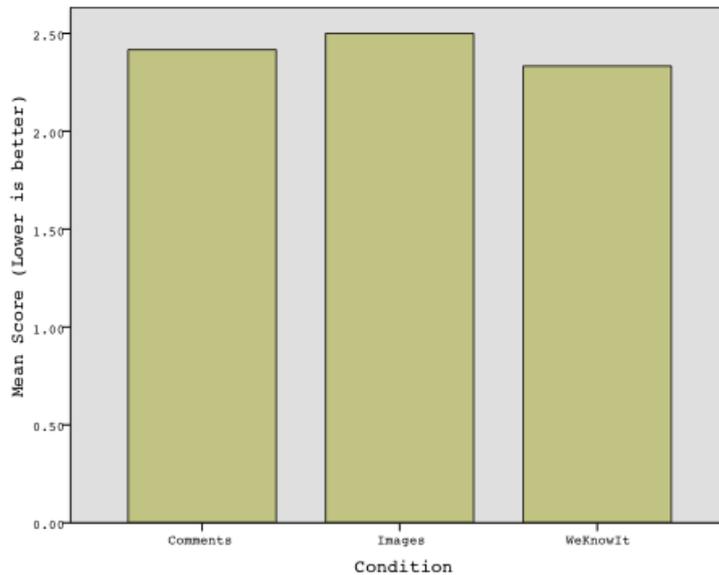


Figure 11: ER - I found it easy to understand what was going on in the city

Thus, there was little effect of the interface on the participants ability to understand what was going on in the city. The WeKnowIt interface is marginally superior to the others but again, the relative simplicity of the task is reflected in the ability of the participant to build an understanding of the incidents.

I felt I understood what was going on quickly

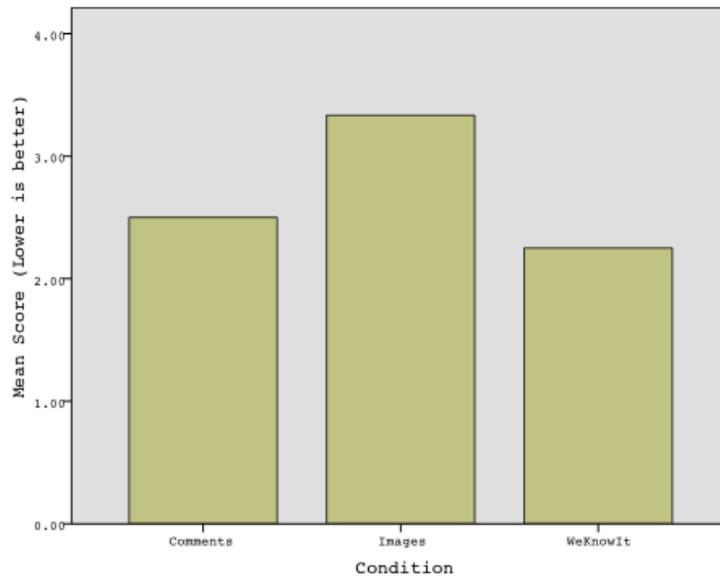


Figure 12: ER - I felt I understood what was going on quickly

As with the ER experts, participants felt that they could process information quickest with the WeKnowIt interface and that the images were the slowest condition to build an understanding.

I found it easy to process all the information available to me

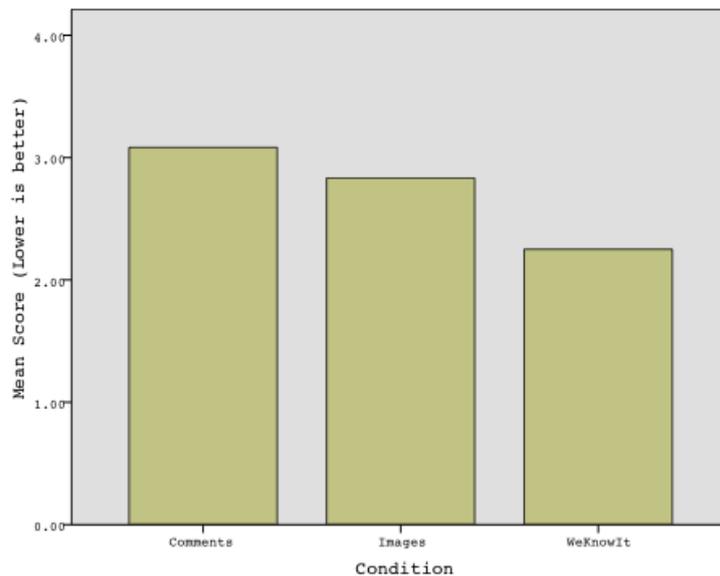


Figure 13: ER - I found it easy to process all the information available to me

Again, the citizens felt that the WeKnowIt interface allowed them to process the information more easily than the other two interfaces. Overall, as with the ER experts, these results are reasonably positive towards the WeKnowIt application.

Interface Rankings

The citizens were also asked to rank the interfaces in order of preference on three scales: Overall, In terms of efficiency and in terms of ease of use. The rankings are presented below.

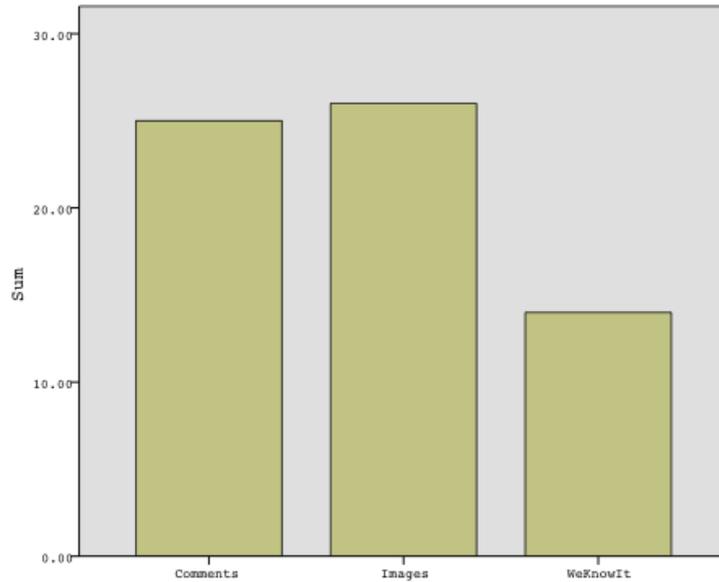


Figure 14: ER - Overall interface rankings (lower is better)

Overall the citizens show a clear preference towards the WeKnowIt interface. This could be due to the ability of the interface to allow citizens to rapidly identify the incident of interest whilst ignoring unrelated information.

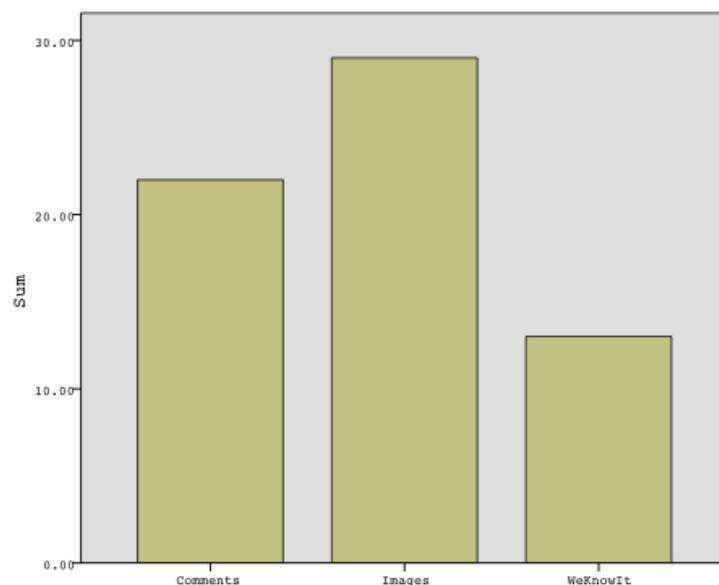


Figure 15: ER – Interface rankings by efficiency (lower is better)

Again the citizens preferred the WeKnowIt application and considered it to be the most efficient overall. The images received a relatively low score;

considering the task it can be argued that images are of less value to citizens than the information of what occurred during the incident.

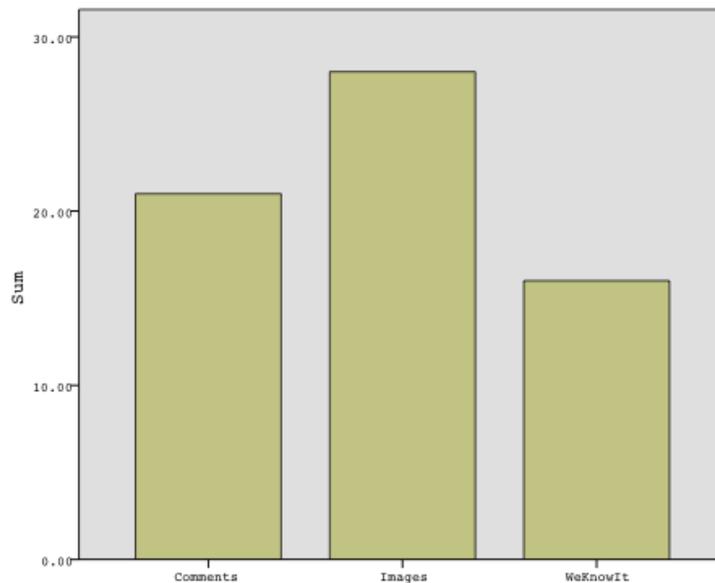


Figure 16: ER – Interface rankings by usability (lower is better)

Again the citizens rated the WeKnowIt interface as being easier to use than the comments and images.

Standard Questionnaire Results

The same standard usability questionnaire was used to assess the citizen response.

Question	Mean Score
I think that I would like to use this system frequently	2.4
I found the system straightforward and simple	2.1
I found the system easy to use	2.1
I would need the support of a technical person to be able to use this system	4.2
I found the various functions that the system performed were well integrated	2.0
I thought that there was too much inconsistency in the system	3.7
I think I could learn to use this system very quickly	1.5
I found the system cumbersome to use	4.0
I felt confident using the system	2.3
I needed to learn a lot of things before I could get going with this system	4.3

Again these results are relatively positive. The relatively ambivalent score for the inconsistency question could be a result of the problems with the tag filtering mechanism.

Focused Questionnaire Results

In addition to the generic usability questionnaire a further set of questions were used to probe some specific properties of the WeKnowIt interface.

Question	Mean Score
I think that WeKnowIt added value to my understanding of the incident	1.9
I found the interface easy to use	2.0
I thought the information was presented clearly	2.2
I could use the interface to focus on information I was interested in	1.9
The interface helped me deal with information overload	2.3

Again the responses to these questions were positive overall and suggest that the interface for WeKnowIt is good. One question that should be addressed in future versions of the ER demonstrator is to improve the presentation of the information and to consider methods of addressing the information overload problem.

Comments

The same open-ended questions were asked of citizens than of the ER experts.

What do you think was the most useful part of the WeKnowIt interface?

The citizens largely appreciated the connection between the location, images and the textual information afforded by the WeKnowIt interface. The citizens were also more appreciative of the use of the tag cloud, presumably since it was more discriminative in the given task and therefore allowed the citizens to filter the available information down to something more manageable.

How do you think the WeKnowIt interface would affect your decisions?

Although a marginal question for the citizens, several of the participants highlighted that they could use the WeKnowIt interface in order to "avoid difficult or dangerous situations". In addition to this, citizens highlighted that the application could help when they needed to make decisions whether to travel or not.

What do you think was missing from the WeKnowIt interface?

Some respondents indicated that they would like to have a more immediate preview of incident information when they hovered their mouse over the icons. In addition, citizens expressed a preference for a bigger map and for a search function. One participant also confirmed that the application should have some indication of the objectivity of the information.

What three pieces of information would you like to see but couldn't?

As before, several respondents indicated that they would like to be able to hover over an icon in order to preview information that it contains. Other users stated that they would like to see information relating to diversions (both car and bus). There were also some minor comments relating to how the information was presented to the user – for example, representing dates in a more human readable format.

2.3.3. Upload Interface Evaluation

To evaluate the upload functionality, the users were asked to upload a single image of a fire. The users were then left to attempt the upload procedure without assistance although, if requested, assistance was given. Generally, however, assistance was not required and the users were able to complete the full upload process without help.

Timings

On average the citizens took 105 seconds to upload an image. Again, the majority of this time was spent on the tagging and localizing screen – specifically attempting to localise the image was perceived to be a complex task for the users.

Questionnaire Results

The table below gives the results given for the questionnaire.

Question	Mean Score
It was clear from the interface how I should upload an image	1.5
It was simple to tag an image	1.8
I understood why I would tag an image	2.4
I could understand what I was able to do with the mobile interface	2.8
I think I could use this interface in a time-critical situation	2.7

Again the responses were largely positive although there is still room for improvement. Primarily, the process of tagging and localising the image should be much simpler and clearer given the scores given here. Given the relatively low score for the fourth question, the purpose of uploading an image should be clarified.

Comments

The respondents were largely positive about the upload process and said that it was generally straightforward and simple to use. The negative comments focused on the process of localising the image which was described as being "too fiddly" and the tagging process which one respondent said "I was not sure if my tag had been correctly identified"

There was also an issue of what types of tags and descriptions were being applied to the image. In terms of the description, the text used was largely a high level description of the image including the type of incident (e.g. fire) and some information about the location where the image was taken. A similar effect was seen in the tags applied to each image – two types of tags were applied to images: descriptive or positional. Thus, a common tag entered (if it was not already automatically suggested) was “Fire” and if a tag was added beyond this, it generally related to the location (street or place) where the image was taken.

2.4. Conclusion

This document presented the results of the evaluation of the first prototype of the Emergency Response demonstrator for the WeKnowIt project. The first prototype of the demonstrator supports the upload of images using an intelligent tagging system and also the exploration of the uploaded data using a visualisation which supports the faceted filtering of properties in the WKI DS.

Both processes were evaluated by two groups of users, representing the target group of users for the WKI ER application: Citizens and ER experts. In both cases the users were asked to upload an image using the upload interface. The time taken, title, description and tags applied to the images were noted and the response to the interface was gathered.

To evaluate the access interface, the WeKnowIt was compared to two other interfaces: one which reflected a minimal amount of unstructured information as a collection of statements about a number of incidents and one which represented the raw data presented in the WeKnowIt interface. The dataset comprised of images and comments relating to three incidents which were co-occurring in the city of Sheffield. The number of images and comments collected corresponding to the severity of the incident. ER experts were then asked to use the interfaces to build an understanding of each of the incidents, their location and their corresponding location. The citizen participants had access to the same data and interfaces but were instead asked to find out more information about one of the incidents (one with medium severity) on the basis of a small amount of information.

Overall both groups of participants were positive about the WeKnowIt application. They were able to complete the tasks required of them both in terms of uploading information, accessing information and also interpreting and interacting with the information in order to build an understanding of what incidents were happening in the city.

2.4.1. Recommendations

With regard to the access interface the following recommendations are made:

- The tag filtering interface was counter-intuitive. Currently, selection of a tag causes it to be struck out in the tag cloud and the corresponding images are removed from the map display. This process has two problems. Firstly, the evaluations have shown that this process is counter-intuitive for most of the participants. The majority of the participants that accessed the information through the tag cloud expected that selecting a term in the tag cloud would highlight the corresponding information in the map. Additionally this meant that if an image is attached to multiple tags, the user must spend time discovering what these tags are in order for information to be displayed.
- None of the participants made use of the evidence list part of the interface (see the screenshots in the appendix). This could be because of its location within the interface but could also be a perception that the incident list doesn't add any value. In either case the interface component should be restructured in order to draw the benefits out for the user.
- The ER experts indicated that they would like to be able to analyse the images in more detail than is provided by the small version presented on screen. This request is in conflict with the general feeling from both groups that the map should be the focus of the interface and should be larger than it currently is. A solution to this problem may be just to offer the ability to view the image as an overlay over the current interface.
- A further problem with the interface as it currently stands is that the connection between the various information displays (e.g. the tag and time displays) is relatively weak. For example, the tag widget displays tags that are present in the WKI-DS regardless of the state of the other components. Some of the participants felt that changes in one widget should be reflected in the other widgets. Whilst this wasn't a serious issue with the current round of evaluations because the amount of data was such that the information overload didn't present itself, in future rounds and versions of the demonstrator this will become a serious problem.
- With regard to the integration of the WeKnowIt system with other systems that are currently used by Emergency Response teams it can be argued that this can be addressed through allowing the other system read access to the WKI-DS. The alternative approach, of finding a way for the WeKnowIt application to access other systems is out of the scope of the project but any barriers to allowing this process to occur should be avoided.

With regard to the upload interface:

- Overall the process was relatively quick and understandable for the users and there was no major sticking point during the upload process which is a positive result of the evaluations.
- The process of tagging, however, whilst not confusing to the users, was often misinterpreted by the participants. There was a confound between the tags that the users wished to add and the tags that were suggested by the intelligent WKI services. In addition to this since the tags looked like buttons participants often clicked on the tags expected this to mean that the tag would be applied to the image. Thus this process should be clarified.
- The process of localising the image was also identified by participants as being a little fiddly. In future versions of the upload process the process of localising the image should be simplified. This can be done either through making more use of the intelligent services or by providing an alternate means of stating the location of the image.
- With reference to the score and time results presented above, whilst participants were marginally faster and more accurate when using the WeKnowIt application, the second prototype should focus on ensuring that these figures are improved upon. This can be done by both addressing the issues outlined above but also by improving the responsiveness of the interface and focusing on technology that will assist the ER experts at identifying critical information and distinguishing between incidents (for example, clustering technologies).

Given the findings of both evaluations, a further recommendation can be made. Both the paper and interface evaluations identified a potential problem with the notion of tagging information within the ER demonstrator. The participants in the paper evaluation and those in the upload evaluation were clear about the concept of tagging but were often unsure about the actual tags that they should apply to the information they were uploading.

Typically the tags that were applied to the information that was uploaded related to information that would be present in the information metadata. So, typically there was a tag relating to the class of incident (e.g. fire) and if there was another tag added by the user it generally related to the location of the incident. Because of this the added value of tags is relatively small since the information contained in the tags is also stored in other metadata related to the image.

Therefore, it is recommended that future versions of the ER demonstrator either move away from tags as a means of annotating uploaded information or attempt to focus the tags towards terms which are complementary rather than overlapping of the extra metadata which already annotate the image.

In addition to the problem with tags, future versions of the demonstrator should seek to address the issue of information overload and focus on services which aim to address this problem. Although the amount of data used in the evaluation was relatively small, some of the responses from the evaluation participants and the results from the timing information indicated that had there been a larger amount of data then users would have had a harder job of discriminating between different incidents. It is currently unclear whether the disconnected filtering approach taking in the user interface of the demonstrator will be sufficient to assist the user in dealing with information overload. The second version of the prototype should therefore focus on integrating services which support the clustering and categorisation of information on facets that the filtering widgets cannot account for.

In addition to evaluating the prototype demonstrator some consideration should be given to an assessment of the evaluation protocol. Whilst for the first prototype the amount of data and procedure used in the evaluation protocol was sufficient for further evaluations the size of the dataset should grow considerably. An additional drawback of the evaluation process is that the information conveyed in the image is locational rather than informative – i.e. the images do not actually contain pictures of fires etc. For future evaluations a solution to this problem should be sought in order to balance the information provided in each condition. In addition to this having incidents occur at different times would allow some further exploration of how users identify information they are interested in.

3. Consumer Social Group Evaluation

The prototype for this case study is composed of two different parts, according to the scenario: travel preparation and mobile guidance. It follows two subsections where each part is assessed by the partner who implemented it, Yahoo for the Travel Preparation and TID for the mobile guidance, following its own methodology, adapted to the needs of each part.

3.1. Travel Preparation

We present the first evaluation for the CSG Web Prototype². The tool has been evaluated by two different groups of users with different profiles:

On one hand, 21 consortium members took part in the evaluation process. Given the nature of its members and his relationship with WeKnowIt, the results of this evaluation are considered as an extra evaluation. This group will be named *Consortium members* in the document.

On the other hand 15 users, not related to the project or Yahoo!, also took part in the evaluation of the prototype. We consider the results gathered from the evaluation of these users as the evaluation results. This group will be referred to *External users* in the document.

It is worth mentioning that the *Consortium members* group did the evaluation before the *External users* group. Although questions and tasks were the same for both groups, the first group experience made us reconsider the mandatory time users needed for the tasks. For *External users* the time was reduced and, therefore, results for questions related to the tasks are not comparable. For these tasks there was a minimum time set that a user should spend on the task. There was no maximum time.

Regarding to the tool used, a generic Survey platform was implemented with the aim to provide an easy way to perform the evaluation from remote locations, and in order to reuse it for future improvements or for different tools of the project.

The evaluation had different goals:

1. Regarding to WP7:
 - a. Evaluate the WP7 prototype in terms of *usability* and *satisfaction*.
 - b. Obtain user explicit feedback to be aware of possible improvements and demands for future versions.
 - c. Obtain user implicit feedback to observe the usage of the tool and thus improve and refine workflows and use cases.
2. Regarding to WP1:

² Publicly available at <http://weknowit.research.yahoo.com/csg>

- a. Collect data for the personalization service implemented by Telefonica I+D.
3. Regarding to WP3:
 - a. Evaluate the effectiveness of media intelligence services and correct integration in the prototype.
4. Regarding to the consortium:
 - a. Invite consortium members to use the tool to generate interest and possible use cases in order to integrate current and future services.

