SOCIETIES
FP7-ICT
Contract no.: 257493

SOCIETIES

Deliverable D8.2

Specification of First Prototype User Trial

Editor:	Patrick Robertson, DLR
Deliverable nature:	Report
Dissemination level: (Confidentiality)	PU
Contractual delivery date:	M12
Actual delivery date:	26/10/2011
Suggested readers:	Third party service developers, trial organisers and participating staff, developers, integrators
Version:	V8
Total number of pages:	97
Keywords:	Services, user – trials, trial specifications, test-beds, user experience, performance logging and metrics, trial locations, user involvement, trial expectations, evaluation

Disclaimer

Impressum

[Full project title] Self Orchestrating Community Ambient Intelligence Spaces

[Short project title] SOCIETIES

[Number and title of work-package] WP8 User Trials and Evaluation

[Document title] D8.2 Specification of First Prototype User Trial

[Editor: Name, company] Patrick Robertson, DLR

[Work-package leader: Name, company] Mark Roddy, Lake Communications

[Estimation of PM spent on the Deliverable] 10

Executive summary

A particular focus of the SOCIETIES project is the development of a robust, scalable and user-friendly platform and to evaluate through strong involvement of end-users, the usefulness and acceptance of the developed CSS software via three user trials in all three user groups: the student community, disaster management and the enterprise user group.

After the paper trials described in D8.1, the first user trial with a platform prototype and first 3rd party services is the focus of the present deliverable. To provide useful input to the second user trial with the final, scalable platform, great importance must be placed on the testing task, especially at trial one. Only in this way, we can ensure that the system produced and demonstrated at the end of the project is of the highest quality and answers the basic value propositions of the project.

The main goal of the SOCIETIES trials is to allow us to evaluate the developed technology and to validate some assumptions about the developed technology. The trials should allow us to:

- *Ask the right questions*: Part of the planning of the trials is to find the precise list of research questions we want to ask, and design the trials in a way that will allow us to ask these questions and get answers from the users and by other means.
- *Observe the right behaviour*: The trials should allow us to put the users in situations that we consider as useful, and observe, through different means, what the users do (objectively) and think (subjectively). Putting users in the right situations means that the right technology needs to be developed and tested prior to the trials.
- *Collect the right data*: Based on the research questions we want to ask, trial planning is to ensure that at the end of the trials we have collected the right data to evaluate and validate in a methodologically sound way. It will be too costly for the project to repeat the process.

Taking as input the scenarios from WP2 and the paper trials from WP8, the technical work-packages 4 and 5 were asked to produce a list of research and technical objectives, which they prioritised for evaluation in the first user trial of the early demonstrators. This list of objectives was validated by the Research & Technical Co-ordination team, to ensure that the core trial objectives (for example value propositions) had been fully captured. In a first attempt at defining and anchoring formal performance tests we have taken the resulting evaluation points prioritised them, linked them to individual tests, and furthermore related them back to all in WP4, WP5, and WP6. The result is a denser network of feedback from the envisioned tests to the work packages, in this critical stage in the build up prior to the first trials.

The main body of this document has shown that the specifications of the three trials will differ to a certain extent between the user groups. The objective of the student trials is a long term, robust usage of the platform of technology expert users, the disaster management users have a particularly high demand for user-friendliness, fault-resistance and autarkic usability, while the enterprise user group focuses on integrating the platform with existing systems for day-to-day operations featuring improved exchange and cooperation for employees. However, there are a number of trial-wide *common* aspects that benefit from three trials/groups shedding light on them.

List of authors

Company	Author
ITSUD	Zhu Wang
HWU	Fraser Blackmun
SINTEF	Babak Farchchian, Jacqueline Floch
DLR	Patrick Robertson, Korbinian Frank
Lake Communications	Mark Roddy
Intel	David McKitterick

Table of Contents

1 Introduction

A particular focus of the SOCIETIES project is the development of robust, scalable and user-friendly platform and to evaluate through strong involvement of end-users, the usefulness and acceptance of the developed CSS software via three user trials in all three user groups: the student community, disaster management and the enterprise user group.

It is foreseen to start with a minimal, yet robust system and gradually increase the extent of the provided functionality over the duration of the project. This approach should provide both qualitative and quantitative results in such aspects as usability, utility or robustness of the system as it evolves.

After the paper trials described in D8.1, the first user trial with a platform prototype and first 3rd party services is the focus of the present deliverable. To provide useful input to the second user trial with the final, scalable platform, great importance must be placed on the testing task, especially at trial one. Only in this way, we can ensure that the system produced and demonstrated at the end of the project is of the highest quality and answers the basic value propositions of the project.

This deliverable now makes a first specification of the trials, the involved actors, temporal and hardware aspects, as well as the definition of 3rd party services necessary for testing and the definition of concrete tests. It shall analyze which components are necessary for the trials, the modalities and the set up necessary for the user trials using the first prototype.

These specifications will differ to a certain extent between the three user groups, as the objective of the student trials is a long term, robust usage of the platform of technology expert users, the disaster management users have a particularly high demand for user-friendliness, fault-resistance and autarkic usability, while the enterprise user group focuses on integrating the platform with existing systems for day-to-day operations featuring improved exchange and cooperation for employees. However, there are a number of trial-wide *common* aspects that benefit from three trials/groups shedding light on them.

A final and updated release of this deliverable (M18), closer to the actual trial period, will finally provide more details on the components of the technical work packages WP4 and WP5 involved in the tests. As their design is still in progress, no *final* conclusions about them can be drawn for the moment. However the importance of this first version of the deliverable is to set clear goals for the implementation of the third party services and the test bed which play key roles in providing robust added value for the users in the first trial.

The remainder of this document is divided into 5 chapters:

- Chapter 2 describes in detail the methodology to use for the testing.
- Chapter 3 will specify the trials for the student community.
- Chapter 4 will be devoted to the trials for the user groups involved in disaster management.
- Chapter 5 will specify the trials in the enterprise user group.
- Chapter 6, finally, will conclude the findings of the first chapters and give an outlook on the second version of this deliverable.

2 SOCIETIES First Prototype Trials Roadmap and Methodology

2.1 Overall Trials Planning and Organisation within the Project

The main goal of the SOCIETIES trials is to allow us to evaluate the developed technology and to validate some assumptions about the developed technology. The trials should allow us to:

- Ask the right questions: Part of the planning of the trials is to find the precise list of research questions we want to ask, and design the trials in a way that will allow us to ask these questions and get answers from the users and by other means.
- Observe the right behaviour: The trials should allow us to put the users in situations that we consider as useful, and observe, through different means, what the users do (objectively) and think (subjectively). Putting users in the right situations means that the right technology needs to be developed and tested prior to the trials.
- Collect the right data: Based on the research questions we want to ask, trial planning is to ensure that at the end of the trials we have collected the right data to evaluate and validate in a methodologically sound way. It will be too costly for the project to repeat the process.

These three questions are related to the three phases of Scenario-Based Design (SBD) as pointed out by Rosson and Caroll (2002) and shown in the figure below.

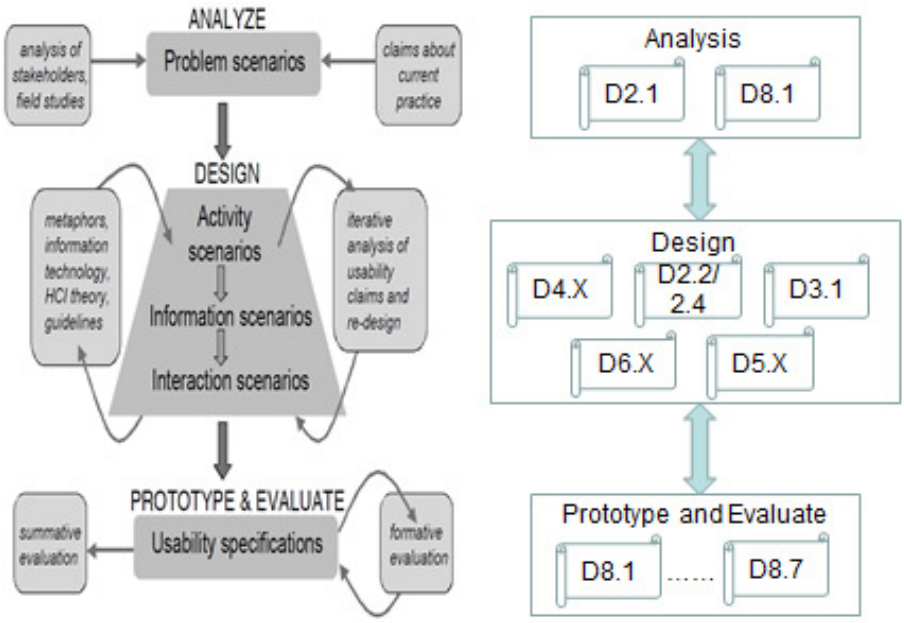

Figure 1: The three phases of SBD (to the left, Rosson and Caroll 2002) and how they related to SOCIETIES deliverables (to the right)

Analysis allows us to think about the right questions to ask. A large part of the analysis in SOCIETIES is done in WP2 (through questionnaires, state of the art survey) and in D8.1 (through paper prototype evaluation). Gradually we have collected a list of research questions that constitute the basis for evaluation and validation.

Observing the right behaviour is enabled by having the right technologies made available to the users, and by exposing the users to the envisioned scenarios in SOCIETIES. The scenarios were defined by WP2 and revised by Task T2.4 based on paper trial evaluations. Currently the project is in the process of implementing the technology that will be used in the trials.

Collecting the right data is performed during the trials. This deliverable is mainly about specifying how the trials will be organized, both in practical terms and in terms of designing research to collect the right data.

Although SBD provides a framework for designing new technology, the research methods employed in the various phases are not strictly specified. SOCIETIES has already made some choices about some of these methodologies but some are still open to discussion. This is illustrated in the figure below.

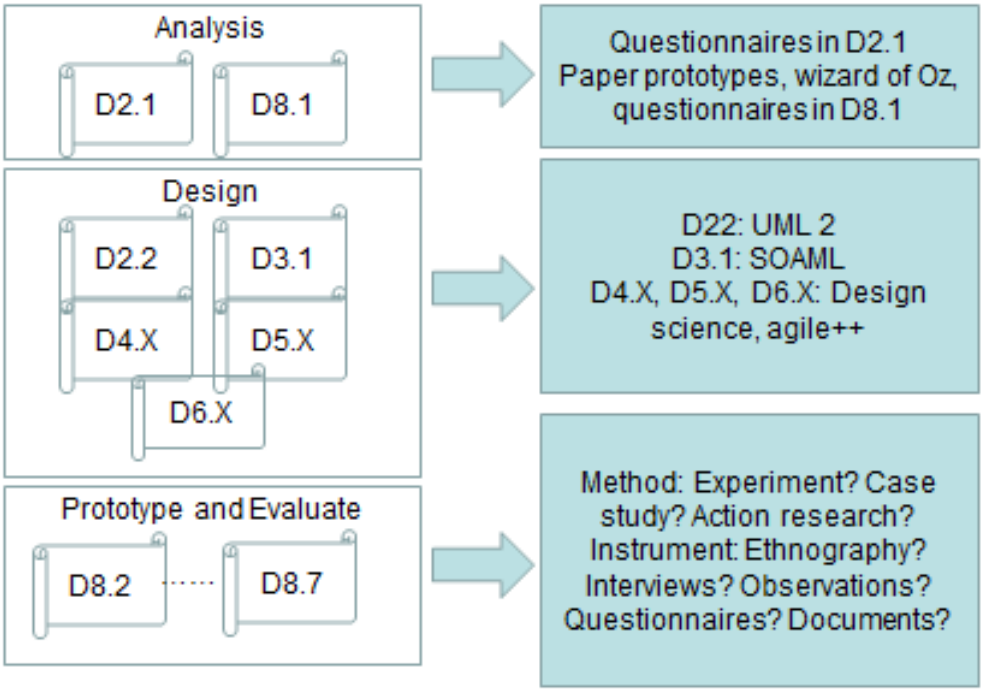

Figure 2: The various research methodologies being used or decided in each phase of SBD

The analysis phase used a combination of contextual methods such as paper prototype evaluation and wizard of Oz evaluation. We also complemented these methods with questionnaires, interviews and open-ended discussions with the participants.

The design phase uses a scenario-initiated development process where scenarios are defined based on the analysis of the end-user involvement and are used as the basis for the design of the technology for the trials (SOAML and UML 2). The technology implementation will most probably follow different variants of agile methods.

The methods that will be used during the trials are being selected as part of this deliverable and future WP8 deliverables. The three different trials will require different methods because of their different nature. For instance, the student trial will span a longer period in time and can make use of ecologically valid and contextual methods, while e.g. the first disaster management trial will resemble a controlled experiment. Our goal is to define the set of questions to be asked, and the set of data to be collected. Based on this input we will be able to make more informed decisions about specific research methodologies to be used.