The document presents the entire evaluation procedure: we start describing the *experimental setup*. Regarding to results, we present the *usability* and *satisfaction* grade achieved, and then we comment on the *quantitative* and *qualitative* results obtained. Finally we analyze the *open question feedback* obtained that is derived into a set of improvements for the tool.

3.1.1. Experimental Setup

A web survey platform was implemented to allow users to evaluate the prototype in a remote location. As gathered data was supposed to be used in the WP1 Personalization step a login system was required.

The evaluation access process had the following steps:

1. Users received an email with instructions on the evaluation. The email was sent to around 50 users for the *Consortium members* group and to around 20 users for the *External users* group. 21 *Consortium members* and 15 *External users* performed the evaluation.
2. Immediately after, users were sent an automatically generated email with a link to the tool and instructions on how to login with a randomly generated username.
3. Once logged in, users needed to accept a Terms of Service Agreement and they were presented the tool.
4. All issues found with the authentication or with the access to the evaluation site were solved by request.

The evaluation process was divided in the following sections:

1. Demographics questionnaire: Standard questions to get demographic data for the participants. (See Appendix 1: CSG – Travel Preparation -Questionnaires)
2. Task 1: Common task to all users to get familiarized with the tool and to collect data for WP1 personalization experiment. It consisted of the following steps:
 - a. Task pre questionnaire
 - b. Task: Prepare a trip to Barcelona (See Task 1)
 - c. Task post questionnaire

3. Task 2: The second task consisted on two different tasks randomly assigned to users. The objective was to test the search capabilities of the tool and test the travel search services provided by WP3.
 - a. Task pre questionnaire
 - b. Task: Find landmarks in a city (See Task 2a and Task 2b)
 - c. Task post questionnaire
4. Post questionnaire with:
 - a. Usability and Satisfaction questionnaire: Standard questionnaire to test how usable the presented prototype is.
 - b. Tool specific questions: Questions about the tool layout and the general interface to get feedback and to establish goals for future improvement.

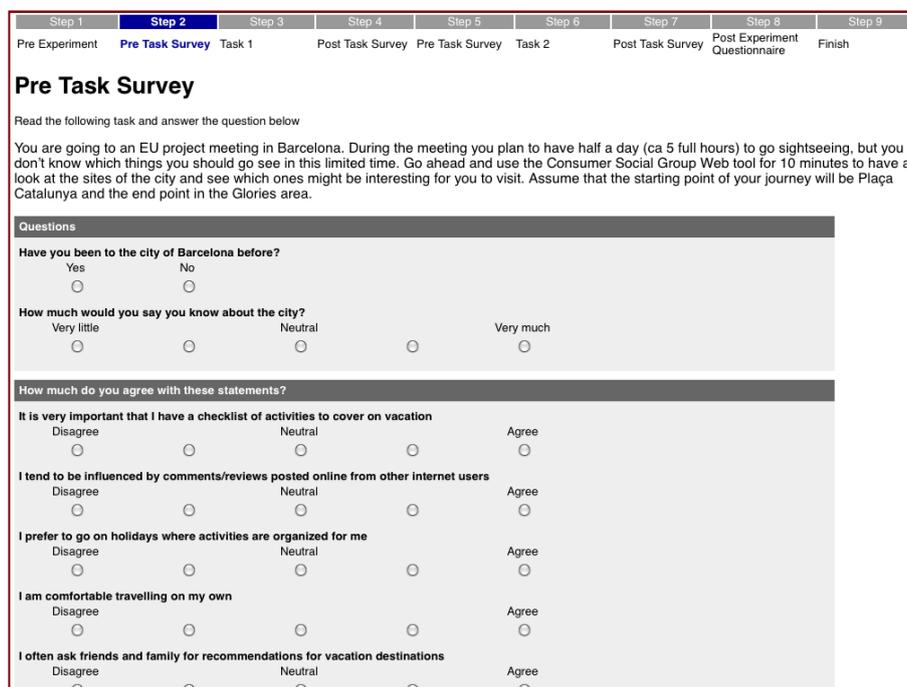


Figure 17: CSG – Web questionnaire

Besides explicit answers given by users using the questionnaires, all user activity (clicks and queries) was logged for mining purposes.

All questionnaires are listed in Appendix 1: CSG – Travel Preparation - Questionnaires.

3.1.1.1. Task 1

Task 1 is about planning a trip to Barcelona and was performed by all users.

Description

You are going to an EU project meeting in Barcelona. During the meeting you plan to have half a day (ca 5 full hours) to go sightseeing, but you don't know which things you should go see in this limited time. Go ahead and use the Consumer Social Group Web prototype for 10 minutes to have

a look at the sites of the city and see which ones might be interesting for you to visit. Assume that the starting point of your journey will be Plaça Catalunya and the end point in the Glories area.

Expected Output

A list of landmarks in Barcelona that could be used for both personalization experiment and the evaluation of the CSG desktop prototype.

Time limit

The participants were asked to spend at least 5 minutes on solving this task and were given a reminder to continue to the next step after 10 minutes.

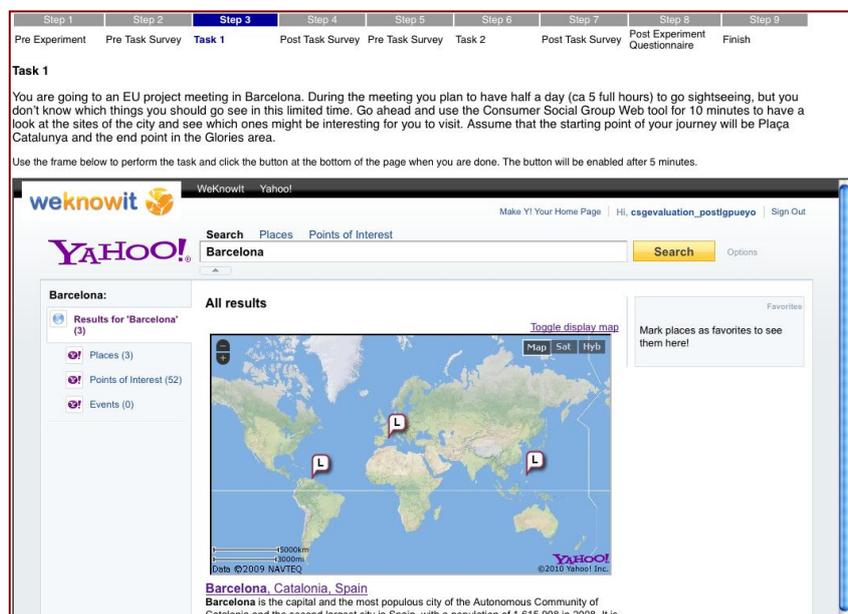


Figure 18: CSG - Task 1 step during the evaluation

3.1.1.2. Task 2a

Task 2a involved finding historical sightings in Rome and was performed by half of the participants.

Description

As an amateur historian, you are collecting all information about historic sightings in Rome, Italy that were built before the 16th century, and which you'll visit during your upcoming visit to Rome. Once you've find such a sighting please mark the corresponding page as a favorite.

Expected Output

A list of POIs in Rome, describing sightings in Rome, built before 1500. This list can be used to measure the task performance in terms of precision and recall.

Time limit

The participants were asked to spend at least 5 minutes on solving this task and were given a hint to continue to the next step after 10 minutes.

3.1.1.3. Task 2b

Task 2 involved finding information about museums in London and was performed by half of the participants.

Description

You are planning on a cultural trip to London. You want to know about which museums are in the city. Once you've find such a sighting please mark the corresponding page as a favorite.

Expected Output

A list of museums in London. This list can be used to measure the task performance in terms of precision and recall.

Time limit

The participants were asked to spend at least 5 minutes on solving this task and were given a hint to move to the next step after 10 minutes.

3.1.2. Usability and Satisfaction

21 *Consortium members* and 15 *External users* evaluated the prototype using the online survey platform.

The SUS Usability and Satisfaction questionnaire³ was used, which allows for the usage of a standard methodology, which outputs a score on a scale from 0 to 100, the greater the score the more usability and satisfaction. Questions for the SUS questionnaire can be found in Appendix 1: CSG – Travel Preparation -Questionnaires.

This questionnaire consists of 10 questions divided in *Positive questions* and *Negative questions* presented alternatively. Users could choose a value from the following 5-point scale:

Option	Value
Strongly disagree	1
Disagree	2
Neutral	3
Agree	4
Strongly agree	5

Positive questions are defined as those where the bigger the value given by the user the better the usability.

³ Brooke, J. (1996) SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & A. L. McClelland (eds.) Usability Evaluation in Industry. London: Taylor and Francis.

Negative questions are defined as those where the lower the value given by the user the better the usability.

Questions 1,3,5,7 and 9 are *Positive questions* while questions 2,4,6,8 and 10 are *Negative questions*. To be able to compare results from both types of questions we will use normalization. Normalizing, in this case, consists of transforming results to homogenize the scores to a single scale where big numbers indicate good usability. In our case, normalization has been done by $NormalizedScore = 6 - response$ for *Negative questions*.

The average SUS score obtained was 68.92 over 100 for *Consortium Members* and 60.66 over 100 for *External users*, which we interpret as a decent overall usability score with room for improvement for the 2nd round of prototypes. The difference in the scores for the two user groups is not statistically significant ($P > 0.85$ using two-tailed unpaired t-test).

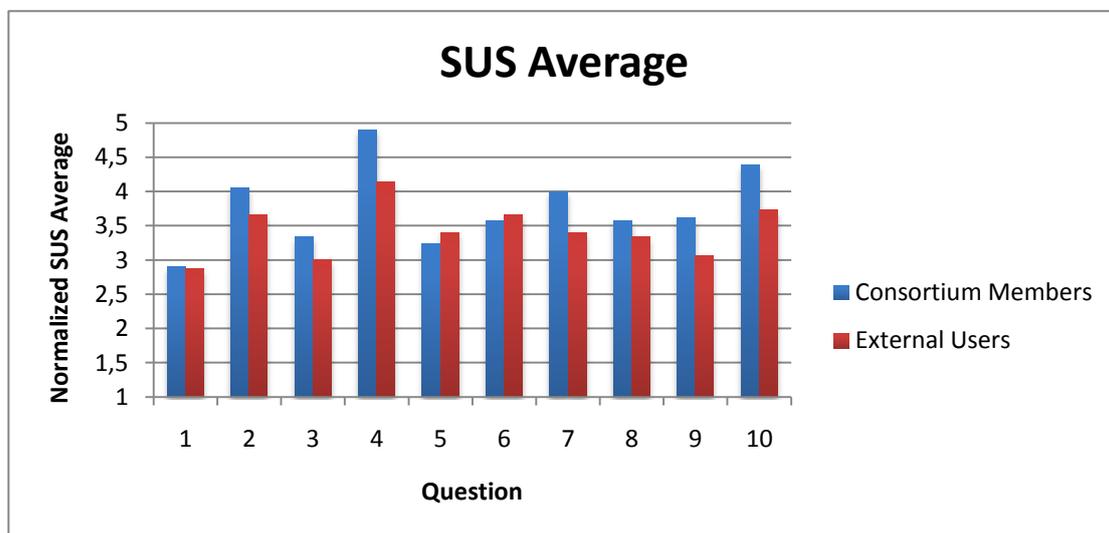


Figure 19: CSG - SUS Average for both groups. Normalized values per question. Bars in Figure 19 show a similar trend over all questions for both groups. These allow for a joint analysis for the results.

The analysis per individual questions shows the following:

1. When normalizing the values for all questions, we can observe that all results except one are over or equal to 3.0 (in a scale from 1 to 5), which gives the indication that the overall usability experience is good and that there are neither positive nor negative peaks in specific questions.
2. Questions 4 and 10 get the best scores. These were the questions:
 - a. (Question 4) I think that I would need the support of a technical person to be able to use this system
 - b. (Question 10) I needed to learn a lot of things before I could get going with this system

In terms of usage, the tool covers users' expectations, as it is easy and intuitive to use.

3. Question 1, a *Positive question*, gets the lowest score (2.8):

- a. (Question 1) I think that I would like to use this system frequently

Given that this is a prototype travel site, the impression is that users know other tools that they are already using and that the CSG Prototype does not offer enough value to introduce a change.

- 4. Questions 3 and 5 do not reach a very high score when compared to others, and are the second and third worse results obtained after Question 1 for the *Consortium Members* group. These were the questions:

- a. (Question 3) I thought the system was easy to use
- b. (Question 5) I found the various functions in this system were well integrated

These results give the indication that the workflow presented was fairly complex and that it could be improved.

- 5. For the *External users* group, question 9 gets poorer performance when comparing to the *Consortium Members* group. This could be explained by the fact that *External users* are not used to test research prototypes, and they expect final products, usable in real world problems. This fact introduces a trend that will be seen during all the evaluation results for this group.

		Consortium members		External users	
		Avg	nAvg	Avg	nAvg
1	I think that I would like to use this system frequently	2.90	2.90	2.87	2.87
2	I found the system unnecessarily complex	1.95	4.05	2.33	3.67
3	I thought the system was easy to use	3.33	3.33	3.00	3.00
4	I think that I would need the support of a technical person to be able to use this system	1.10	4.90	1.87	4.13
5	I found the various functions in this system were well integrated	3.24	3.24	3.40	3.40
6	I thought there was too much inconsistency in this system	2.43	3.57	2.33	3.67
7	I would imagine that most people would learn to use this system very quickly	4.00	4.00	3.40	3.40
8	I found the system very cumbersome to use	2.43	3.57	2.67	3.33
9	I felt very confident using the system	3.62	3.62	3.07	3.07
10	I needed to learn a lot of things before I could get going with this system	1.62	4.38	2.27	3.73

Table 1: SUS Average results per question.

3.1.3. Quantitative Analysis

3.1.3.1. Number of Favorites

The web application provides a way to store currently viewed locations as favorites. A click on the link over the map displayed sends the current location to the server, stored for the current user and presented on the right hand side for a later quick access. Favorites can be deleted or accessed at any point, as they are always shown on the favorites list

The number of favorites marked during the tasks can show the quality of results presented to users. Although in Task 1 users were not explicitly asked to store favorites, they used the functionality, which indicates that it is a useful functionality.

Although the second task explicitly asked for favorite marking, there are users that didn't mark anything and didn't report any technical issue. Table 2: Number of favorites marked by user for *Consortium Members* shows the results without counting those users.

	Task 1	Task 2	Task 2a	Task 2b
Max	10.00	12.00	12.00	6.00
Min	0.00	2.00	2.00	2.00
Average	2.95	5.61	7.44	3.78
StDev	2.75	2.95	3.02	1.23

Table 2: Number of favorites marked by user for *Consortium Members*

	Task 1	Task 2	Task 2a	Task 2b
Max	8.00	7.00	7.00	5.00
Min	0.00	3.00	4.00	3.00
Average	1.44	4.78	5.50	4.20
StDev	2.37	1.31	1.50	0.75

Table 3: Number of favorites marked by user for *External Users*

Regarding the *Consortium Members*, the different numbers for Task 2a and 2b could be explained by the difficulty of finding results for both tasks. While Task 2a had a large number of possible results ("historic sightings in Rome") Task 2b was a narrower topic with less possible or, at least, known results ("museums in London"). We will come back to this point later when we look at user satisfaction/experience for Task2 later in this document, where we see that both Task 2a and 2b had 'neutral' satisfaction and Task 2b was perceived as more difficult than Task 2a.

For *External Users*, we see that in general fewer favorites were marked and that the trend for different tasks is the same: Task 2b got less favorites marked. Although groups had different time constraints to complete the tasks, we see that the trend is similar. A reduction on the task mandatory time resulted in less favorites marked.

3.1.3.2. Zone click analysis

User queries and user clicks were logged for mining purposes. In the tool workflow, when a Place⁴ is presented 5 points of interest are presented on the left hand side. The points of interest are ranked using the output of the WP3 *GetPOIs* service. Therefore, the click analysis aggregated over different users can give us an indication of how well the points presented help the user navigate a given Place.

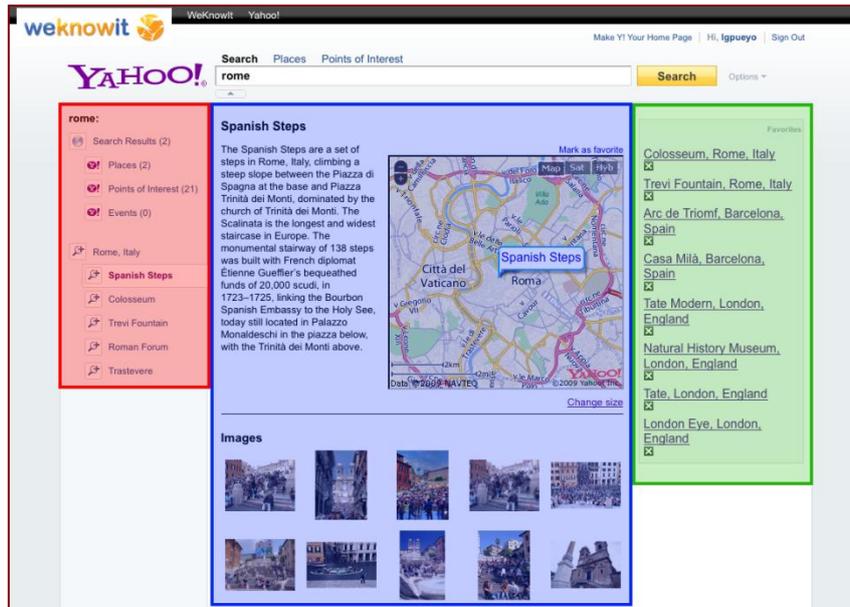


Figure 20: CSG - Tool click zones.

The following tables present the relation between clicks on components in the page versus clicks on the suggestion bar. In Figure 20 the application zones are presented with a colored layer: The red zone corresponds to the suggestion list while the blue zone compresses other logged components in the page, including maps, lists of Points of Interest and search results. Components in the page include the following items compressed in the blue zone of Figure 20:

- Points of interest of a given location
- Points marked in the map

In general, users use the tool in two ways:

1. To explore different Places (high ratio)
2. To explore one single Place (low ratio)

The nature of the tasks promoted the second behavior, as they were focused on exploration of points of interest belonging to a single Place. This behavior is not ideal for testing how good suggestions are⁵, as

⁴ We use the term *Place* to refer to a country, state or city and *POI* to refer to a point of interest within a Place.

⁵ This is done separately from the CSG Web prototype.

probably an hypothetical user would like to explore many points of interest, not only a summarized (suggested) list.

	Zone	Average	Ratio
Task1	Page	9.05	2.30
	Suggestion	3.94	
Task2A	Page	6.45	1.34
	Suggestion	4.80	
Task2B	Page	8.70	4.35
	Suggestion	2.00	
Task2	Page	7.52	1.88
	Suggestion	4.00	

Table 4: Average number of clicks on different parts of the tool for the *Consortium Members* group

	Zone	Average	Ratio
Task1	Page	6.85	2.28
	Suggestion	3.00	
Task2A	Page	5.14	2.14
	Suggestion	2.40	
Task2B	Page	4.33	0.72
	Suggestion	6.00	
Task2	Page	4.77	1.59
	Suggestion	3.00	

Table 5: Average number of clicks on different parts of the tool for the *External users* group

Nevertheless some insights can be extracted from the data:

- The number of clicks in general is not high in all of the cases. This is probably caused by the fact that many users did find their objective with a few clicks. A Precision/Recall study can be conclusive to say if they found what they were looking. This is outside the scope of the evaluation of this prototype.
- Task 1 was intended to be a familiarization step. It is then normal that the ratio is higher than for Task 2A
- Task 2B shows a really high ratio for *Consortium Members* and a high ratio for *External Users*, meaning that users used the suggestion bar much less than the other page components. This could be due to the nature of the task, as users knew exactly what to look for and they didn't need suggestions.
- In general, we observe low ratios, and this phenomenon indicates that suggestions are useful for the user.

Regarding to the evaluation process for this section, clicks on the Map control were logged as result list elements, and it would be interesting to have this data in the context of clicks comparison. This issue has been marked as future work for next evaluation.

3.1.3.3. Number of Favorites progress on time

Using the click logs, we could also observe the progress of number of favorites marked per user with respect to time for Task 2.