2.2 Conceptual framework for research design

In addition to the methodological framework discussed in the previous section, we have also defined a conceptual framework for aligning the central concepts and the knowledge to be created by the project. The following figure shows an overall view of these concepts:

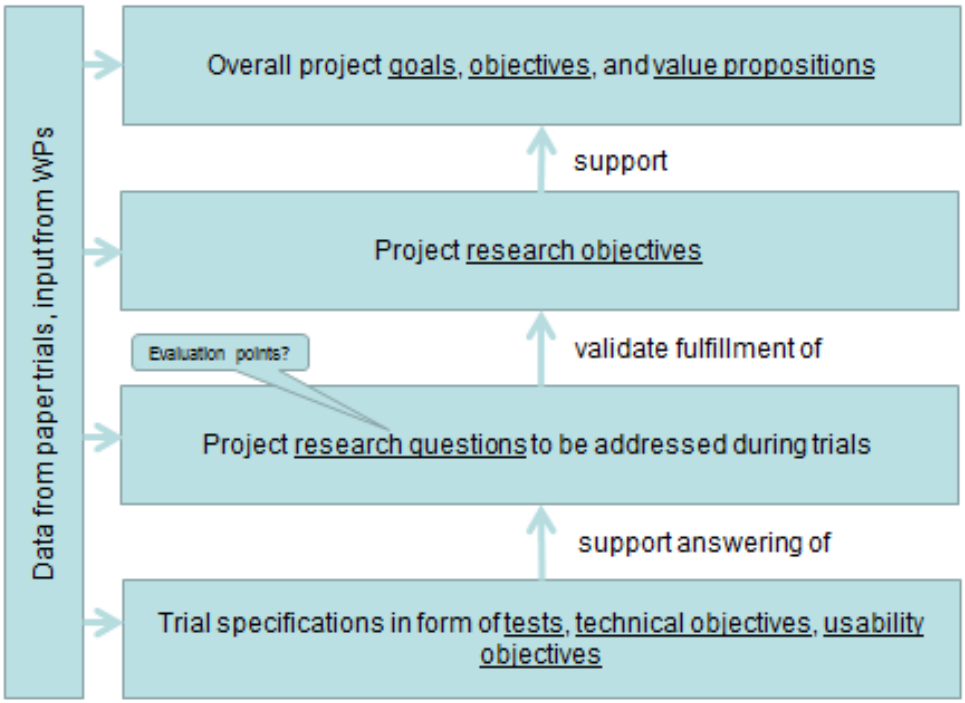

Figure 3: Conceptual framework for aligning knowledge-related concepts

The top-most concept is used to accommodate for project-wide goals, objectives and value propositions. The input at this level is mainly from the description of work (DoW) and from an on-going discussion in the project about making project value propositions clearer. An example project objective is: "*To facilitate the creation, organisation, management and communication of communities via CSS/CIS, where pervasive computing is integrated with social computing communities*".

Based on the top level of project goals, the project has defined a set of research objectives. Research objectives are mainly topic areas that are covered by the objectives of the project. The project objective example above will have a research objective such as: "*Develop CSS/CIS platform to facilitate creation of communities*". Research objectives are also largely defined by the DoW.

Each research objective is then decomposed to a set of research questions. These are research questions that we want to answer during the trials. Research questions will produce evaluation data for the developed technology, and will validate (or not) our assumptions. Examples of research questions related to the above research topic are: "*What are the costs/risks for our users (in terms of privacy, system training, system errors,….)?*" or "*What are the best user interaction metaphors for representing our major concepts (e.g. CSS, CIS)?*".

Once the research questions are defined, the definition of tests, technical objectives and usability objectives will finalize the specification of each trial. Tests are mainly description of functions that a user should be able to access. Each test will collect some data about a research question. E.g. if we want to know about a privacy mechanism, that mechanism needs to be testable by the user during the trials. Some tests are related to platform functionality while some are mediated through 3.party services that the project will develop (WP6). For each trial we also have technical and usability objectives based on e.g. the type of the users involved or the number of the users.

The above concepts are being integrated into our SBD framework using a uniform notation (UML 2), which is widely used in the project. Tool support is important in order to be able to trace research questions, to prioritize, and to manage the evolution of the software. Figure 4 shows how the above concepts are being codified in our UML 2 tool for easy management and follow-up. To the left in the figure is the project goal, then one of the objectives is to support that goal, which is then followed by a number of research questions related to that objective. To the right we see the technical artifacts (platform components and 3rd party services) that will enable answering the research question during the trials.

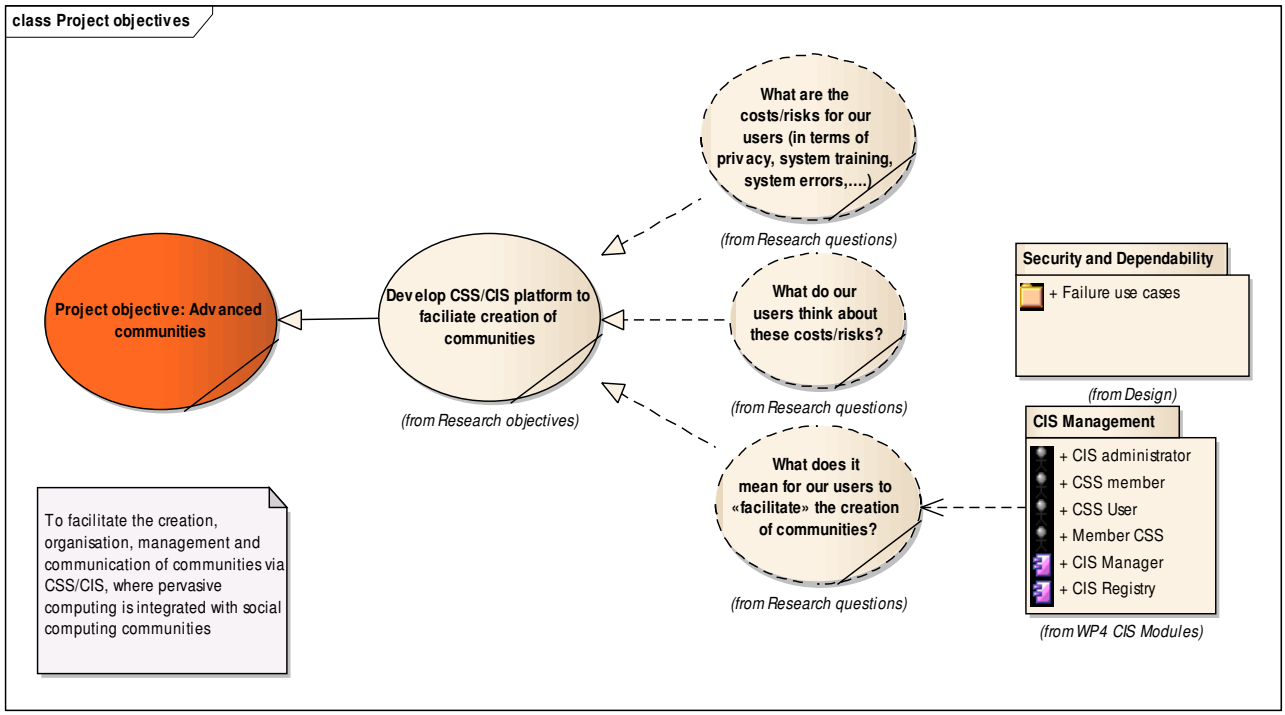

Figure 4: UML representation of project objectives and their projection onto key research questions from an end-user point of view

2.3 First Trial Objectives

Overall, D8.1 suggests the following three main findings for the evaluations of the Societies project:

1. The project wants to **find relevant people/members for a CIS**, data mining to find relevant people/things. These are the core values. Related to that, the **privacy concerns** of users have to be taken into account.
2. **Automation**: Our paper trials suggest that users are sceptical towards automation. They accept automation, but want clarity in its nature.
3. Across all three groups, esp. DM domain: can see **use of the backend community**, but see a real problem with trust. Use existing trust regimes. We should build protocols that leverage trust relationships

The individual objectives of the first trials in the three user groups are given in sections 3.2, 4.2 and 5.2. They differ, as the trial settings have been chosen differently from user group to user group in order to give the evaluation different focuses.

2.4 Methodology for Creation of Evaluation Tests

In our Paper Trials in April and May of 2011 we presented to our User Communities low-fidelity graphical representations of our storyboards. During these presentations a number of questions were inserted that related to the storyboard scene, and we asked our users to give us their individual feedback to these questions. We summarised these responses in Annex B of the deliverable D8.1 “Paper Trial Evaluation Report”, which were then passed over to task 2.4 to revise our original scenarios accordingly see Figure 4 below.

The technical work-packages 4 and 5 were asked to produce a list of research and technical objectives, which they prioritised for evaluation in the first user trial of the early demonstrators. This list of objectives was validated by the Research & Technical Co-ordination team, to ensure that the core trial objectives (for example Value Propositions) had been fully captured.

Task 8.2 used these inputs to build and prioritise the test tables in sections 3.3.1, 4.3.1 & 5.3.1 and in sections 3.11.1, 4.11.1 & 5.11.1. Methodologies for the creation of these tables has been detailed in these sections.

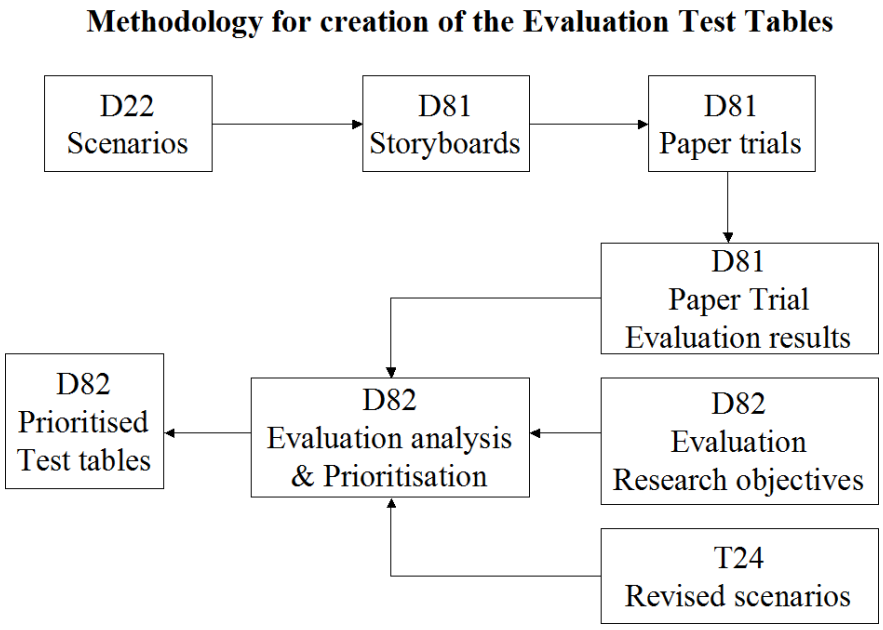

Figure 5: Inputs and Logical Flow of the Evaluation in D8.2

2.5 Critical issues

It appears that the project has a tricky "conflict of interests" situation to navigate, but these are exactly the challenges that call for research in this area of social and pervasive computing.

On the one hand the SOCIETIES project appears to need a "positive" response to proposed technologies from users to give the project permission to 'go-ahead' to develop technologies, such as automation, sharing of preferences etc. So to date our work has an emphasis on 'good' scenarios that deliver value to end-users. Whilst, on the other hand, for good user centred design, and also for ethical reasons, we should allow users every opportunity to be aware of all of the potential risks involved in engaging with these novel technologies so that they can make informed choices and give well balanced feedback, so we should give equal emphasis to 'negative' scenarios.

In D8.1 the methodology used, was to a large extent determined by uncertain and limited user access. If we are to truly engage users, how are we going to motivate them? Our users are not stakeholders in the sense that they would be motivated by having input to technology changes that would be otherwise solely input by management in the case of traditional Computer Supported Cooperative Work (CSCW). Longer user studies, than those feasible in a project initial phase (see D8.1), would be required to begin to evaluate trust for example, and such a study would also be better for some user engagement before, and after, as well as during the trial, would require significant and dedicated user input. However, we feel that these questions regarding trust can be better addressed and answered when users are able to interact with real services and features, and within the context of their human interactions.

A question that might also be asked is if we need to employ and reward, some key users to act as lead informants and distributers of knowledge. This, however, would raise the question of how to handle the resulting power relationships.

To date our engagement with users has to be leveraged off trusted human relationships. While these relationships have allowed us access to users, we should be aware of a previous power relationship that is also inherited into the roles of researcher and user in a trial setting. This is particularly pertinent in the case of the student group, where the students are giving responses to people who are perhaps lecturing them or seen as important in the faculty. Students are already attuned to trying to give responses to impress lecturers, i.e. to give them the answers they want, in order to achieve good grades, there is a risk that this may

unintentionally affect and bias the results of our trial. We should therefore, be aware and consider how these pre-existent relationships have the potential to influence results, and if we need to mitigate against this, in our selected user engagements, and in our documents.