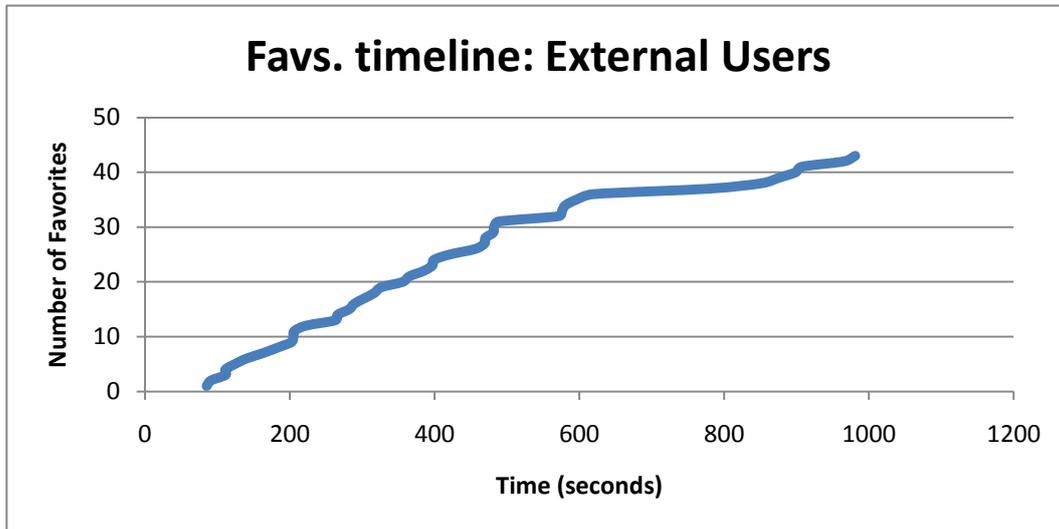


Figure 21: CSG - Number of favorites marked across time for *External Users* group

The analysis for all users raised the presence of some outliers that didn't allow seeing the real progress. These outliers correspond to users that stopped marking favorites during a period of time or marked its first favorite long time before they entered to the task page (more than 1 hour).

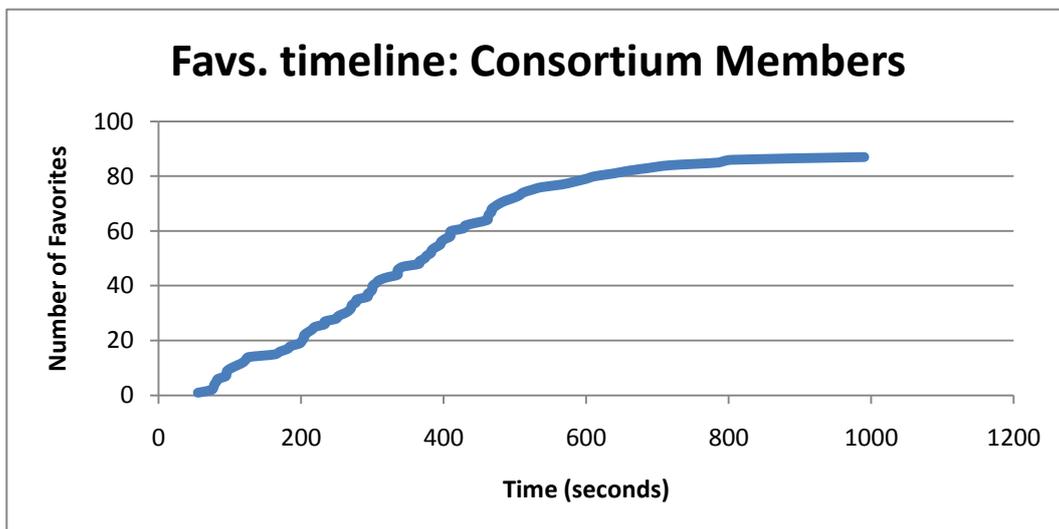


Figure 22: CSG - Number of favorites timeline for the *Consortium Members* group

Figure 21 and Figure 22 show the number of favorites on time without outliers. The number of favorites grows constantly until it becomes constant, as the users might have found all favorites. It is worth to compare the same behavior removing the *First Favorite Time*, which is the time when a user marks its first favorite. We can consider this time as

when the user finished reading the task description and actually started to complete the task.

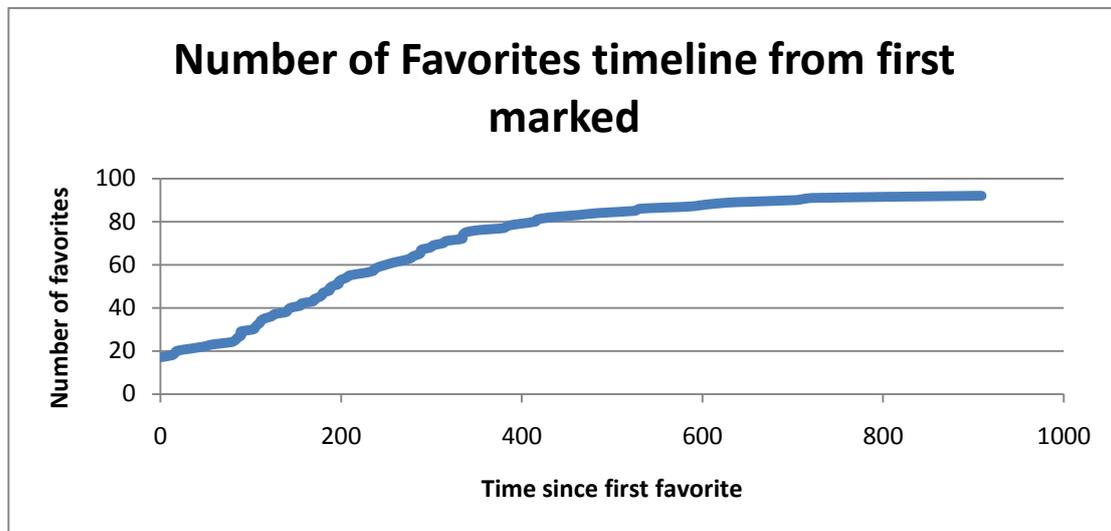


Figure 23: CSG - Number of favorites across time for *Consortium Members* group

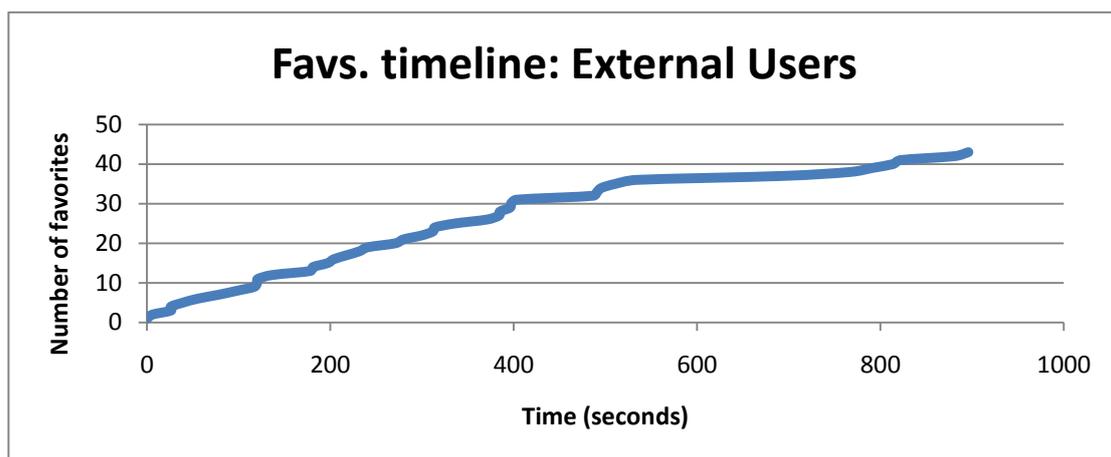


Figure 24: CSG - Number of favorites across time for *External Users* group

This is what is shown in Figure **23** and **24**, where we see that for the *Consortium Members* group in 5 minutes (300 seconds) users have found 75% of the favorites marked (69 out of 92) while in 8 minutes (487 seconds) *External Users* have found 75% (32 out of 43). For *Consortium Members*, 90% of the favorites were found after 7.5 minutes (461 seconds). For *External Users*, 90% of the favorites were found after 14 minutes (811 seconds).

The differences between groups could be explained by the fact that *External Users* are not used to perform evaluation tasks. Besides that we can observe in Figure **26** that two *External Users* used a long time but the major part of the users chose their favorites in less than 5 minutes (300 seconds).

If we take marking of favorites as one of the measures of how good the application is for search, we should take these numbers as target to

improve in next releases of the tool. Besides, the number of favorites marked is another metric we can establish as baseline.

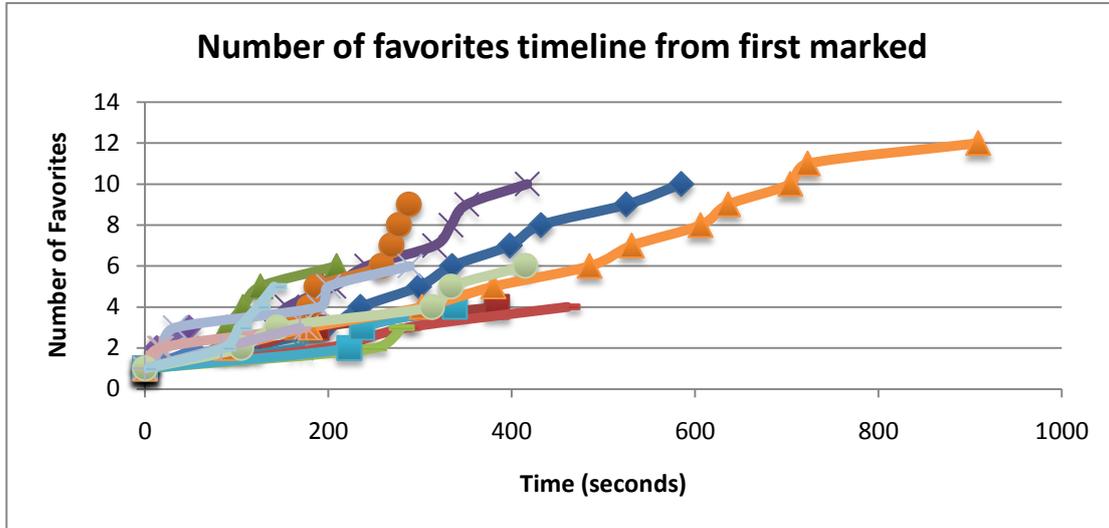


Figure 25: CSG - Number of favorites across time for *Consortium Members*

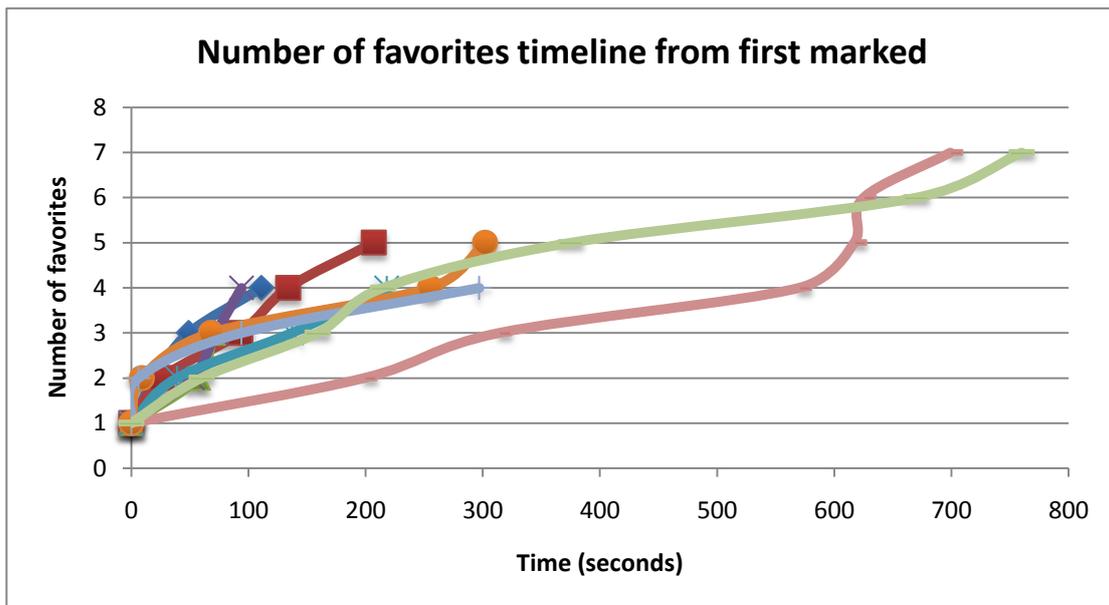


Figure 26: CSG - Number of favorites across time for the *External Users* group

3.1.3.4. First query time

The first query performed by the users will give us an idea of how much time it took for them to read and understand the task.

	Task 1	Task 2	Task 2a	Task 2b
Max	432.00	80.00	80.00	54.00
Min	14.00	5.00	7.00	5.00
Average	82.24	26.67	34.91	17.60
Stdev	83.06	22.95	24.49	17.00

Table 6: First query time for the *Consortium Members* group

	Task 1	Task 2	Task 2a	Task 2b
Max	272.00	112.00	112.00	79.00
Min	20.00	4.00	4.00	5.00
Average	109.33	45.44	58.00	32.88
Stdev	66.96	33.37	36.25	24.44

Table 7: First query time for the External Users group

As we see in Table 6: First query time for the *Consortium Members* group **and** Table 7: First query time for the External Users group

, it took less time to start working for the second task than for the first one. This could be explained by the training done with the first task: users are more experienced with the evaluation workflow and they can start quicker. It is also worthwhile to observe the difference between the two second tasks. Task 2a (historical sightings in Rome) get higher starting time than Task 2b (museums in London). This may indicate that the latter task was easier to understand.

3.1.4. Qualitative Analysis

3.1.4.1. Post Experiment Questionnaire

Besides the standard usability part, the questionnaire contained some specific questions about the interface, with the aim of knowing the opinion of some core parts of the workflow. The organization of an interface that integrates different sources of information is a key to let users navigate through all the information quickly and effectively, and detecting problems or symptoms of problems early is crucial when integrating more sources of information in the tool. The results of the first 3 questions are shown in Figure 27 and Figure 28.

Question 1: *The organization of the navigation bar on the left-hand side of the travel exploration portal is clear.*

Results shown in Figure 27 show that there is 0% “Strongly disagree” and 50% of the answers are positive for *Consortium Members*. However there isn’t a clear agreement about the organization of the navigation bar. Improvements in the prototype should keep this fact into account in order to make the left part more useful. The results are much better for *External Users*, as 80% valued the navigation bar as a clear solution to navigate.

Question 2: *The break up of search results in the navigation bar is useful to quickly narrow down the search results.*

More than 50% of the users expressed agreement with the question. However, some users disagreed and for the *Consortium Members group* some even strongly disagreed. We’ll investigate this issue in more detail when preparing for the second round of prototypes

Regarding to *External Users*, less than 10% showed a disagreement and none of them strongly disagreed. An overall analysis of the qualitative questions for this group tells us that this question gets the lowest result.

Question 3: *The organization of a location page is clear. A location page shows the detailed information of a selected location, point of interest, or an event.*

The results for this question are very positive, meaning that the tool effectively presents information from different places in a structured way. The challenge will be to continue with this level of agreement when adding more functionality.

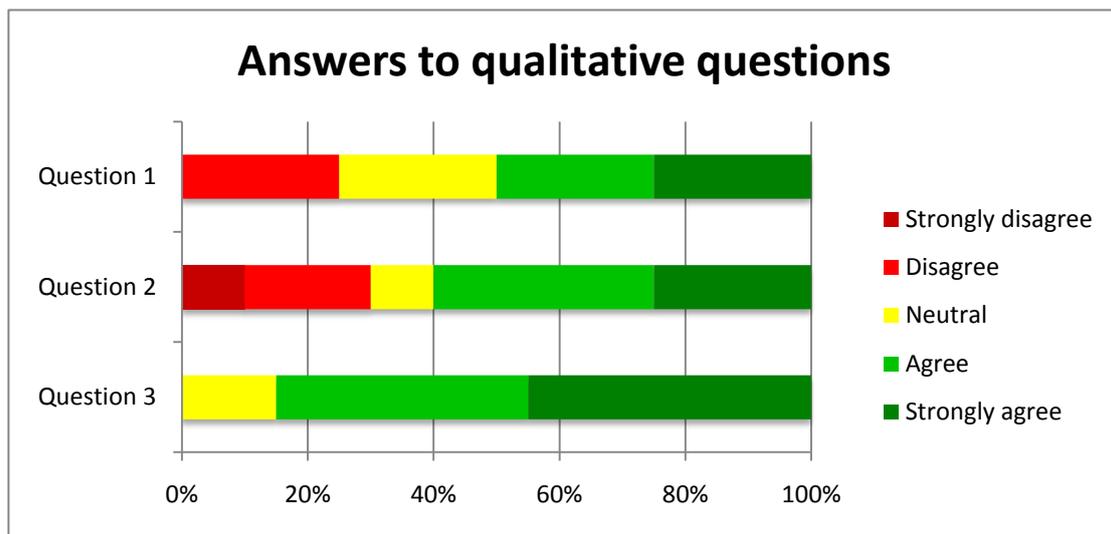


Figure 27: CSG - Answers to specific questions about the interface for *Consortium Members*

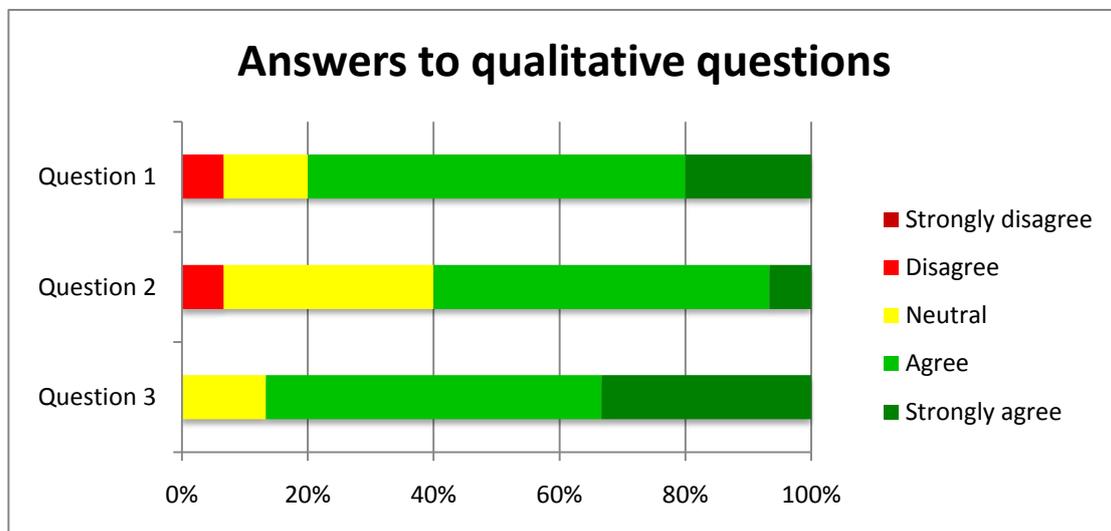


Figure 28: CSG - Answers to specific questions about the interface for *External Users*

Answers to questions 4 and 5 can be seen in Figure 29 and Figure 30.

Question 4: *I have used the breakup of search results presented in the navigation bar to focus my results on Places, Points of interest and Events.*

The answer is generally positive. Nevertheless we observe a similar percentage of negative questions to the disagreeing users in Question 2. Again, a study of the behavior of this class of users should be done to understand what trend they follow without the need of the break up in the results.

Moreover, we can say that this functionality differs from the presentation done by many well-known search engines, where a list of results is presented, with no clustering.

Question 5: *I felt the need to consult additional sources, besides the information presented in the travel portal.*

The results for this question are not as good as they should be if the service would be able to provide enough information. The *External Users* group shows a clear demand to consult more sources (more than 70%). The future work should be focused on the improvement of the amount and quality of information given, as intuition tells us that the service does not add enough value to the users to switch from usual travel portals to the service. On the other hand one should take into consideration that the CSG prototype is only a prototype and it should be expected that the users would want to consult other applications.