The rest of this document will present our picture of the first trials with these critical issues clearly in mind. It is envisaged that these critical issues be revisited in the final version of this document, and in the trial evaluation process.

3 Student Group Trials

3.1 Summary of the trials

The student user group trials will take place at the Heriot-Watt campus in Edinburgh, Scotland. They will involve various groups of students from the Computer Science and Information Systems courses at Heriot-Watt. One confirmed user group consists of all the Computer Science and Information Systems students who are in their 3rd year come October 2012. Their trial is expected to last for at least three months from October to December, and possibly continuing until late in the academic year, which could be as far as June 2013. It is possible that other user groups will be trialled at the same time, such as two separate groups of first and fourth year students, but this is not confirmed. Also, more student groups may be targeted in the following academic year from October 2013. This has yet to be considered in detail, however, although if it does happen, it is likely that the confirmed user group of third year students from 2012-2013 will be used. They would then be in their fourth year, where they would be undertaking dissertation projects, which they may be encouraged to use their CSSs for.

For the student trials, every student in a given group will receive a CSS device (e.g., a smartphone) which provides all the individual user functionality of a SOCIETIES CSS. Each group member will have the opportunity to communicate with the CSSs of fellow group members, as well as create, join, and send/receive invites for CISs. Another key aspect of these trials is that students will be given the tools needed to create their own third-party services, or 'apps', for their CSSs, which can in turn be shared with the others both individually and through CISs.

Aside from some confirmed details described above, the exact details for what groups will be involved, and what times their trials will take place, is still under discussion.

The storyboard of student group has been presented in the first user evaluation (i.e. the paper trial documented in D8.1).

3.2 Objectives of the Student Group trials

Objectives of the first Student Group evaluation user trial:

1) Research Objectives: where we construct demonstrators that can be used to measure the value of our system.
2) End User Objectives: where we evaluate the usability of our system from our users' perspective.
3) Technical Objectives: where we implement integrated prototypes of the SOCIETIES system so that it can be evaluated from a practical perspective.

3.2.1 Research evaluation objectives for our Student Community

- *S-RO1: To extract context rich data sets so as to create communities of interest.*
- *S-RO2: To investigate how interaction with physical resources, using physical context can be used for initiating and supporting social interactions.*
- *S-RO3: To investigate how are relevant members (people) of a CIS discovered.*
- *S-RO4: To investigate how we can address trust issues, in particular, how student end-users trust the integrity and accuracy of the services.*
- *S-RO5: To understand how much automation is accepted by end-users and how the end-users trust this automation.*
- *S-RO6: To investigate the characteristics of communication and collaboration in a university scenario.*

- *S-RO7: To understand the impact of privacy concerns to the functionality of the system.*
- *S-RO8: To investigate how we can provide personalization services, how users accept the provided services.*
- *S-RO9: To get logging data for performance metrics and better understand the user in the context of their use of the services.*
- *S-RO10: To extract user-feedback for the field of social computing.*

3.2.2 User evaluation objectives for our Student Community

- *S-UO1: To investigate what levels of usability and what user interaction metaphors should be supported.*
- *S-UO2: To evaluate an easy-to-use privacy layering interface.*
- *S-UO3: To evaluate the end-user experience and user-acceptance of the presented system.*

3.2.3 Technical evaluation objectives for our Student Community

- *S-TO1: To investigate the level of interoperability constraints of the software components within real-world environments.*
- *S-TO2: To evaluate the performance and scalability of the designed system (but not solely of the implemented prototype).*
- *S-TO3: To understand the user's desire for future applications and technical solutions of the presented system.*

3.3 Analysis of the Student Group Paper Trial Evaluation

The Student Group have taken part in a two-part Paper Trial involving an initial storyboard session followed by an immersive environment session.

3.3.1 D8.1 Student Group Storyboard Responses

During the initial storyboard session, we presented the student storyboard slides to a group of 15 students. Throughout the presentation the students were asked a series of questions, each relating to some concept that had just been presented and the students responded in a multiple choice fashion using their keypads to input their responses. After all of the slides had been presented, the students were asked which of the scenes was the most memorable.

A data set was collected and analysed to form the storyboard responses. We summarised these responses in Annex B of the deliverable D8.1 “Paper Trial Evaluation Report” and used these responses to build the following tables.

- The first column gives a reference to the scenarios, which were detailed in deliverable D2.2 “Scenario description, use cases and technical requirements specification”.
- The second column gives a summary of the paper trial feedback from our users for the referenced scenario.
- The third column presents a summary of the recommendations from the Student Group Paper Trial, which was relayed back to the consortium.
- The fourth column gives the unique identifier for the scene/the table row, marking it as an input to the future User Trial tests, and the number of this identifier maps directly to the identifier for the corresponding User Trial test relating to that scene (so input I1 is the input for test S1, etc.)..

- The final column lists the evaluation objective references (identified in section 1.2) that relate to that row's test input.

Source in D8.1	Summary of user feedback from D8.1	Summary of D8.1 discussion and recommendation	Input ID	Objective Reference
US_01 Scene 1	The majority of students answered positively to all the questions from this scene. They would join a Freshers community, they would like to be told information about things, such as where Freshers tend to eat, they would agree to being automatically added to a community for their Degree course, and they would accept being forced into special communities in emergency situations.	This indicates the usefulness of automation services under some specific conditions, such as study and emergency situation.	I1	S-RO5 S-RO7
US_01 Scene 4	Most students would be interested in going to an event like the proactive disco. Most students actually would not mind if their music preferences were shared for something like the proactive disco.	This forms part of our system's ability to create "context rich data sets" and we could test this to see how much data they would be prepared to give us.	I2	S-RO7 S-RO8
US_01 Scene 6	A few students reject the usefulness of a cooking service outright, but not everyone is sure they would find it useful, as only slightly more students answered 'Yes' than 'Maybe'. Students can think of at least some areas where they would benefit from a community service. Most, however, only see a few such areas, so they probably would not use lots of different community services.	It's the actual service that needs revision, in terms of usefulness, not the notion of community services, which is clearly seen by the results as having benefits to our student community. The positive response to Q8 of the questionnaire clearly indicates that the student community saw value in community services; it was just that they did not see value in the actual "cooking service".	I3	S-RO8 S-UO3
US_01 Scene 7	This is the first point where the majority reject a pervasive concept: the students generally would not like services to start automatically on their behalf, even if the system is sure the service would be beneficial. The students would want to give confirmation before the service started. Most students would appreciate suggestions to introduce themselves to new people based on shared aspects, although many of them would still prefer making friends traditionally. On the opposite side of things, most students would have no problem with	The area of automation around services needs further study. This can be tested using control groups in our first field trial. The usability of this automation needs to be tested, because it is one of the "corner stones" in pervasiveness, whereby the technology should be seen to disappear into the background. We need to ascertain why in some instances auto-joining a community is acceptable (as in scene 1) but auto-starting a service is not.	I4	S-RO4 S-RO5 S-UO1 S-UO3

	somebody else being the one to receive the suggestion to introduce themselves, as long as the student was in a happy mood.	Auto-starting a service is a key feature in our system. We need to understand the usability constraints about this from our users point-of-view.		
US_03 Scene 7	The vast majority of students would join a job-oriented community like the Student Shopping Service, if it were available. Nobody said they would reject it outright. The response would perhaps suggest students would be interested in ways that SOCIETIES could help them find work or make money.	We should test this	I5	S-RO6
US_03 Scene 4	Students were asked if they would like to be invited to join a sub-community based on shared criteria, to which most were favourable and nobody rejected it. Then they were asked if they would like to be automatically added to sub-communities, and here the majority say they would want to give confirmation first. This is in contrast to the answer given in scene 1, where the students would want to be automatically entered into the Degree course community.	Discuss further with a view to possibly extending. **The scenario description is fine but the project needs to consider the usability surrounding auto-joining a community on the user's behalf. This needs to be tested using control groups etc in the first user field trial.** There is a conflict in the results from this scene and the first scene, in terms of the system auto-joining a community on the users behalf.	I6	S-RO1 S-RO3 S-RO4 S-RO5 S-RO7
US_01 Scene 5	There is a split of opinion as to whether the device should predict their behaviour and make appropriate suggestions based on this. By and large the student community would support system monitoring, in order to make better suggestions based on behaviour, and most approve of it (assuming any data is stored privately and securely). The majority of students would find the automated actions that would come from predictions useful in daily life.	Given the split of opinion as to whether the device should be able to predict user behaviour, we might want to consider testing this further. We need to fully understand the implications for this split of opinion.	I7	S-RO1 S-RO4 S-RO7 S-RO8 S-RO9
US_04 Scene 5	Almost all students felt they would appreciate technology to help them with their studies by allowing them to join communities dedicated to it. Students would therefore seem to like things that help them with their academic life, in addition to helping them find jobs as shown in previous scenes.	We should develop such kind of services and test them	I8	S-RO6

3.3.2 D8.1 Student Group Wizard of Oz Responses

The immersive trial took the form of a 'Wizard of Oz' role-playing session, where the trial participant acted out a scripted day in the life of a hypothetical SOCIETIES user, while being asked questions on what they

were seeing at certain points. The trial took place in a specially designed area at Heriot-Watt University, which was set up using large cardboard stands and tables to create a path for the user to follow. Each of these trials took between 10 and 15 minutes to complete, and 13 trials took place in total over a 2 week period from the 9th to the 19th of May 2011.

A data set was collected and analysed to form the Wizard of Oz responses. We summarised these responses in Annex B of the deliverable D8.1 "Paper Trial Evaluation Report" and used these responses to build the following tables.

Source in D8.1	Summary of user feedback from D8.1	Summary of D8.1 discussion and recommendation	Input ID	Objective Reference
US_01 *Scene 7* **Wizard of Oz :** Yes **Hotspot: 5**	The response to the question in the "Wizard of Oz" trial is especially interesting in that it appears to go against the response to a similar question in the storyboard event, where most students indicated that for any automated decision, no matter how sure the system was about it, they would like to be informed beforehand and to give approval before the decision was taken. It should be noted though that the question primes our participants by saying "To minimize pop-ups..", which is clearly an incentive to suggest that they should trust the system so as to minimize pop-ups.	The new scenario needs to be described in a way that shows that initially there are pop-ups but that over time the system will learn to take decisions on users behalf, but only if the user wants this to happen. This is another potential system usability issue and needs to be tested in future trials	I9	S-RO4 S-RO5
US_03 *Scene 1* **Wizard of Oz :** Yes **Hotspot:** 1	The majority of participants are at least receptive to being informed about the actions of others in their pervasive communities, and making use of this information in forming their own decisions. Some of those who responded 'No' to the clothing questions explained that they did not think clothing was something that would influence them, and the fact that nobody responded negatively to the more general question shows they were all receptive to the idea of information on others' actions influencing them in their own lives.	We should develop such services and test them	I10	S-RO6 S-RO8 S-RO9 S-RO10
ES_01 *Scene 3* **Wizard of Oz :** Yes **Hotspot:** 2	Almost all participants said they would like university screens to display content specific to their interests on walking past them. Furthermore, it is interesting to see the different responses depending on who is nearby when the screens show adverts. The majority of participants said it would influence what they wanted the screens	The responses to this immersive experience indicate that the participants liked the idea of adaptive displays but that privacy in general is a concern, although it is noted that this changes according to who is in the vicinity at the time. Again, we should probably test this further in the	I11	S-RO2 S-RO6 S-RO7 S-RO8

	to show when they were on their own, with friends, or with strangers. Some people felt they were happier with friends seeing personal information than strangers, while some said the exact opposite. Some remarked they would like the screens to adapt to groups of people to reflect collective interests in some way.	first user trial. The idea of adapting the contents based on group interests could make for interesting future analysis. It is clear that privacy is a concern when using open displays and the content displayed on them.		
ES_01 *Scene 3* **Wizard of Oz :** Yes **Hotspot:** 2	A slight majority is comfortable with minor personal information, such as name, appearing on advertisement screens in public areas. While having control over what personal information appeared on such a screen is a "must" have. It was suggested that user aliases could be displayed instead of their actual names, and it was also said that showing a name is acceptable in a place where nobody would know who you are anyway.	Personal information and privacy in general is a concern. Also, having control over what personal data is displayed is very important. A future test could be developed by using participants personal data displayed in various settings, such as; on their own; with friends; with strangers; with a mix of both – and test user reaction to these settings. The scenario could be revised to reflect these new settings. Also, there needs to be a future test that evaluates the system learning feature and how this might impact participants concerns about privacy, etc. The results from Q9 in the "Wizard of Oz" trial clearly show that users want to have control over what personal data is displayed on open displays.	I12	S-RO2 S-RO7
US_03 **Wizard of Oz :** Yes **Hotspot:** 3	Everybody joined the community for one of their personal interests. Almost everybody said they would have joined even if barely any of their friends were in it, suggesting an openness to join communities that interested them, regardless of who the members were (which is backed by their own statements).	This validates the value propositions of SOCIETIES, we should test these services.	I13	S-RO1 S-RO3 S-RO10
US_02 *Scene 5* **Wizard of Oz :** Yes **Hotspot:** 4	Most people agreed the augmented reality glasses were not very usable in their current form, even if wireless. Complaints mainly revolved around the uncomfortable fit of the glasses, which were very bulky and put pressure on the nose, and some remarked that they would be more acceptable if they resembled normal glasses. All bar one participant thought it would be useful for the glasses to display information on people they encountered while wearing them; some	The project probably wants to explore and test other types of user interface. Examples include: (1) Audio information, through an earpiece; (2) Private displays, smartphone, tablets, contact lenses, etc. It is evident from the responses that the participants liked to discreetly see in real-time information about other people, although it was interesting that	I14	S-RO2 S-RO10