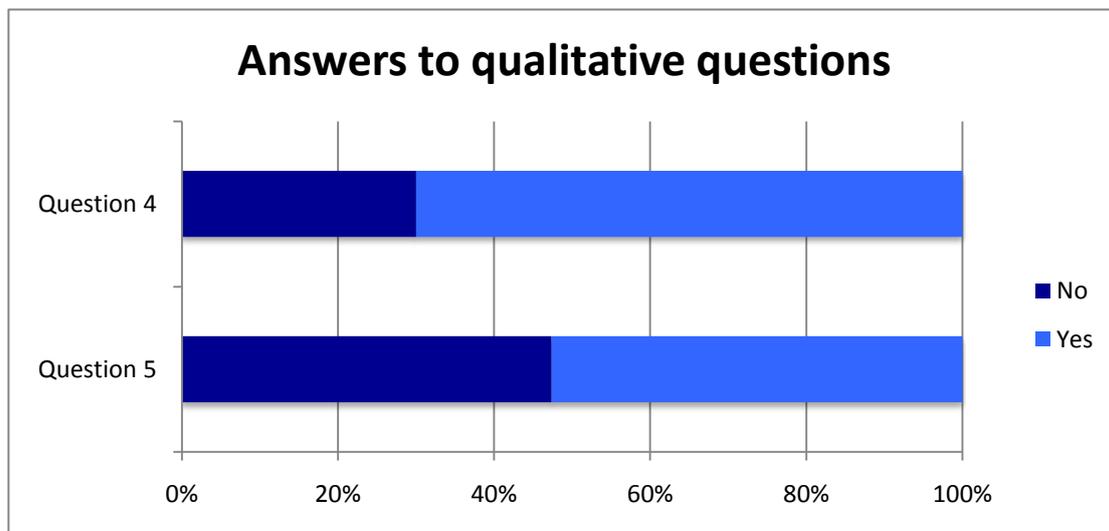


Figure 29: CSG - Answers to specific questions about the interface for *Consortium Members*

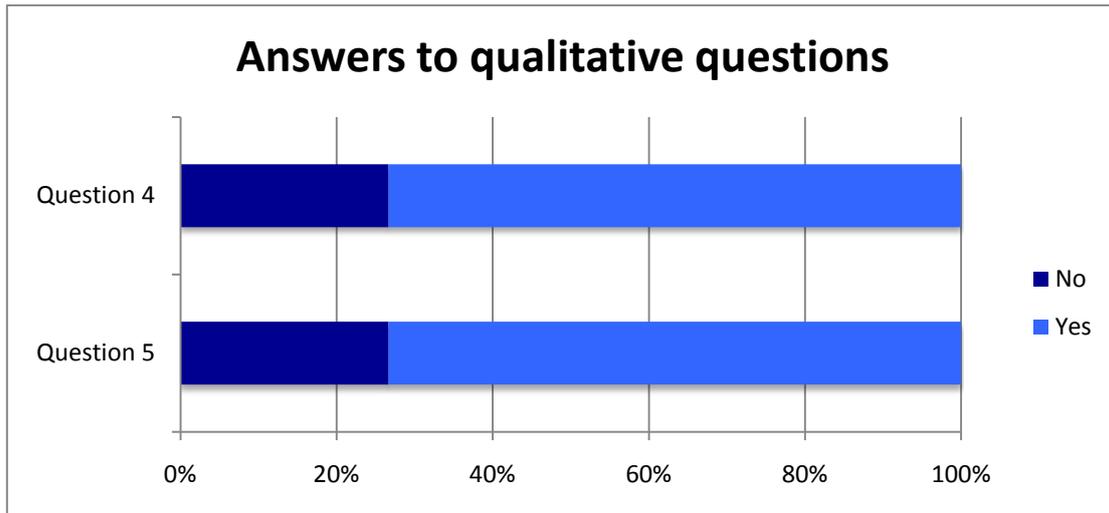


Figure 30 : CSG - Answers to specific questions about the interface for *External Users*

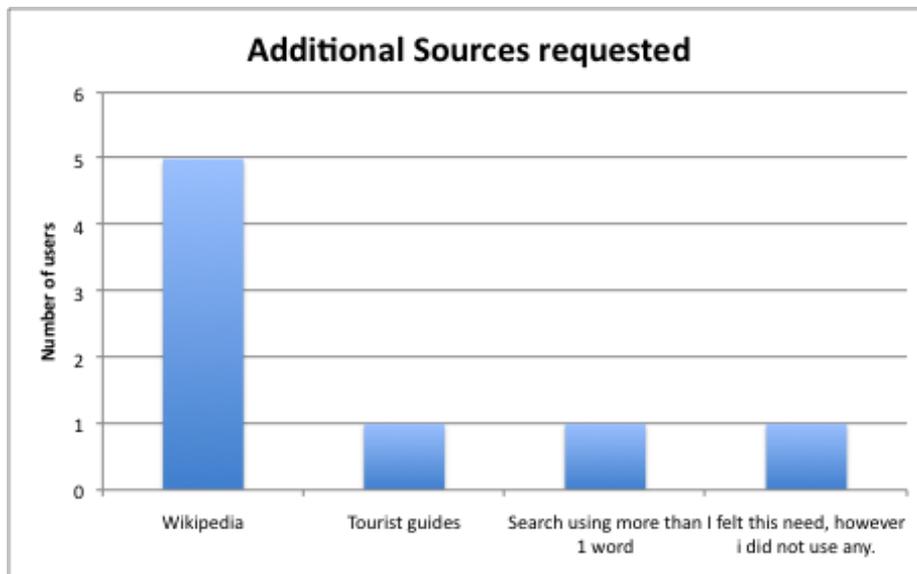


Figure 31: CSG - Answers to "additional sources needed" question for *Consortium Members*

Question 6: If yes, which ones?

It is interesting to note that even if all information was based on Wikipedia most users mentioned Wikipedia as the source they wanted to consult. We interpret this that they probably wanted links to full content of Wikipedia since we only showed the first section of each Wikipedia article.

3.1.4.2. Task 1

Since the main goal of Task 1 was to collect data for personalization experiment and to give users the opportunity to get familiar with the system we will not analyse in detail the outcome of the questionnaire answers. Answers to open questions asked after Task 1 can be seen in the section Open Question Feedback which can be found below.

3.1.4.3. Task 2

For Task 2 we compare the user satisfaction/experience for the two types of tasks. In this evaluation we do not distinguish between *Consortium Members* and *External Users*, but compare the difference between the users who solved each of the two tasks. Four questions were asked about the user satisfaction and user experience in the post task questionnaire.

		Mean		Median		T-test
		2a	2b	2a	2b	P
Q1	How satisfied are you with your performance of completing the task?	3.1	2.9	3.0	3.0	0.7584
Q2	Did you find more or less historical sightings/museums than you expected?	3.6	2.4	3.0	2.0	0.0072
Q3	How satisfied are you with the support of the CSG tool when completing the task?	3.1	2.4	3.0	2.0	0.0369
Q4	How difficult/easy was it to find historical sightings/museums?	3.4	2.0	4.0	2.0	0.0004

Table 8: Mean and median user satisfaction/experience for Task 2

Table 8: Mean and median user satisfaction/experience for Task 2 shows the mean and median user satisfaction/experience for the two types of Task 2. Appropriate answers were on a 5 point scale where ‘Very dissatisfied/difficult/few’ is represented by 1 point and ‘Very satisfied/easy/many’ is represented by 5 points. A ‘Neutral’ experience is represented by 3 points. We see from the table that the user experience was close to ‘neutral’, which is rather disappointing and we will need to improve the system for the 2nd round of prototypes.

If we compare tasks 2a and 2b we see that the satisfaction/experience is consistently worse for task 2b. For task 2a the experience is slightly better than ‘neutral’ but slightly worse than ‘neutral’ for task 2b. The difference is statistically significant for questions Q2 and Q4 ($P < 0.01$ for unpaired, two-tailed T-test). Based on answers to open questions (see details in next section) we think that the limited search functionality and POI clustering is mostly to blame for the worse user experience in task 2b since it was a more specific task than 2a. I.e., a smaller portion of London POIs are museums, compared to the portion of historical sightings among the POIs in Rome. We plan to improve the search functionality and POI classification in the 2nd round of prototypes.

3.1.5. Open Question Feedback

In the questionnaires we asked four open questions about what users found missing or could be improved in a later version of the tool. The questions were:

- How suitable do you think the CSG tool is for this task - please provide additional feedback. (Asked after Task 1)

- Do you have any additional comments the systems performance for completing this task? (Asked after Task 2a and Task 2b)
- What other types of information would you like to have included? (Asked after the experiment)
- Before finishing the evaluation, do you have any additional comments? (Asked after the experiment)

Table 9 shows an aggregated overview of the feedback from the users. The table shows results for both *Consortium Members* and *External Users*. We focus on feedback that can be used to improve the prototype in the next round of prototyping. Each row of the table shows a short description of the issue and the frequency of its mentioning. We break the feedback into three groups: data feedback, backend feedback and frontend feedback – depending on which part of the system the feedback addresses.

Module	Issue	Frequency
Data	Tourist information (Restaurants, hotels, opening hours, etc.)	17
	Transport information (public transport, travel times, etc.)	13
	User ratings, comments, and reviews	10
	Suggested routes through a city	3
Back end	Better search (better search functions and result filtering)	20
	POI categorization (Museums, parks, metro stations, etc.)	18
	Suggest related POIs (spatial, typological, personalized, etc.)	11
	Social planning and sharing of trips	3
	Removal of irrelevant POIs or photos	3
Front end	Improved map display (larger map, more POIs in map, etc)	12
	Include map based route planning	6
	Improved organization of results (was confusing or inefficient)	5
	Improved map interaction (scrolling, zooming, etc.)	3
	Enable back-button functionality	3
	Improved favorites interaction (not working or unclear)	3
	View favorites on map	2
	Improve organization/interaction with images	2
	Include snippets for POIs in result list	2

Table 9: Tool improvement suggestions based on aggregated outcome of answers to open questions

3.1.5.1. Data

The most frequent feedback was that the users wanted more tourist information, such as opening hours, official websites and transport connections. Many users also expressed their interest in getting comments and ratings from other users. A few users mentioned the functionality of having suggested routes through a city.

3.1.5.2. Back end

The most frequent back-end feedback was on POI categorization and improved search options. The search related feedback included better textual search capabilities, adding filters on search results and visual search. Many users mentioned that suggestion of related POIs would be useful, either nearby POIs, POIs of the same type, or POIs that would fit a personalized profile. Some users also mentioned that they would like support for planning trips with friends or sharing trip descriptions.

3.1.5.3. Front end

The most frequent front-end issue was improved map display and interaction, such as better zooming, including thumbnails, showing a set of objects in a map, etc. Some users found the result organization and favorites management confusing. Other users mentioned easily solved issues such as enabling back-button functionality, including more POIs in the navigation bar, adding snippets for POIs, etc.

3.1.5.4. Using feedback for 2nd round of prototypes

We will take the given feedback into account when working on the 2nd round of the prototypes. We will discuss with other partners in other work packages of the project about providing services for addressing these issues. We will give high priority to issues mentioned by many users, while also giving priority to scientifically interesting problems that will significantly improve over the state of the art.

3.1.6. Conclusions

Regarding to the CSG Web prototype, the evaluation process was carried out with two different groups: *Consortium Members* group gave us useful information that can be used in future collaboration and services addition in the project. Moreover, the fact that they are familiar with the project before hand gives to their opinions a more mature view, and is considered very useful to perform changes to services and interface organization.

External Users, on the other hand, showed their opinions for a product and not for a research prototype. Their opinion is especially important to observe how a tool like CSG would impact as a real product, and this

helps us to not lose the focus: CSG is a tool for planning a trip, and it should aim to help users to get the most complete view of places they want to visit. To do evaluations with users that know little or nothing about the project has the risk of getting really bad results and the advantage that these results are not biased towards our interests. In this case we can say that the results were really good and the feedback obtained very useful.

As a result, we can say that the overall usability of the tool is good with room for future improvement. Users showed the following broad areas of improvement:

- Navigation: The workflow to access the information is good, but can be improved. Our target for future releases will be to do the frontend more intuitive.
- Information shown: Users underlined the need to consult more information sources, and this will bring us to include more information related to travel and tourism in the next release.

In terms of using the evaluation to assess the quality of the WP3 services that were integrated in this first round of prototypes we conclude that it is essential to improve the search functionality of the services as well as developing methods for classifying POIs.

The results of this report should be seen as early feedback on a tool that is under development. The message is that it is a promising tool with room for improvement.

Finally, the analysis of all comments and responses allows us for the identification of specific improvements for future releases:

Type	Issue	Priority	Cost	Score
Interface	Improve Favorites display	5	2	20
Interface	Improve Map interaction	5	2	20
Interface	Improve application dynamism and workflow	4	2	16
Evaluation	Log map clicks and, in general, all events in further evaluations	3	1	15
Sources	Add information from WeKnowIt system	5	3	15
Interface	Improve navigation bar	4	3	12
Sources	Add information from external sources	4	3	12
Interface	Add more information in the Map	4	3	12
Evaluation	Perform Precision / Recall analysis for Task evaluation	2	1	10
WP3	Improve search functionality	5	4	10
WP3	Improve POI classification	5	4	10
Sources	Add more tourist information	4	4	8
Functionality	Add snippets for POIs	4	4	8
WP3	Add POI categorization information	4	4	8
Interface	Do navigation bar optional (disagreeing users)	3	4	6
Sources	Add ratings and comments	3	5	3
Interface	Find new ways of result clustering and evaluate them	2	5	2
Functionality	Allow for social planning	2	5	2

Table 10: CSG – Improvements for the CSG Web Prototype

3.2. Mobile Guidance

The user trial and evaluation of the Mobile Guidance Application was done according to the task T7.2.3. “Evaluation and user trial” specified in WP7 – Case Study for the Consumer Social Group.

Given that the prototype was at an early stage, it has been considered more useful to perform a participatory evaluation, whose purpose is to identify areas for improvement and which does not require independent evaluators.

Two different evaluation experiments have been implemented: one involving TID staff with experience in European projects and another involving the WeKnowIt Project partners.

3.2.1. Evaluation

The object of evaluation has been the mobile guidance application. It is described extensively at deliverable D7.4.1 Initial Consumers Social Group Case Study Implementation.

The functionality implemented in this version is represented by an UML use cases diagram in next graph:

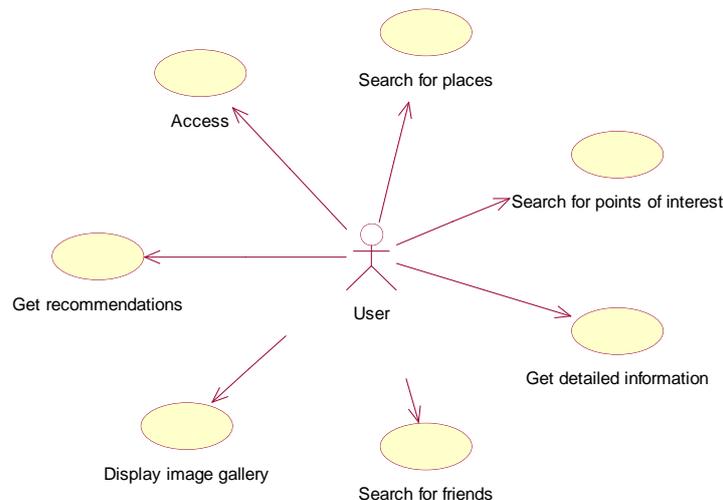


Figure 32: CSG - Mobile guidance: uses cases

In order to help identify the interface of the application used for the evaluation, some snapshots are provided below, in Figure 33.

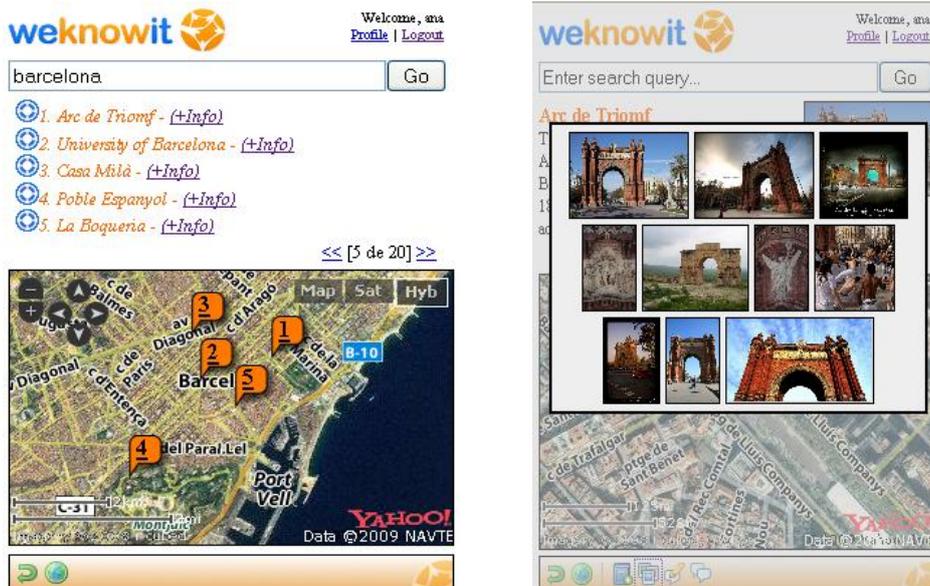


Figure 33: CSG - Search for POIs and Image gallery

The user interface shown above works over the deployment shown in next graph.

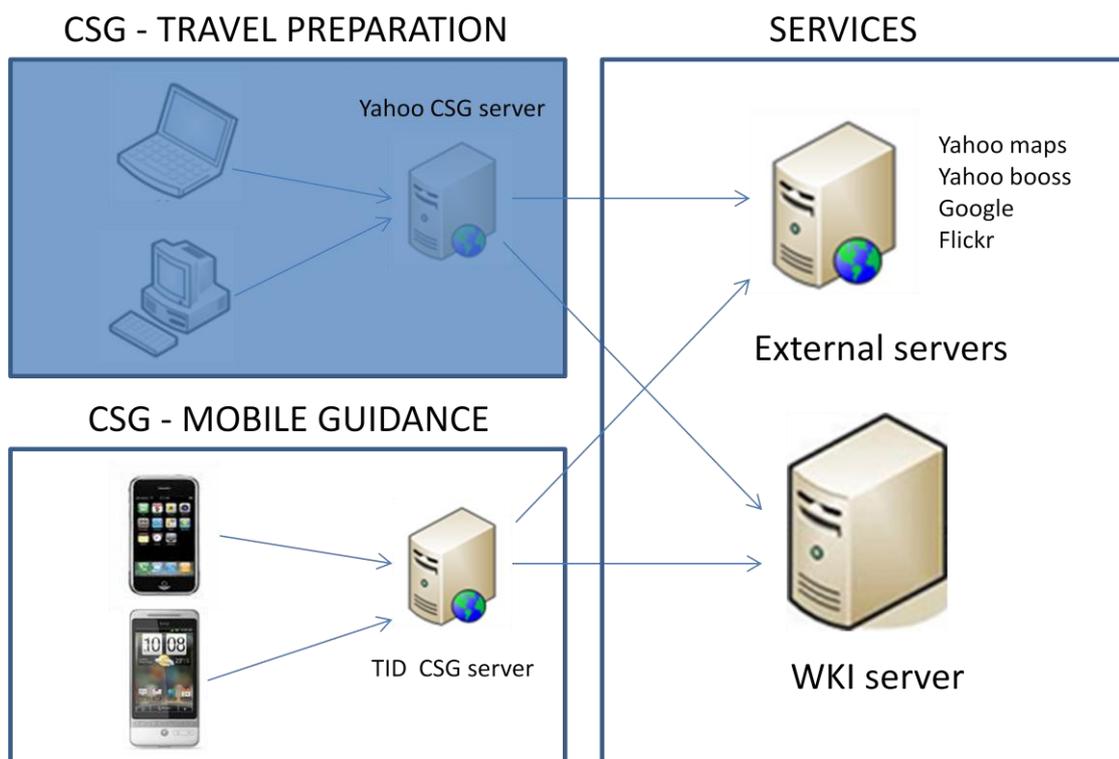


Figure 34: CSG – Mobile Guidance deployment

3.2.2. Evaluators

Two evaluations were performed, each implemented by a group of evaluators. The first evaluation has been performed by a group of 4 people at TID. These people have decided to voluntarily contribute to the evaluation. The evaluation was held in Madrid on 16 December 2009.

TID evaluators’ team was composed of people with previous experience in prototypes evaluations, with profiles as project managers or researchers.

The second evaluation has been performed by a group of 21 people during the project meeting held in Barcelona in January 2010. This second team had a wide range of profiles: developers, researchers, project leaders and to head of units.

In both cases, it may be argued the lack of independence towards the evaluator, but, as previously informed, the purpose of the evaluation is to find areas of improvement; this configuration is well accepted by the formal methodologies.

3.2.3. Evaluation Process

TID’s evaluation

Place: Madrid

Date: 16 December 2009

Number of users-evaluators: 4

Equipment: Servers and 4 mobile phones

Preparatory meeting: it was maintained in order to explain the project purpose and time plan, the system architecture, and the Consumer Social Group Prototype. The Mobile Guidance Application’s functionality was demonstrated on field.

Routes available to users: each user-evaluator had a different route. Time taken to follow the routes was around 2 hours.

Route #1



Route #2



Route #3



Route #4



Consensus meeting: The users-evaluators had a post-experiment meeting, where they exchanged experiences and considerations and finally produced an evaluation and listed some recommendations.

Partner's evaluation

Place: Barcelona

Date: 21 January 2010

Number of users-evaluators: 22

Equipment: servers and 22 mobile phones

Preparatory meeting: Since the project partners had the project context in mind, no further explanation was given for this purpose; and a time slot was kept in the agenda to demonstrate the mobile guidance application.

Routes available to users: according to recommendations; each user-evaluator had a different route. The time taken to follow the evaluation route was 5 hours.

Questionnaires: the users-evaluators filled questionnaires where their opinions and recommendations were gathered.

3.2.4. Evaluation Outcome

TID's outcome

TID's evaluators did a consensus meeting in order to share and discuss their perception and experiences; and give their outcome.

Before stating their evaluation, the team made a series of considerations, which were taken into account:

- Consideration to the time frame since TID joined the consortium April 2009 to the time the prototype is delivered December 2009

- Consideration to being first iteration, and found adequate having a first useful prototype addressing some available services from the WeKnowIt architecture
- Consideration to addressing users needs: need of touristic information, need of recommendation about touristic places, need of support for physical navigation on field, need of communication with touristic group

Outcome:

Availability: All items of functionality were available

Usefulness: Positive on all items

Easy to use: Positive

Comments: difficulties to handle the map

Easy to understand: Positive

No comments

Appearance: Positive

Comments: Integration of the application in the mobile phone hosting it as a native application would increase its usability and appearance.

Responsiveness: Positive.

Comments: Sometimes lack of responsiveness happened when downloading maps. Focus on the frequency when maps are downloaded. People don't move so quickly on field.

Suitability: Positive

Comments: Current functionality is found suitable but needs improvements. Suggestion to better address user needs.

Missed functionality: Positive

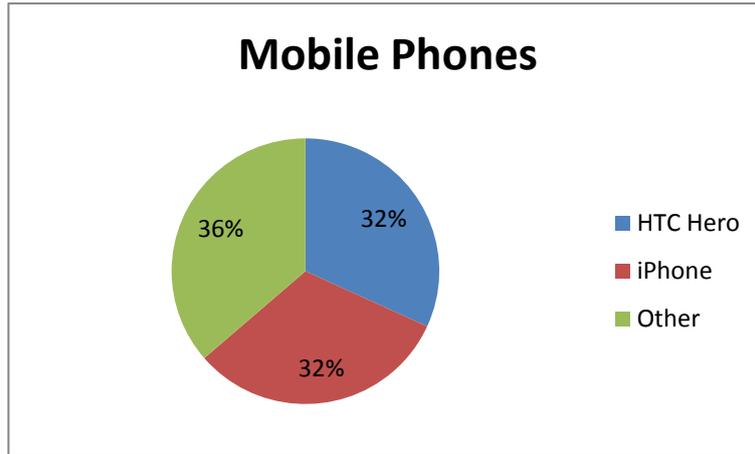
Comments: Maps should offer functionality to centre it on actual user position. Users should be able to add temporary Points of Interest. There should be categories of points of interest (where to have a break, where to take a taxi, etc). Automatic path search to a point of interest would be useful. Broadcast messages or pictures to the group.

Collective Intelligence: Positive

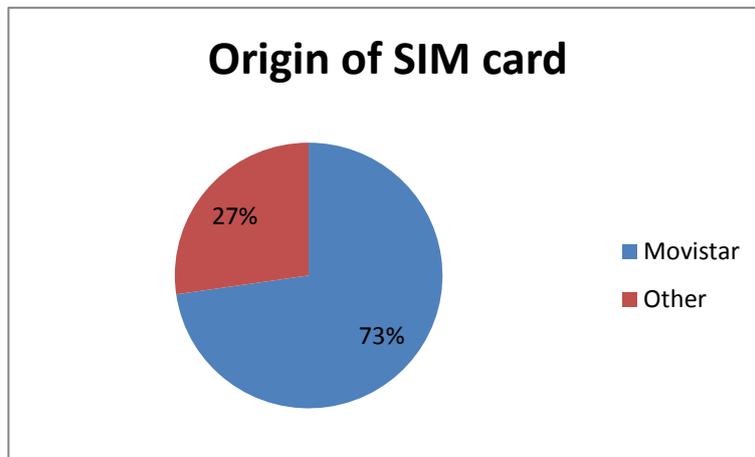
Comments: Evaluators believe this tool, when completed, can improve the experience of a group of consumers on a touristic trip.

Partners' outcome

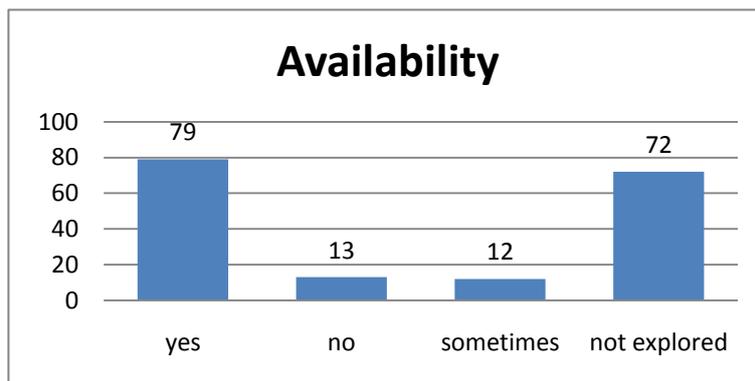
Q1 – What model of mobile did you use in the evaluation?

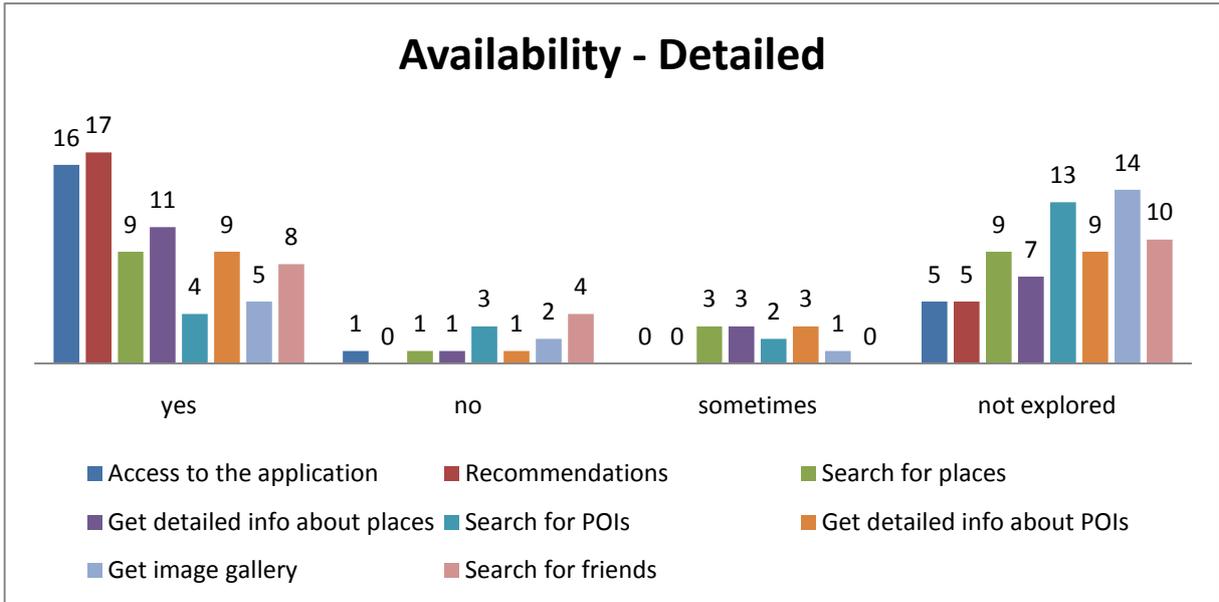


Q2 – Was your SIM card provide by TID?

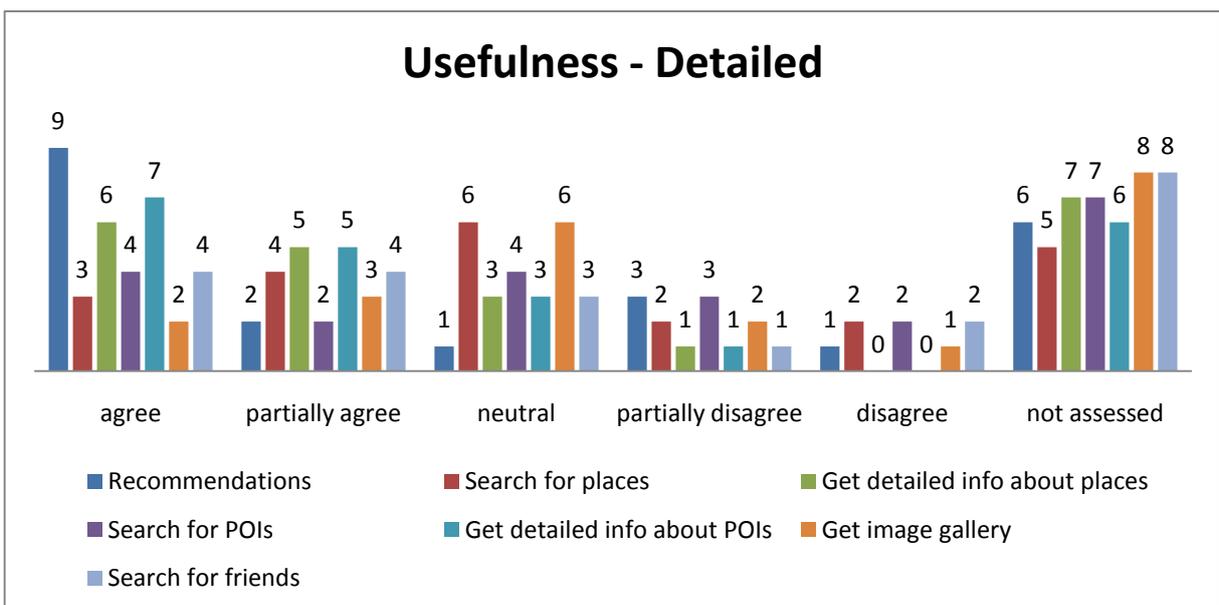
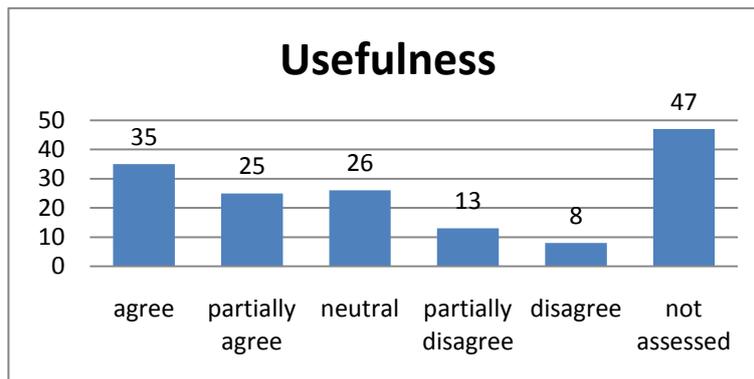


Q3 – The following functionalities were available in the Mobile Guidance Application

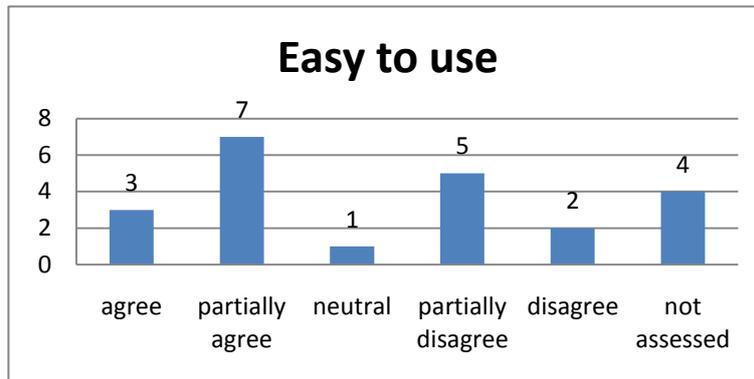




Q4 – The following functionalities were useful in the Mobile Guidance Application



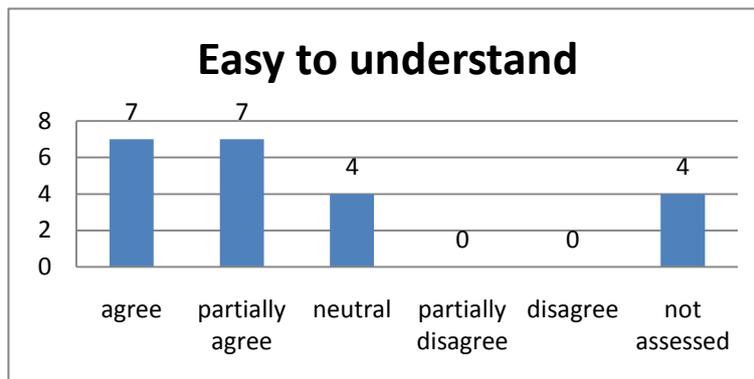
Q5 – The mobile guidance application was easy to use



If you disagree, please, tell us what limitations and difficulties you came across

Summary of comments: bottoms and links were too small, which made operation difficult, specifically with the map. It currently appears in hybrid mode (map and satellite picture); it is suggested to present only the map. Search is very limited.

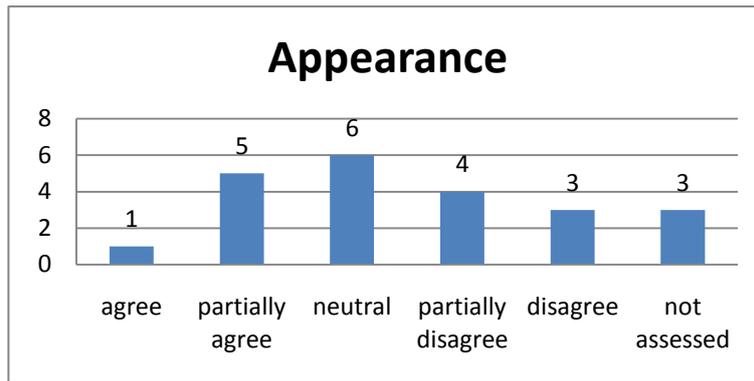
Q6 – The mobile guidance application was easy to understand



If you disagree, please, tell us what items, behaviours or elements were confusing

Summary of comments: There was a question to understand on what basis recommendations were given.

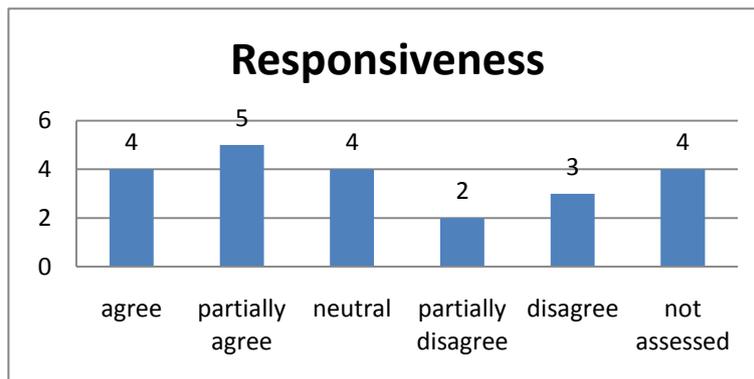
Q7 – The appearance of the mobile guidance application was appealing



If you disagree, please, make suggestions or identify which areas or elements were unappealing

Summary of comments: Some evaluators simply indicated that it was not pretty or did not look good. Others gave better guidance pointing out that a wide orientation would help or that icons and font style were nice, but layout and font size too small.

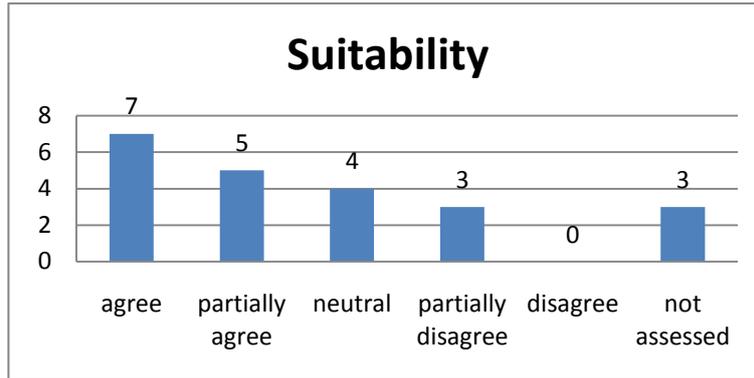
Q8 – I found the mobile guidance application to be responsive



If you disagree, please, give details on the situations where you experienced lack of responsiveness. (Did you consider network coverage?)

Summary of comments: Some evaluators experienced lack of responsiveness: *"It was generally slow. At some point the application froze but started working again after a few minutes"; "Some interruptions happened because of network problems"* but others had better experience: *"The application was responsive even though it can be further optimized, e.g. by caching recently downloaded info"; "It was OK most of the time. It stopped responding for a period but then it was up again"*

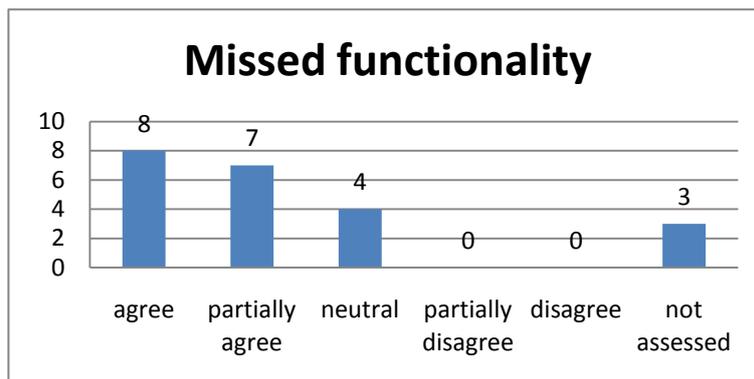
Q9 – All items of functionality were suitable for mobile guidance



If there is no level of agreement, please identify which functionalities were not suitable.

Summary if comments: few comments were received; image thumbnails were not found suitable. There was a positive comment: *"Users and friend locations were nice"*; and a suggestion: *"Map navigation +/- and moving should work with gestures"*

Q10 – I missed other functionalities that may be suitable for mobile guidance



If you agree, please, make suggestions for additional functionality

Summary of comments: There were many suggestions.

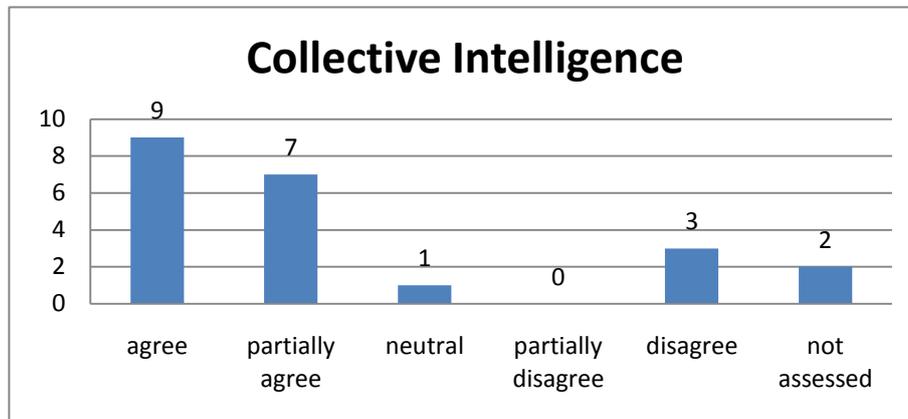
About the route: Find path to POI or optimal route planning, including information about public transport, such as subway map or relevant buses. Visualize routes, what you have done with timestamps, what is ahead. Allow pictures taken to be displayed in the route. Export your route to a gpx format. Bookmark locations

About the POIs: better search for POIs, list of nearby POIs and attractions, information and news about POIs such as opening hours; distances to POIs, Include as POIs restaurants, coffees, etc.

About recommendations: route-based recommendations; recommendation of events; recommendations for groups rather than individuals

About communication: with friends – support for group travellers; emergency alerts

Q11 – I believe consumer’s users groups on a touristic trip can increase their collective intelligence by using the mobile guidance application

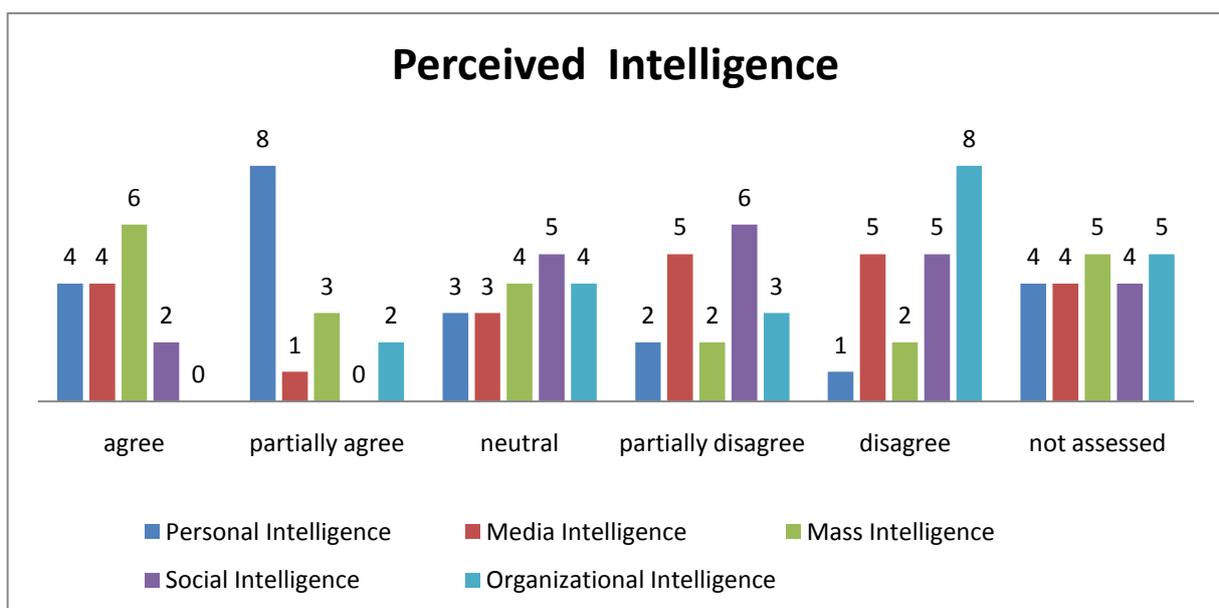


There is important acceptance on this statement but two negative comments were received:

"My general comment is that any pocket grid + map are much more useful. The only nice thing I noticed is the GPS location on the map – this can be useful"

"The user can improve their experience by exploiting the results of the analysis resulting to Collective Intelligence. They can also contribute to the generation of CI"

Q12 – I perceive the following levels of intelligence in the behaviour or content of the mobile guidance application



Others Comments:

We would appreciate any other additional comments on the evaluation process or the mobile guidance application itself.