	expressed privacy concerns on this point, stressing that the kind of information appearing should be **limitable** by the person being described. Everyone thought it was useful that the glasses and earphones were more discreet than the screens, and consequently everyone was also happy with the glasses displaying more personal and private information than the screens. The unanimous response to both these questions suggests the participants would prefer discreet, personal displays of information, which wouldn't be viewable by others.	what should be seen needs to be under the control of the observed.		
US_01 *Scene 3* **Wizard of Oz :** Yes **Hotspot: 5**	Almost all students would appreciate exam timetable information showing on university screens. Most would also appreciate variety in the content of such screens: both course-related information and personal advertisements are desired, but the majority of participants have concerns about what information is displayed.	Possibly revise scenario to include mixing personalized advertisements with University info, although this is not a "must" have. One of the questions is "open-ended" (i.e. Q24) and it is suggested that this could be tested further. Having said that it could be regarded as a previous similar concern regarding privacy. It is not quite clear about the result in terms of the "factors" that would effect what was being displayed.	I15	S-RO2 S-RO6 S-RO10
US_04 *Scene 5* **Wizard of Oz :** Yes **Hotspot: 5**	Everybody joined the exam study community. When asked why most explained that essentially they were studying for that exam, so it made sense to join a community that could help them. Nobody had a problem giving their course timetable data; some people had problems giving their mobile phone or their location being given, but it is notable that less people had problems with location than with their phone. Most people are comfortable with their location being broadcast to the exam study community. Combining this with the general response to the that it would not matter if few friends were already members, the implication is that most of the students do not mind letting communities made up of non-friends have access to their location.	Revise – the original scenario did not include the factors outlined in the questionnaires (i.e. it should now). Might be interesting to test further whether participants really do surrender their location data as easily as the inference to the survey suggests. Just to test the inferred results again.	I16	S-RO1 S-RO6 S-RO10

3.3.3 D8.1 Student Group General Discussion Response

Group Discussion Recommendation	Test discussion or Description	Input ID	Objective Reference
The cardboard box artwork community was intended to cater directly to no particular participants' interests, but to be a 'random' community unrelated to them. Most did not join the community, and of those, most would not have been persuaded to join by any kind of incentive. Based on what participants said about their decisions, it would seem the participants join communities because they cater to their interests or because they intrigue them.	**NB:** this was not taken explicitly from one of our D2.2 scenarios This scenario covers the reasons why people join communities. We might want to give visual indicators or scores to communities (e.g. TripAdvisor scores etc). This is an interesting "control group" scenario, which might have benefits for our usability testing in future trials.	I17	S-RO10
These questions followed on from the previous cardboard box artwork community questions and were meant to evaluate whether knowing about other community members would affect your willingness to join the community. The most interesting result was that the participants who decided not to join the community originally, changed their minds when they knew that a lot of their friends were already community members.	**NB:** this was not taken explicitly from one of our D2.2 scenarios This is not one of our original scenarios so we might want to consider making it one, as previously suggested. We can probably devise some interesting tests here for future trials.	I18	S-RO3 S-RO10

In both trials, most students express open-mindedness to many of the kinds of pervasive ideas being conveyed to them. From the perspective of SOCIETIES requirements and concepts, this is a good deal of positive reinforcement to our current position, which should help with future development decisions. There are some points of contention, however, where the results indicate a clash between current SOCIETIES ideas and student opinions. The main one relates to automated decision making: while most students approve of this, they are not keen on this happening without any notification and with no control over the decision. When pressed on this issue in the immersive trial, some students acknowledged they would not mind full automation on very trivial issues, but they would want to give approval of any automated decisions that they considered non-trivial.

Something that comes out more in the immersive trial, where students could speak their mind and express themselves, is that different students have different opinions on various issues, so some are more open to certain ideas that others are against, and vice versa. When explaining themselves, it sometimes becomes clear they are concerned about specific consequences of the ideas, rather than about everything the idea entails, and some of their concerns would be guarded against in SOCIETIES. One student, when asked about personalised adverts being displayed for them, says they "can't see it working" because there could be multiple people around you. In reality, SOCIETIES could deal with this by showing adverts that appeal to collective interests, or by only showing adverts to lone individuals. Another student remarks they would not want adverts to appear in university because when they are heading to lectures they are not interested. Again, the system would be able to deal with this by only showing adverts to students when they don't have upcoming lectures timetabled.

In summary, the results garnered from the student trials support existing SOCIETIES concepts and requirements. Both trials produced similar results on key issues, showing a convergence of student opinion. Such issues include: 1) privacy and information disclosure, where most students are happy to disclose basic information, while for more personal information they should be consulted first; 2) students would likely join social communities that interest them; 3) pervasive services, such as playing preferred music or a cooking service, are perceived as useful for at least some purposes. The points on which the general student opinion differs from pervasive ideals do not contradict the requirements, although they should be given due consideration when we are elaborating the relevant concepts, mainly: 1) automated decision making and sub-community creation/membership, where a desire for user confirmation or notification of such decisions was often expressed; and 2) while it is considered acceptable if the system suggests to greet somebody based on shared interests or intents, as well as the person's mood, not so many students would use this to start friendships, instead preferring to meet new people "the old-fashioned way".

3.4 Temporal aspects

The trials aim to address students over academic term time.