- *Overall the application was good, and I believe that most issues noted above will be removed when it is implemented as a native iPhone application*
- *Very nice and well-organised experiment. It actually evaluates effectively the overall application from a user point-of-view. Additional requirements might be the evaluation of individual services (e.g., the satisfaction from the WP services – How satisfied is the user from the outcome of the POI ranking)*
- *Most of the data of the application (POIs, descriptions, and images) should be stored locally on the device so that not many server requests are issued. The login should take place only once in the beginning of the day and last for many hours. Log-in should not be requested even when the application shuts down. The maps were slow. Faster map access is necessary for the phone to be useful for touristic navigation*
- *The users takes a photo and the application uses the current position and image analysis to return information about the POI*
- *For the interface you can check existing iPhone applications like: Dpplr and Everytrail*
- *Lots of power consumption (battery empty after 3 hours)*
- *Application crashed three times*

3.2.5. Conclusions

The mobile guidance application that integrates collective intelligence services provided by WeKnowIt system has been evaluated within its first implementation cycle. Two different evaluator's teams performed participatory evaluations, which are oriented to find areas of improvement, have assessed the features of the prototype.

The first evaluation team from TID was composed of 4 members with joint experience in European projects, research, system design and architecture, and project management. They held a field evaluation experiment on 16 December 2009 in Madrid (Spain). Their feedback was the outcome of a consensus meeting.

The second evaluation team was formed by 22 people contributing to the WeKnowIt Project in a field evaluation experiment held on 21 January 2010 in Barcelona (Spain). Their feedback was collected by means of a questionnaire.

Different mobile phones were used in the evaluation exercise: HTC Hero, HTC Magic, iPhone, Motorola Milestone and Samsung Galaxy.

The evaluators found addressed the following user needs: touristic information and recommendations, field navigation guidance and group

communication, which were provided by three kinds of intelligence: Personal intelligence, Mass intelligence and Social intelligence.

The items of functionality considered for evaluation were: get recommendations of Points of interest; search for places; search for Points of Interest; search for friends; get detailed information about places; get detailed information about Points of interests; get image gallery about Points of interest.

The services integrated that were providing content and behaviour to the several functionality items come from personal intelligence, mass intelligence, and social intelligence.

The following dimensions were explored in the evaluation: availability, usefulness, usability, simplicity, appearance, responsiveness, suitability, completion and collective intelligence.

The evaluation outcome is displayed in Figure 35. Most dimensions received positive feedback, which means complete or partial agreement on that dimension.

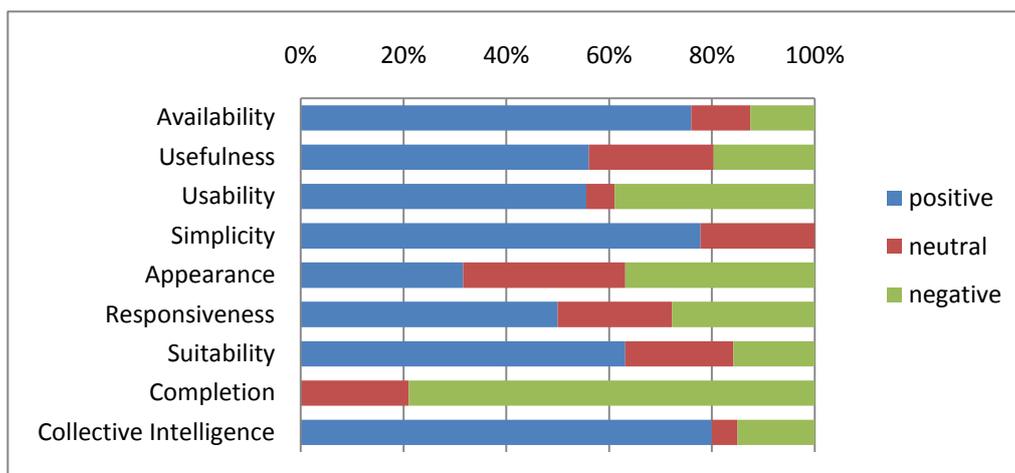


Figure 35: Summary Evaluation Outcome

Most evaluators were positive on the ability of collective intelligence applications to provide useful services to the group of users using the mobile guidance application. The evaluators also found suitable the functionality made available in the first-iteration prototype, at the same time they clearly indicated the application needed improvement, and pointed out several areas: route, points of interest, recommendations and group communication.

Route: Find the optimal path to a Point of interest; include information about public transport, such as subway map or relevant buses. Visualize routes, what you have done, what is ahead. Allow pictures taken to be displayed in the route. Export your route in graph format.

Points of interest: improved search of POIs; list of nearby POIs, and attractions; information and news about POIs such as opening hours, distances to POIs; include kinds of POIs such as “where to take a break”, “where to lodge”, etc

Recommendations: improved recommendations based on the route; recommendation of events and recommendations for groups.

Group communication: broadcast messages to your travellers group; broadcast emergency alerts.

Technology: implement the module as a native application of the mobile phone

4. References

- [1] WeKnowIt Annex I – “Description of Work”, ver.1, FP7-215453, 26 Oct. 2007
- [2] WeKnowIt “D7.6.1 Initial consumer and emergency response use case evaluation protocols”, Oct. 2009
- [3] Evaluation Methodology Basics, E. Jane Davidson – Sage Publications, Inc.
- [4] WeKnowIt “D7.3.1 Initial emergency response case study implementation”, Nov. 2009
- [5] WeKnowIt “D7.4.1 Initial consumers social group case study implementation”, Nov. 2009

5. Appendix 1: CSG – Travel Preparation - Questionnaires

Demographics questionnaire

General Information

1. What is your age?
 - a. <20
 - b. 21-30
 - c. 31-40
 - d. 41-50
 - e. >50
2. Gender
 - a. Male
 - b. Female
3. Where do you live now (country)?
4. If you are a parent how many children (18 years or under) are there in your household?
 - a. None
 - b. My children are older than 18
 - c. 1
 - d. 2
 - e. 3+
5. How frequently do you use a search engine (Google, Yahoo! Search, Bing etc.)?
 - a. Never
 - b. Monthly
 - c. Weekly
 - d. Daily
 - e. Often each day

Travel

1. How many international trips have you taken in the past year?
 - a. None
 - b. 1
 - c. 2-3
 - d. 4-5
 - e. 6-10
 - f. 11+
 - g. Don't know
2. Do you combine business and holiday trips?
 - a. I don't travel for business
 - b. When I can
 - c. Most of the time
 - d. Sometimes
 - e. Never
3. Have you used the Consumer Social Group Web tool before this evaluation?
 - a. Yes
 - b. No
4. If you use online trip planners (e.g. Yahoo Travel, Dopplr, TripIt, Witur), what do you use them for? (Multiple options valid)
 - a. I do not use online trip planners
 - b. Get inspiration on where to travel

- c. Research destination information
 - d. Research hotel information
 - e. To read traveler reviews
 - f. To coordinate trips with other people
 - g. To help organize myself
 - h. Other (text box)
5. Which of the following activities describe what you most commonly do when travelling for pleasure? (Multiple options valid)
- a. Art
 - b. Architecture
 - c. Sightseeing
 - d. Museums
 - e. Beach
 - f. Outdoor activities
 - g. Spa
 - h. Ski/Snow
 - i. Dining
 - j. Golf
 - k. Shopping
 - l. Nightlife
 - m. Music
 - n. Theme Park
 - o. Volunteering
 - p. Other (text box)

SUS Questionnaire

Strongly disagree Strongly agree

1	2	3	4	5
---	---	---	---	---

	1	2	3	4	5
I think that I would like to use this system frequently					
I found the system unnecessarily complex					
I thought the system was easy to use					
I think that I would need the support of a technical person to be able to use this system					
I found the various functions in this system were well integrated					
I thought there was too much inconsistency in this system					
I would imagine that most people would learn to use this system very quickly					
I found the system very cumbersome to use					
I felt very confident using the system					
I needed to learn a lot of things before I could get going with this system					

Task 1, Pre questionnaire

1. Have you been to the city of Barcelona before?
 - a. Yes
 - b. No
2. How much would you say you know about the city?
 - a. Very little
 - b. Little

- c. Neutral
- d. Much
- e. Very much

(How much do you agree with these statements?)

3. It is very important that I have a checklist of activities to cover on vacation
 - a. Totally disagree
 - b. Disagree
 - c. Neutral
 - d. Agree
 - e. Totally agree
4. I tend to be influenced by comments/reviews posted online from other internet users
 - a. Totally disagree
 - b. Disagree
 - c. Neutral
 - d. Agree
 - e. Totally agree
5. I prefer to go on holidays where activities are organized for me
 - a. Totally disagree
 - b. Disagree
 - c. Neutral
 - d. Agree
 - e. Totally agree
6. I am comfortable travelling on my own
 - a. Totally disagree
 - b. Disagree
 - c. Neutral
 - d. Agree
 - e. Totally agree
7. I often ask friends and family for recommendations for vacation destinations
 - a. Totally disagree
 - b. Disagree
 - c. Neutral
 - d. Agree
 - e. Totally agree
8. What would you like to get out of sighting-seeing in Barcelona? (optional)
 - a. Free text option

Task 1, Post questionnaire

1. How do you feel about the time limit for the task?
 - a. Not enough time
 - b. Few time
 - c. Neutral
 - d. Enough time
 - e. Too much time
2. How complete an idea do you have of the places you wanted to visit?
 - a. No idea
 - b. I know a few places I want to visit
 - c. Almost complete
 - d. Complete
3. Do you think 5 hours will be enough to see the things you want to see?
 - a. Not enough time
 - b. Few time

- c. Neutral
 - d. Enough time
 - e. Too much time
4. How suitable do you think the CSG tool is for this task - please provide additional feedback, if any.
- a. Free text option

Task 2A, Pre questionnaire

1. Have you ever visited Rome?
 - a. Yes
 - b. No
2. How familiar are you with historical sightings in Rome?
 - a. Very unfamiliar
 - b. Unfamiliar
 - c. Neutral
 - d. Familiar
 - e. Very familiar
3. Are you interested in historical sightings?
 - a. Very uninterested
 - b. Uninterested
 - c. Neutral
 - d. Interested
 - e. Very interested
4. Have you searched for historical sightings in Rome before?
 - a. Never
 - b. Rarely
 - c. Occasionally
 - d. Sometimes
 - e. Frequently
5. Have you searched for historical sightings before?
 - a. Never
 - b. Rarely
 - c. Occasionally
 - d. Sometimes
 - e. Frequently
6. Please list all historical sightings in Rome that you can think of without any aid.
 - a. Free text option

Task 2B, Pre questionnaire

1. Have you ever visited London?
 - a. Yes
 - b. No
2. How familiar are you with museums in London?
 - a. Very unfamiliar
 - b. Unfamiliar
 - c. Neutral
 - d. Familiar
 - e. Very familiar
3. Are you interested in museums?
 - a. Very uninterested
 - b. Uninterested
 - c. Neutral

- d. Interested
 - e. Very interested
4. Have you searched for museums in London before?
 - a. Never
 - b. Rarely
 - c. Occasionally
 - d. Sometimes
 - e. Frequently
5. Have you searched for museums before?
 - a. Never
 - b. Rarely
 - c. Occasionally
 - d. Sometimes
 - e. Frequently
6. Please list all the museums in London that you can think of without any aid.
 - a. Free text option

Task 2A, Post questionnaire

1. How satisfied are you with your performance of completing the task?
 - a. Very dissatisfied
 - b. Dissatisfied
 - c. Neutral
 - d. Satisfied
 - e. Very satisfied
2. Did you find more or less historical sightings than you expected?
 - a. Much less
 - b. Less
 - c. About the same
 - d. More
 - e. Much more
3. How satisfied are you with the support of the CSG tool when completing the task?
 - a. Very dissatisfied
 - b. Dissatisfied
 - c. Neutral
 - d. Satisfied
 - e. Very satisfied
4. How difficult/easy was it to find historical sightings?
 - a. Very difficult
 - b. Difficult
 - c. Neutral
 - d. Easy
 - e. Very easy
5. Do you have any additional comments the systems performance for completing this task?
 - a. Free text option

Task 2B, Post questionnaire

1. How satisfied are you with your performance of completing the task?
 - a. Very dissatisfied
 - b. Dissatisfied
 - c. Neutral
 - d. Satisfied

- e. Very satisfied
2. Did you find more or less museums than you expected?
 - a. Much less
 - b. Less
 - c. About the same
 - d. More
 - e. Much more
3. How satisfied are you with the support of the CSG tool when completing the task?
 - a. Very dissatisfied
 - b. Dissatisfied
 - c. Neutral
 - d. Satisfied
 - e. Very satisfied
4. How difficult/easy was it to find museums?
 - a. Very difficult
 - b. Difficult
 - c. Neutral
 - d. Easy
 - e. Very easy
5. Do you have any additional comments the systems performance for completing this task?
 - a. Free text option

Tool specific questions

1. The organization of the navigation bar on the left-hand side of the travel exploration portal is clear.
 - a. Strongly disagree
 - b. Disagree
 - c. Neutral
 - d. Agree
 - e. Strongly agree
2. The breakup of search results in the navigation bar is useful to quickly narrow down the search results.
 - a. Strongly disagree
 - b. Disagree
 - c. Neutral
 - d. Agree
 - e. Strongly agree
3. The organization of a location page is clear. A location page shows the detailed information of a selected location, point of interest, or an event.
 - a. Strongly disagree
 - b. Disagree
 - c. Neutral
 - d. Agree
 - e. Strongly agree
4. I have used the breakup of search results presented in the navigation bar to focus my results on *Places, Points of Interest* or *Events*
 - a. Yes
 - b. No
5. I felt the need to consult additional sources, besides the information presented in the travel portal
 - a. Yes

- b. No
- 6. If yes, which ones?
 - a. Free text option
- 7. What other types of information would you like to have included?
 - a. Free text option
- 8. Before finishing the evaluation, do you have any additional comments?
 - a. Free text option

6. Appendix 2: CSG – Mobile Guidance – Questionnaire

1. What model of mobile phone have you used in the evaluation?

HTC HERO	iPhone	Other (please, specify)

2. Was your SIM card provided by TID?

yes	not (please, specify provider)

3. The following items of functionality were available

- Access to the application
- Recommendations
- Search for places
- Get detailed information about places
- Search for Points of Interests
- Get detailed information about POIs
- Get image gallery about POIS
- Search for friends in the group

yes	no	sometimes	not explored

4. The following items of functionality were useful

- Recommendations
- Search for places
- Get detailed information about places
- Search for Points of Interests
- Get detailed information about POIs
- Get image gallery about POIS
- Search for friends in the group

agree	partially agree	neutral	partially disagree	disagree

5. The mobile guidance application is easy to use

agree	partially agree	neutral	partially disagree	disagree

- If there is no level of agreement, would you please inform what areas or elements were difficult?

6. The mobile guidance application is easy to understand

agree	partially agree	neutral	partially disagree	disagree

- If there is no level of agreement, would you please inform what items, behaviours or elements were confusing?

7. The appearance of the mobile guidance application is appealing

agree	partially agree	neutral	partially disagree	disagree

- If there is no level of agreement, would you please make suggestions or spot what areas or elements were unappealing?

8. The mobile guidance application is responsive

agree	partially agree	neutral	partially disagree	disagree

- If there is no level of agreement, would you please give details on the lack of responsiveness?

(Did you check network coverage?)

9. All items of functionality are suitable for mobile guidance

agree	partially agree	neutral	partially disagree	disagree

- If there is no level of agreement, would you please identify what were not suitable

10. You have missed other functionality items suitable for mobile guidance

agree	partially agree	neutral	partially disagree	disagree

- If there is any level of agreement, would you please make suggestions

11. You believe consumer users groups on a tourist trip can increase their collective intelligence by using the mobile guidance application

agree	partially agree	neutral	partially disagree	disagree

- Please, make any comment you wish

12. You perceived any of the following levels of intelligence in the behaviour or content of the mobile guidance application

- Personal intelligence
- Media intelligence
- Mass intelligence
- Social intelligence
- Organizational intelligence

agree	partially agree	neutral	partially disagree	disagree

13. We would appreciate any other additional comment on the evaluation process or the mobile guidance application itself.

7. Appendix 3: ER – Evaluation Material

Screen Shots of Paper Mockups

The full set of mock-ups is shown in deliverable D7.3.1 – these are the screens used in the paper evaluations.

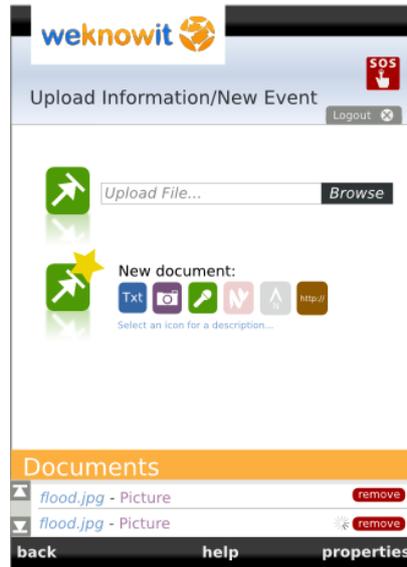


Figure 36: Mobile upload screen

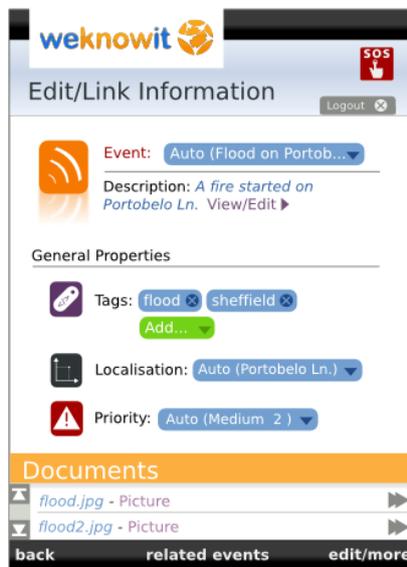


Figure 37: Mobile tagging screen

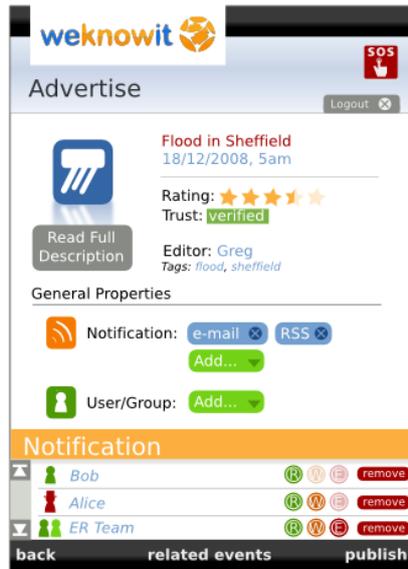


Figure 38: Mobile advertising screen

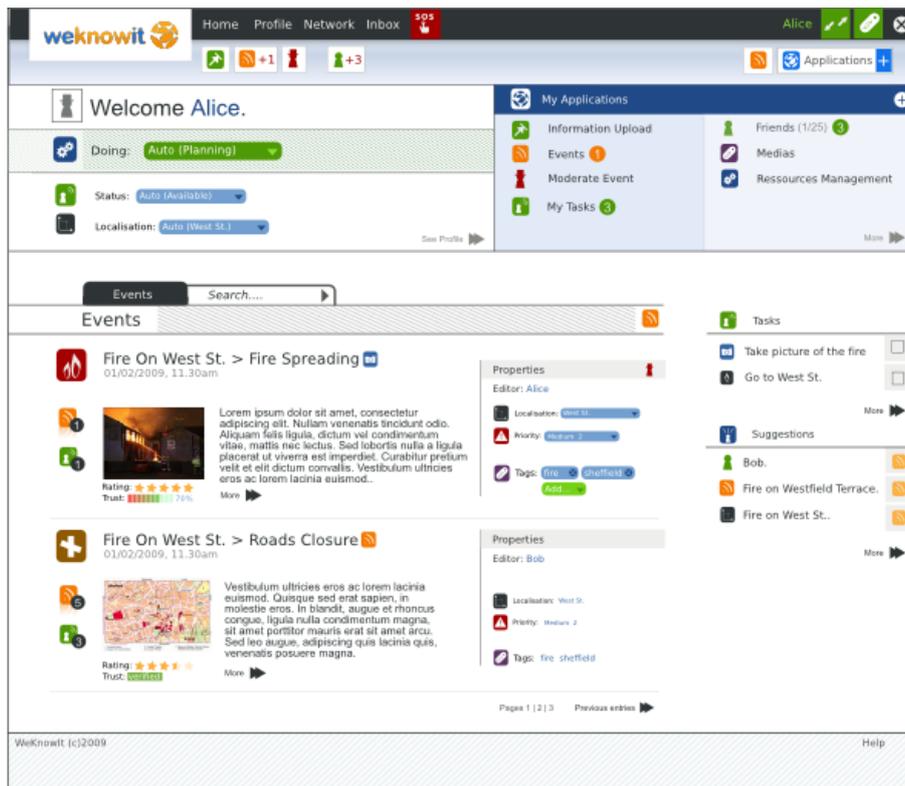


Figure 39: Desktop Homepage

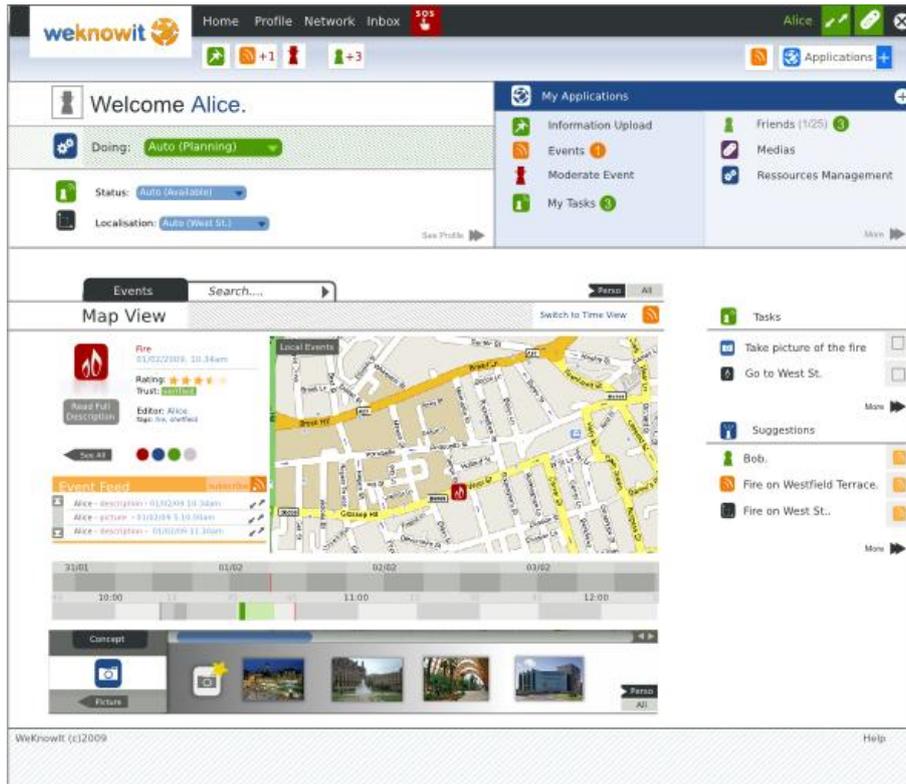


Figure 40: Desktop incident view

Screen Shots of Evaluated Application

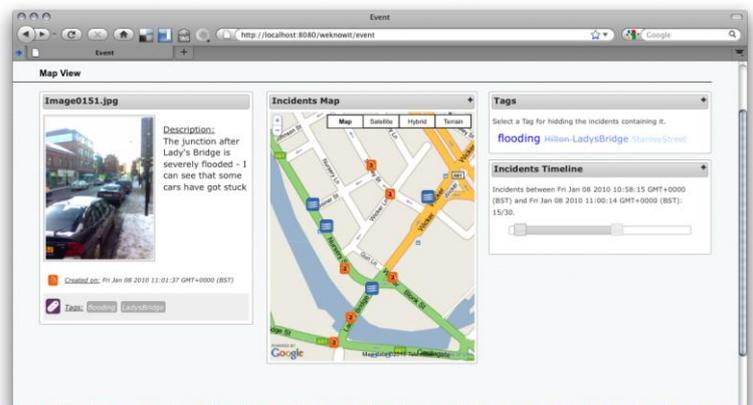


Figure 41: Incident view

This screen shows the main incident view for the ER application. In the top left a small view of the image can be seen with the tag and time filtering on the top right. The map in the middle allows the user to select one of the icons to see more information about that piece of evidence. Additionally there is a list of images at the bottom of the screen.

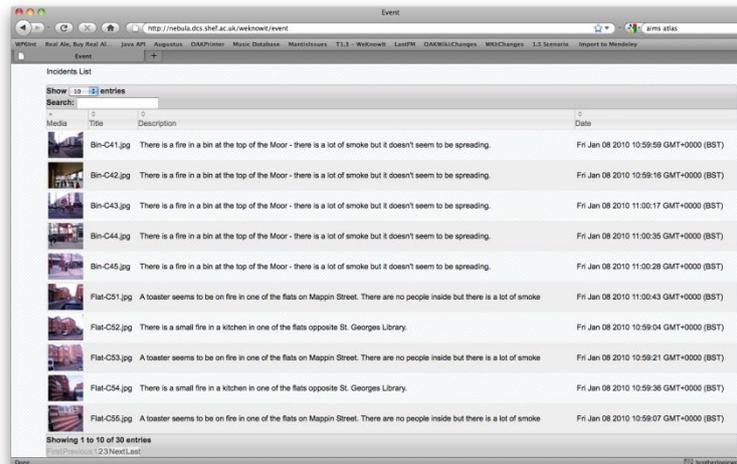


Figure 42: Incident list

The incident list shows a list of all the images and associated information at the bottom of the main application screen.

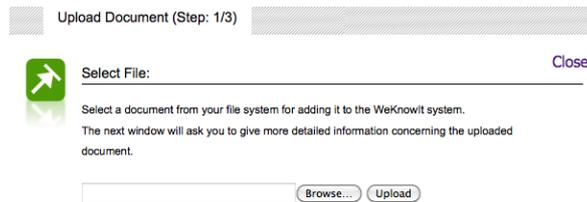


Figure 43: Upload; File select

This screen shows the file selection option to begin the upload process.

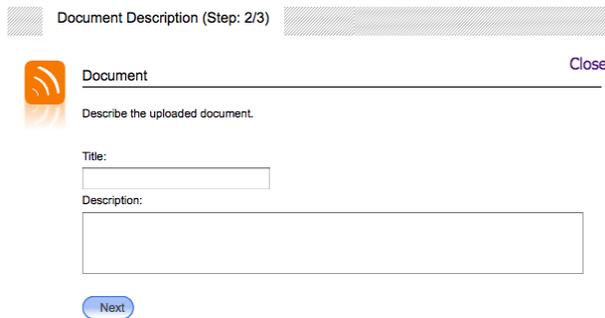


Figure 44: Upload; Title and Description

This screen allows the user to add a title and a description to the image.

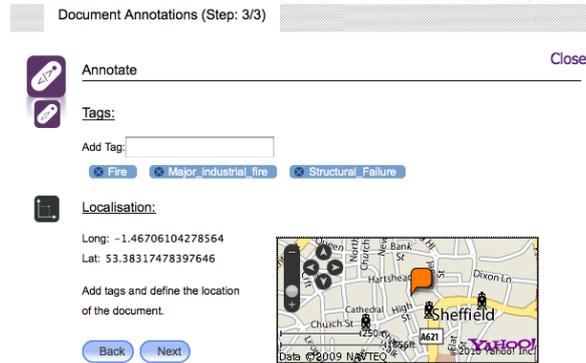


Figure 45: Upload; Tag and Localise

This screen allows the user to tag and localise the image.



Figure 46: Upload; Complete

This is the final screen in the upload process.

ER Data Collection

Data was collected relating to four incidents: Fires, Floods, Treefalls and Explosions. Only the initial three datasets were used in the evaluations; the fourth was collected for demonstration purposes. Details of the data is given below:

Data Set 1: Fire

High Severity: Fire in a shop

Sheffield: Melody Clothes shop on the Moor

1. The clothes shop in the middle of the Moor is on fire - the smoke can be seen from the top of Fargate
2. There is a large fire in the Melody shop on the Moor. It is out of control and looks dangerous!
3. There is a shop on fire on the corner of the Moor and Rockingham Gate - all the clothes have gone up in smoke!

Medium Severity: Kitchen Fire

Sheffield: Student flats on Mappin Street

1. A toaster seems to be on fire in one of the flats on Mappin Street. There are no people inside but there is a lot of smoke
2. There is a small fire in a kitchen in one of the flats opposite St. Georges Library.

Low Severity: Fire in a bin

Sheffield: Fire in a bin at the top of the moor

1. There is a fire in a bin at the top of the Moor - there is a lot of smoke but it doesn't seem to be spreading.

Data Set 2: FloodingHigh Severity: Severe Flooding

Sheffield: Flooding on the Wicker

1. The junction after Lady's Bridge is severely flooded - I can see that some cars have got stuck
2. The crossroad at the start of the Wicker is submerged with water!
3. There is an awful lot of water in the Wicker - Even the buses are getting stuck.

Medium Severity: Road Flooding

Sheffield: Furnival Road Flooding

1. The junction outside the Hilton hotel is flooded - the water is very deep though no-one appears to be in trouble.
2. There is some flooding on Furnival road, people are unable to leave the Hilton Hotel!

Low Severity: Minor Road Flooding

Sheffield: Junction at Stanley Street

1. The junction at Stanley street and Joiner street is flooded - there is a fair amount of water but cars can still get through

Data Set 3: Tree FellsHigh Severity: Major road blocked, car damaged

Sheffield: Tree blocking Hanover Way, appears to have struck a car

1. A large tree has fallen over and struck a car on Hanover Way.
2. A car has been crushed by a tree outside the Pryor building on Hanover Way - there seems to be some people still in the car
3. Hanover Way has been blocked by a tree - there seems to also be a few casualties as far as I can see

Medium Severity: Semi-major road blocked, no damage

Sheffield: Tree blocking Charter Square Roundabout, no damage but road blocked

1. The charter square roundabout has been blocked by a fallen tree - it doesn't seem to have done any damage but traffic is backing up
2. The roundabout outside Grovesnor house has a tree blocking it.

Low Severity: Tree fallen in public place, no damage, no blockage

Sheffield: Tree fallen in the square at flat street - no damage

1. A tree has fallen down in Fitzalan square - it doesn't seem to have caused any injuries

Data Set 4 : Training Data - ExplosionsHigh Severity: Explosion in a busy area

Sheffield: Explosion in a shop at the top of Fargate

1. There has been an explosion in a clothes shop at the top of Fargate - the windows of the store are covering the pavement
2. It sounds like a bomb has gone off in the H&M store in town - there is glass everywhere

- An explosion has happened on Fargate, there seems to be quite a few casualties

Medium Severity: Small explosion in side street

Sheffield: Old pub on Trippet Lane

- There has been an explosion in the Old Green Room on Trippet Lane - there is glass on the road
- That disused pub on Trippet Lane has had it's windows blown out.

Low Severity: Very minor explosion, likely small firework

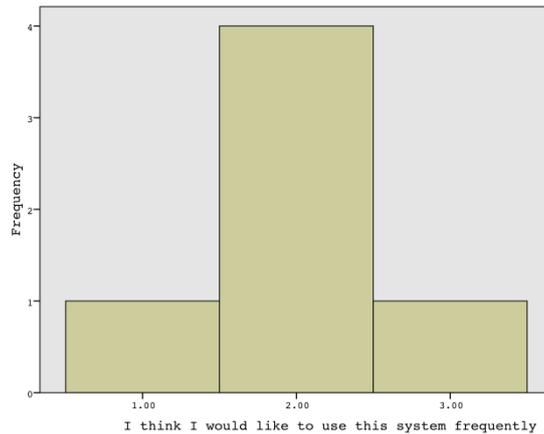
Sheffield: Minor bangs near a shop at the bottom of Fargate

- I have heard some bangs at the bottom of Fargate - it might be kids mucking about but I'm not sure.

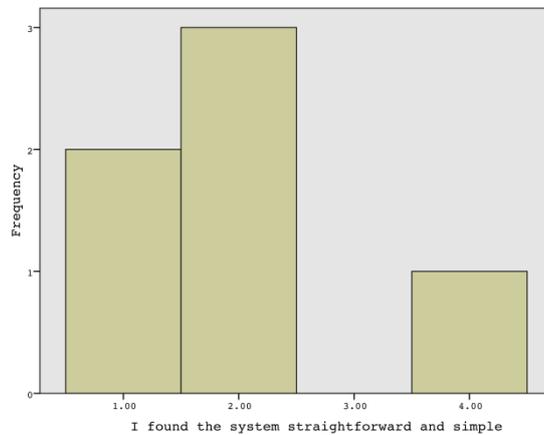
ER Experts Results

Post Experiment Questionnaire Results

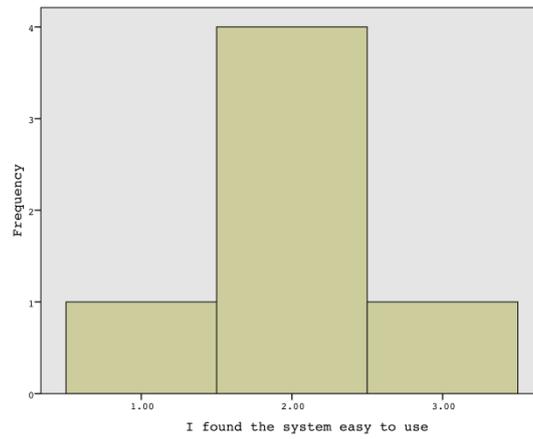
- I think that I would like to use this system frequently.



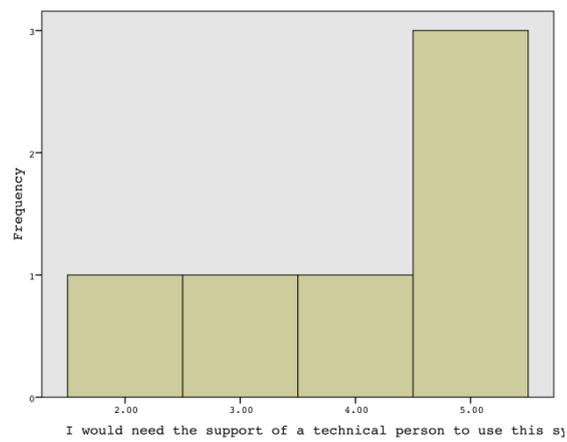
- I found the system straightforward and simple.



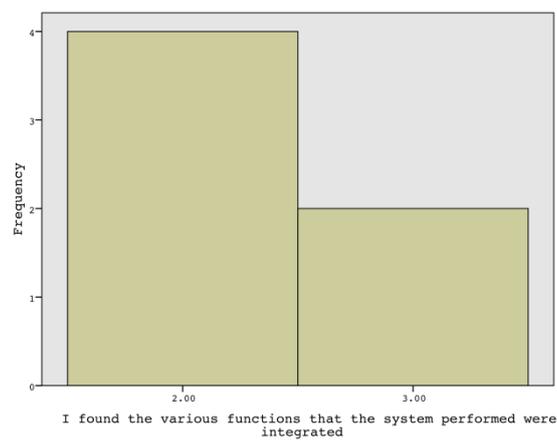
- I found the system easy to use.



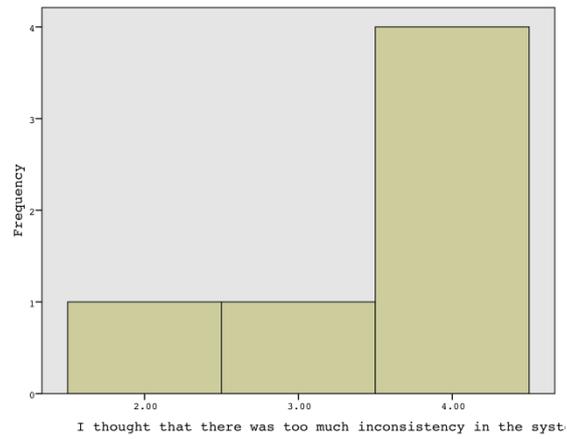
4. I would need the support of a technical person to be able to use this system.



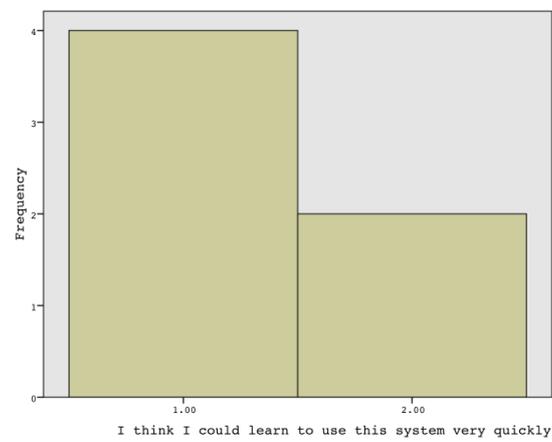
5. I found the various functions that the system performed were well integrated.



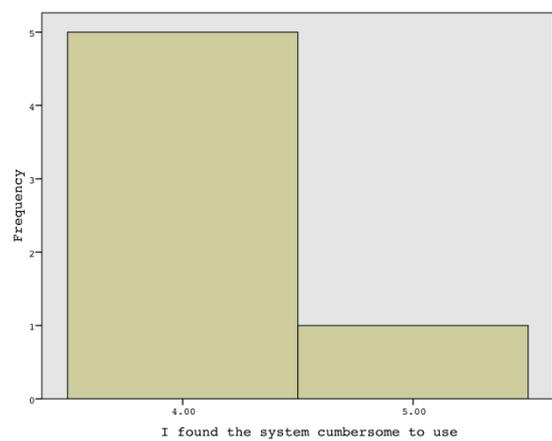
6. I thought that there was too much inconsistency in the system.



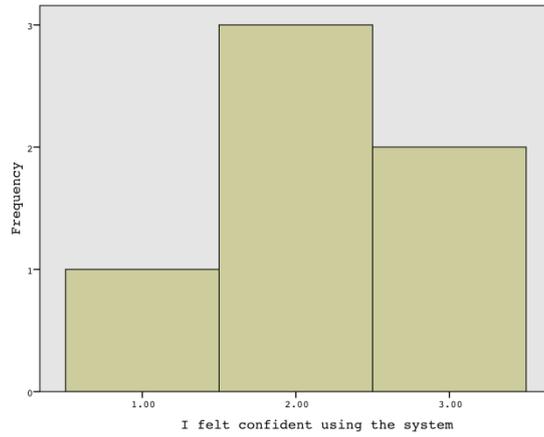
7. I think I could learn to use this system very quickly.



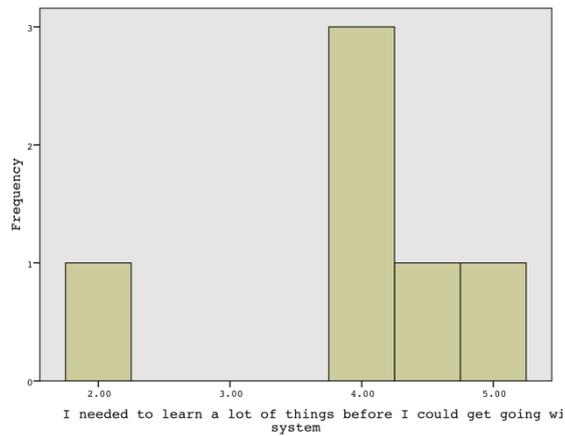
8. I found the system cumbersome to use



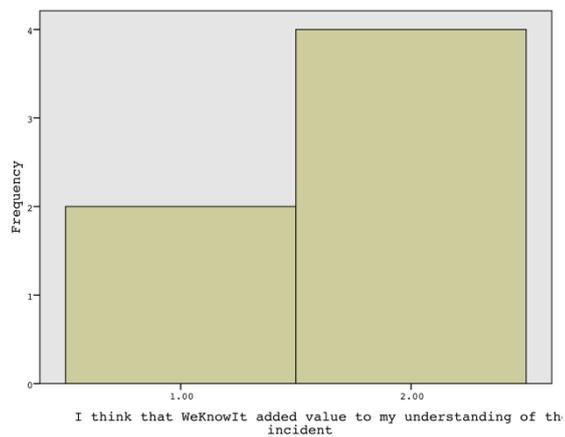
9. I felt confident using the system



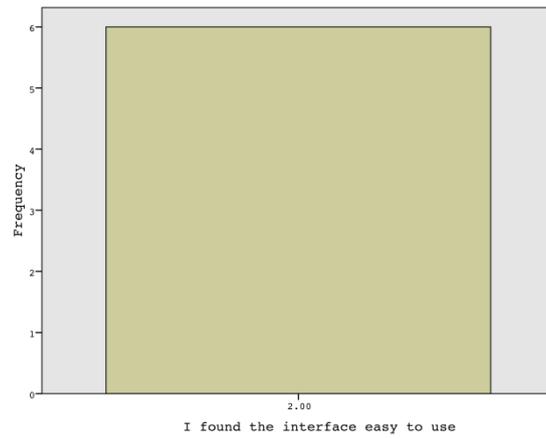
10. I needed to learn a lot of things before I could get going with this system



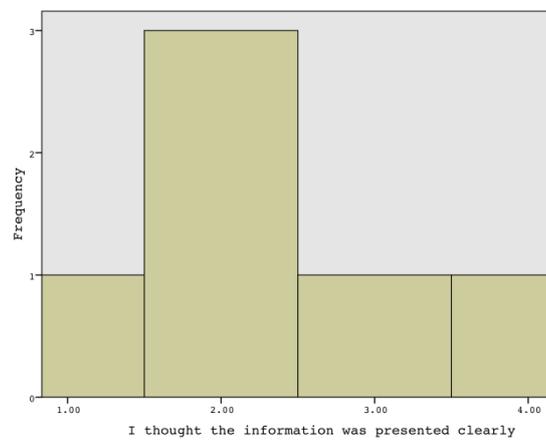
11. I think that WeKnowIt added value to my understanding of the incident



12. I found the interface easy to use



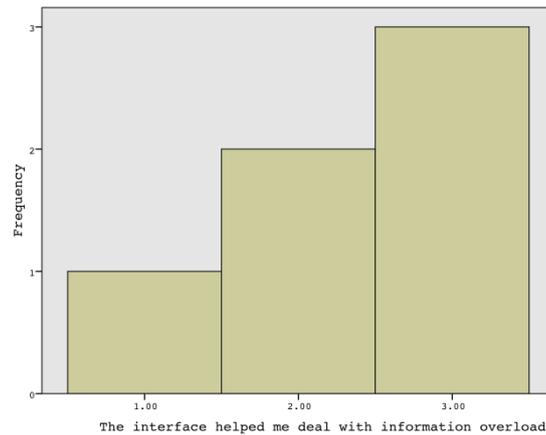
13. I thought the information was presented clearly



14. I could use the interface to focus on information I was interested in



15. The interface helped me deal with information overload



General Comments

What do you think was the most useful part of the WeKnowIt interface?

- Visual image related to a location
- Ability to correlate images with mapping
- Easy To Use
- To be able to contextualise through pictures what was actually happening
- Seeing the photos and comments together on a map and being able to select the area I wanted to focus on
- Situational Awareness - Ability to link info / image to map

How do you think the WeKnowIt interface would effect your decisions?

- Give you a visual picture of the event. Better informed.
- Could help in resource allocation and speed of response
- Yes
- A useful tool to be used in conjunction with other information coming from other sources
- I think it would aid my decision making by being able to grasp understanding of an issue more quickly
- Improve initial assessment and deployment of initial resources

What do you think was missing from the WeKnowIt interface?

- The image needs to be large i.e. when location is clicked on - image should grow to fill the screen
- Help Pages
- Nothing
- Time
- Its probably more due to my lack of understanding but it wasn't clear that there is a list of the incident at bottom of screen

What three pieces of information would you like to see but couldn't?

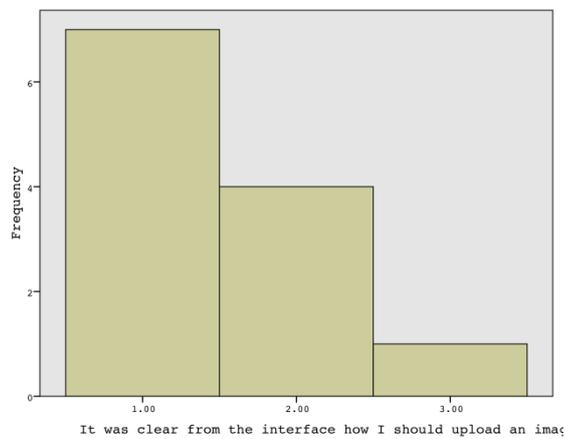
- Weather Conditions
- Ability to add info to maps - control part of incidents
- Can these be loaded into other GIS applications?
- None
- Easy To Use
- To be able to contextualise through pictures what was actually happening
- Time Table

- Compass Bearing might assists
- Buildings and road names
- Police or confirmed information shown on a map in a different colour
- Am unclear if people don't leave an exact tag whether the information would be picked up
- Live imagining - CCTV etc.
- Could be source of misinformation - filter?
- Weather - wind direction / speed etc

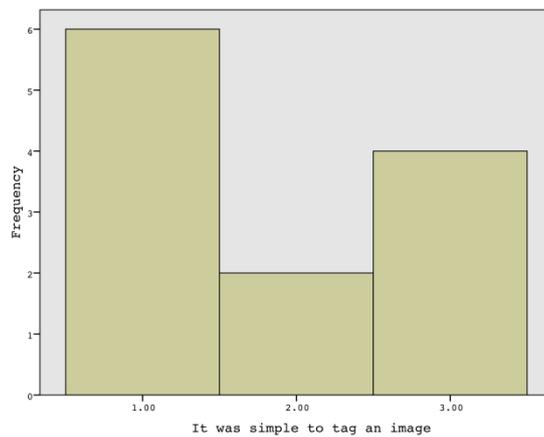
ER Citizen Evaluations

Upload Results

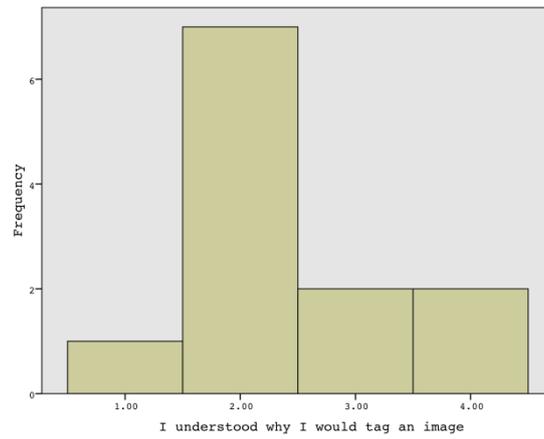
1. It was clear from the interface how I should upload an image



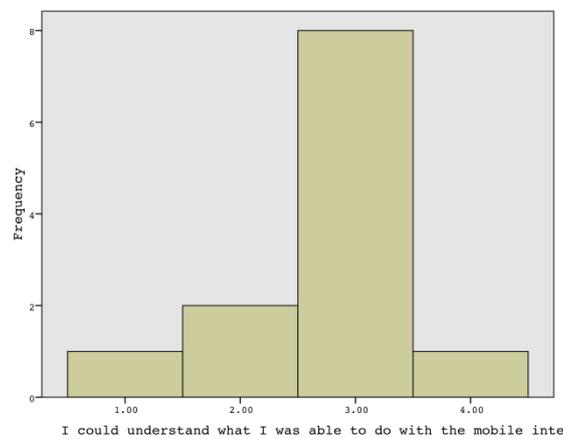
2. It was simple to tag an image



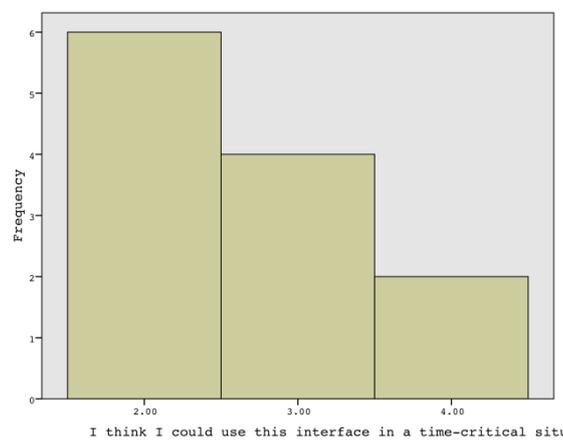
3. I understood why I would tag an image



4. I could understand what I was able to do with the interface

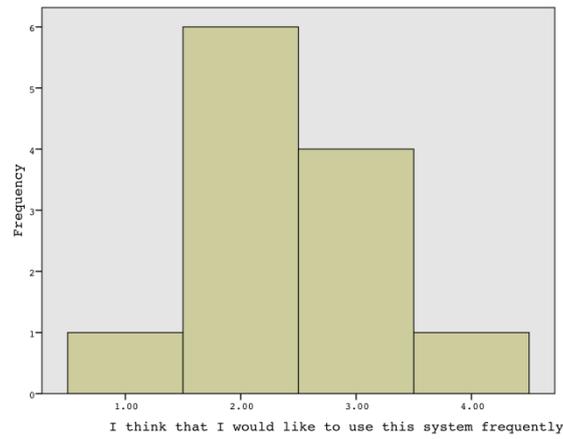


5. I think I could use this interface in a time-critical situation

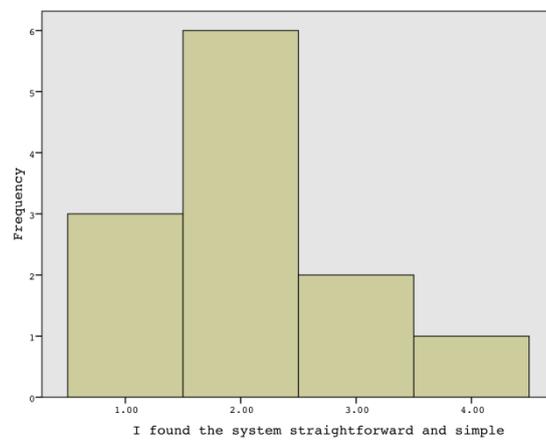


Post Experiment Questionnaire

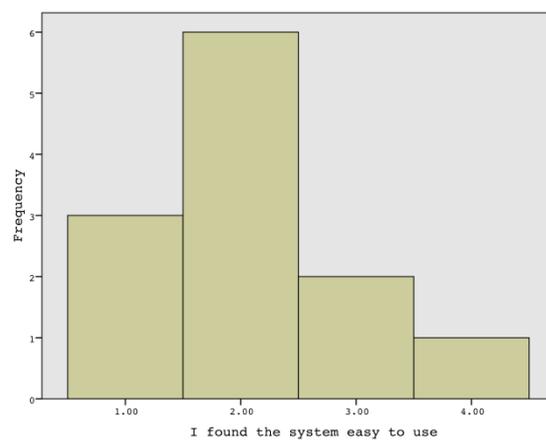
1. I think that I would like to use this system frequently.



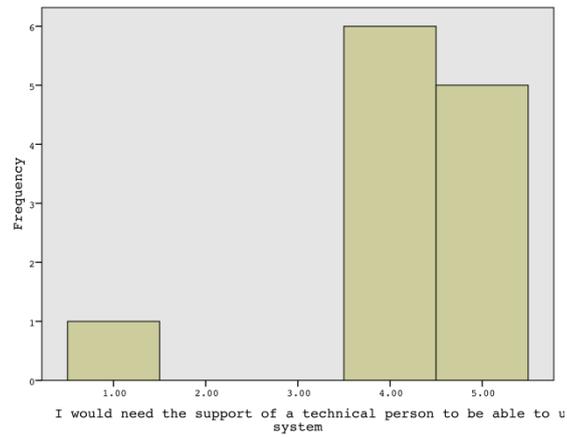
2. I found the system straightforward and simple.



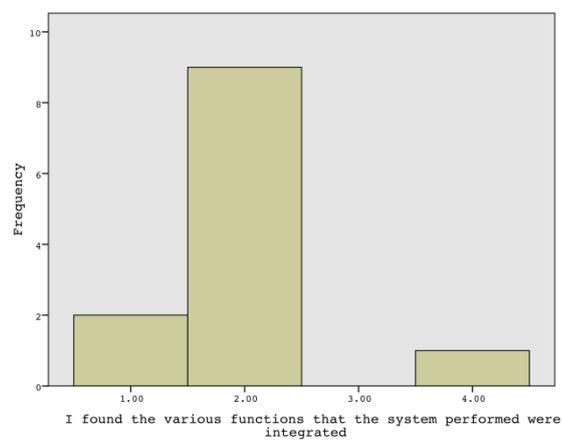
3. I found the system easy to use.



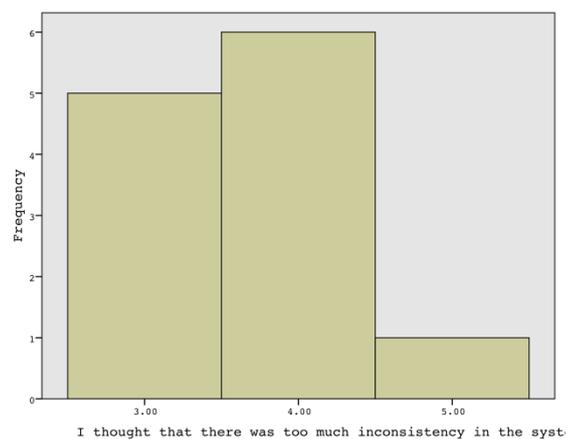
4. I would need the support of a technical person to be able to use this system.



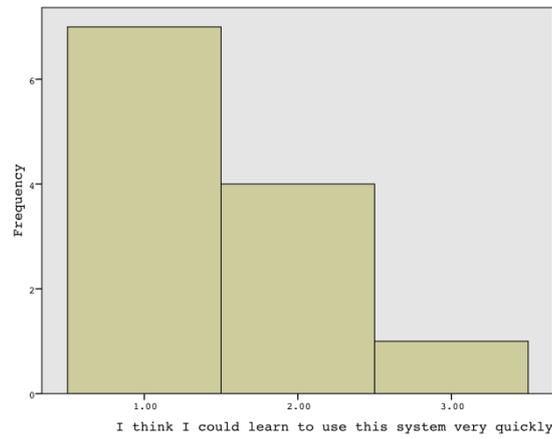
5. I found the various functions that the system performed were well integrated.



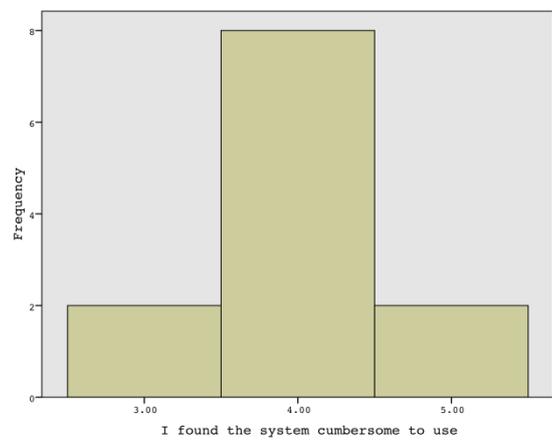
6. I thought that there was too much inconsistency in the system.



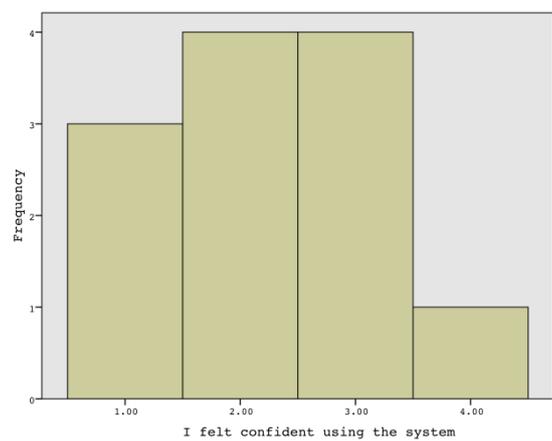
7. I think I could learn to use this system very quickly.



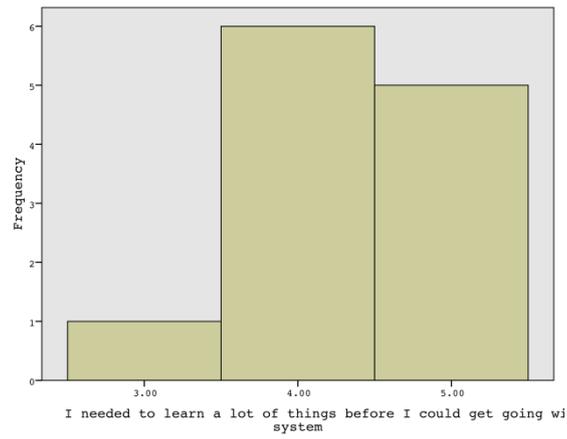
8. I found the system cumbersome to use



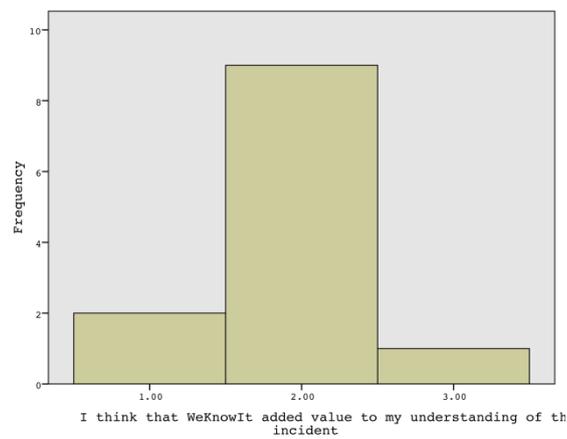
9. I felt confident using the system



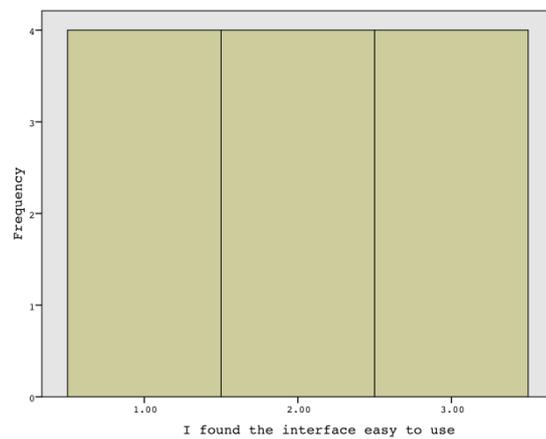
10. I needed to learn a lot of things before I could get going with this system



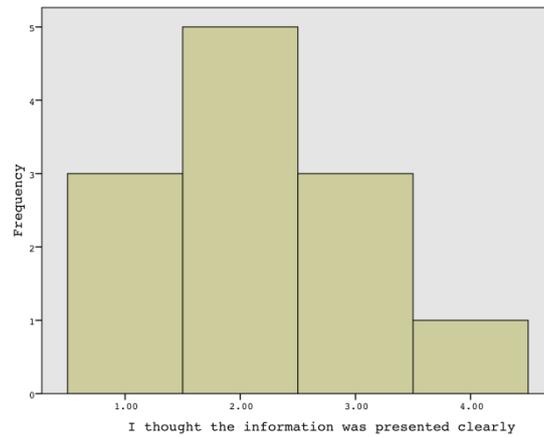
11. I think that WeKnowIt added value to my understanding of the incident



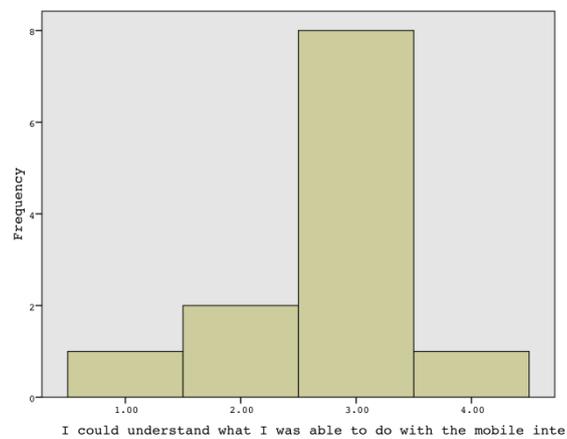
12. I found the interface easy to use



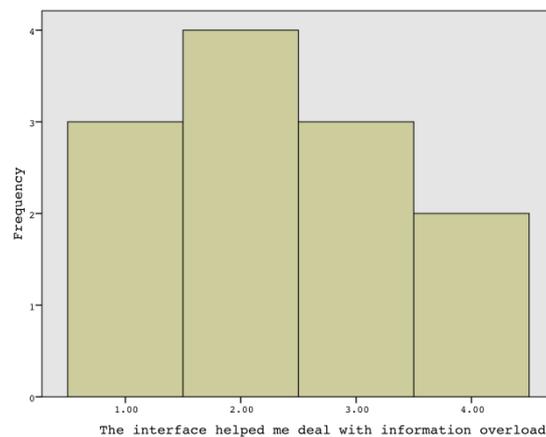
13. I thought the information was presented clearly



14. I could use the interface to focus on information I was interested in



15. The interface helped me deal with information overload



General Comments

1. What do you think was the most useful part of the WeKnowIt interface?
 - Interactive map, timeline, and tagging system allows me to browse a list of things that potentially interest me and easily narrow down the info
 - The Map - you could instantly see where the incident was
 - Linking of pictures to location
 - Tag cloud

- Map
 - Maps and corresponding images
 - Combination of information on incident, plus pictures and maps
 - Combine both comments and images so you get a better idea of incidents
 - Selecting the tags of events that interested me
2. How do you think the WeKnowIt interface would effect your decisions?
- It can help me identify useful information accurately and timely
 - You could avoid difficult or dangerous situations
 - I could avoid affected areas
 - It wouldn't
 - All alternative routes to be considered quicker
 - Help with decision to travel
 - I could use it to plan an alternate route
 - Lack of quality on comments
 - It would allow me to either prepare for potential dangers or re-plan my activities
3. What do you think was missing from the WeKnowIt interface?
- A preview function. Hover cursor over spots on the map triggers a related image, or tag to pop up
 - Addresses of the incident
 - Keyword search
 - Search
 - Hard to say from using it only once but nothing seemed missing
 - A better (clearer / bigger) map
 - Objective information as is user submitted can be easily wrong or misunderstood
 - Search box
4. What three pieces of information would you like to see but couldn't?
- Severity of incident: the information can be accurate in terms of what, where when but I don't know how bad it is but have to guess
 - Popularity of incidents: Suppose I want to find out what is the biggest thing that happened today. The system tells me what happened, but not how much attention it has attracted
 - Details of an incident: Can you link the incident to news reports or discussion by other users
 - Address of the incident
 - Pictures on the tags of incident type
 - Rating of the incidents severity
 - Bus route diversions
 - Simpler way to see time and date
 - Highlighting of current map icons
 - Preview of icon -> photo on map
 - Associate tag with visualisation (colour / text) -> next to icon
 - Related contributions
 - Maps and corresponding images
 - Combination of information on incident, plus pictures and maps
 - A hovering tab on each incident
 - Street names without having to zoom in

- Reasons
- What to do
- Objective Info
- The severity of the incident in a colour coded form
- Whether I can contact the authorities / who to contact
- User comments