Currently, one confirmed user group consists of all the Computer Science and Information Systems students who are in their 3rd year come October 2012. Other groups may be involved at the same time, but regardless, this first trial period is expected to last at least three months and at least 30-50 users should be involved, each of which will receive a CSS device and use it through daily life. We are hoping that the trial can be continued until late in the academic year which could be as far as June 2013.

A second trial period may be undertaken starting from October 2013 and continuing for as long as project time allows. This would most likely involve the confirmed user group from the first trial period, who would now be in their fourth year, and they may be encouraged to use their CSSs for the dissertation projects they will be undertaking at the time.

Since the trials will be carried out through the student's daily life, it is not possible to define more precise time flows in advance; the only constraints are that they take place during term time so the student group is accessible and generally on campus.

3.5 Physical / location aspects

The student user trials will take place at the Heriot-Watt University (HWU) campus in Edinburgh, Scotland and surrounding areas.

In the HWU campus, the following devices will be deployed to capture information or provide services:

1. Wireless network infrastructure in the HWU campus
2. RFID indoor location system in the pervasive computing lab at HWU
3. Bluetooth indoor location system in the pervasive computing lab at HWU
4. Plasma screens in the HWU campus
5. Directed speakers in the pervasive computing lab at HWU

Meanwhile, a CSS device will be assigned to each student, likely to be Android smart phones.

3.6 Third party services

The trials will incorporate the following third party services to be implemented by WP6. These are mapped to the storyboard scenes for clarity.

Service Name	Priority	Scene	Service Description	Evaluation Points	Paper trial feedback taken into account	Compelling Points and Improvement potential
Navigation Service	High	4 - Directions and Introductions	Directs individual to some location through both outdoor and indoor environments. (E.g., Harry is leaving his dorm room to attend his first lecture. His CSS identifies his intent to attend the lecture so the navigation service is automatically started to direct Harry to the lecture room.)	WP5 User Intent		The service should provide a map of the area, and markers indicating where the destination and user are. The user's marker should update in realtime as they move. No user interaction should be necessary
Passerby identifier service	High	4 - Directions and Introductions	When this service is active, whenever the user comes in close proximity of another CSS holder, the user can check their device screen which will discreetly show details on the CSS holder that are publically available. The service provides options for the user to personalise this, such as only displaying info on people they have listed as a contact, or on people they don't interact with regularly via CSS.	WP4 Usability of setting up platform	Most students would appreciate suggestions to introduce themselves to new people based on shared aspects, although many of them would still prefer making friends traditionally. On the opposite side of things, most students would have no problem with somebody else being the one to receive the suggestion to introduce themselves.	Information on passersby should be displayed discreetly as notifications on a user device. User interaction is limited to the user entering filter settings on what data to show and for who
Calendar Service	High	8 - Studying for exams	Manages a calendar for inidivual user detailing all their meetings, events, timetables, dates to remember etc. This service acts as a context source.	Preference Management		The calender could appear to the user like the SOCIETIES wiki calendar. The user should be able to enter bullet-point entries in each date on the calender, and tag them with metadata that makes them more useful as context sources

Content Sharing Service	High	8 - Studying for exams	Allows individuals to share content with other community members. This might be documents, photos, videos, etc.	WP4 Sharing an "item" with feedback		A filechooser and 'submit' button should be all that's needed to send a file. A text window should be on the same screen to display feedback from file receivers.
Meeting organiser Service	High	8 - Studying for exams	Automatically organises face to face meetings among community members based on some criteria and availablity.	WP5 Automatic communities WP5 Preference Learning	Almost all students felt they would appreciate technology to help them with their studies by allowing them to join communities and have face-to-face meeting dedicated to it. Students would therefore seem to like things that help them with their academic life, in addition to helping them find jobs.	Should allow the user to manually organise meetings, otherwise it should just give notifications to the user on what meetings it has organised, with alerts when the meeting time approaches
Screen display service	High	N/A	Students express privacy concerns with what is displayed on screens, and want to see different things at different times. A service that allows each student to specify what kind of thing they would want to see on university screens may be desirable. They can choose at any time whether they want to see exam timetables, course timetables, university news, student union/society news, etc.	WP5 Preference Management WP4 Scalability of Platform		The user should be shown a simple list of choices for what to be shown on screens, with no other interaction
Insant Messaging Service	High	8. Studying for exams	Enables instant messages to be sent between two or more community members			The user should be able to easily select which CSS to chat with from 1) a list of CSSs known to be acquaintances of the user, 2) a list of local CSSs, 3) a search function for CSSs. For chatting itself, the services should provide a chat history box, a text input box, and information on the other CSS.

3.7 User aspects

3.7.1 Number of users

We expect at least the following numbers of users:

- 3rd year students: 30-50.

These students have already been involved in an initial feedback session, a paper trial and an immersive paper trial. Further interaction and feedback sessions are planned. More user groups may be trialled at the same time as the 3rd year students, but they are unconfirmed and we cannot reasonably estimate how many more users they could add. In any case, the trial should take place during term time so the student group is on campus and accessible.

3.7.2 Person-on-person monitoring / mentoring

Societies staff will give tutorial sessions to each user group before their trial begins to introduce how to use the CSS device and 3rd party service development.

We suggest that a questionnaire on technology acceptance be given to users before the trial. For this we will use standardized models.

3.7.3 User Feedback

Users will be able to provide feedback iteratively during the trials by communicating directly with the Societies support staff. Formally, they will be interviewed after the trials. The interview will address the same line of questions as the paper trials.

3.8 Observation / Monitoring

Monitoring during trials will be mainly based on the life logging application embedded in the CSS devices. Meanwhile, we plan to deploy wireless network infrastructure, RFID indoor location system, Bluetooth indoor location system as well as Plasma screens to monitor users.

Students will be invited to write first-person accounts of their experience during their trial. Personal Monitoring during the person-on-person mentoring will allow the documentation of the users during their service usage with audio / and or written notes.

Additional observation and documentation will include still photographs, screenshots (auto-generated from the applications at specific actions), log files (auto-generated from the applications and the Societies platform at specific actions).

3.9 Formal Performance Metrics and Feedback for WP4, 5 & 6

3.9.1 Student Group Test Tables

In the following test tables we have taken the analysis from the tables in Section 3.3 and created a list of tests, in order of priority. The first column gives the unique test identifier. The second column gives the test prioritisation. The next column gives a summary of the test description. The last four columns are attempting to capture more detailed requirements for each of these tests, as they relate particularly to WP4-6.

Test ID	Priority	Test description	Input	WP4 Relation	WP5 Relation	WP6 Relation	Objective Reference
S1	H	This test should be conducted at the beginning of the user trial. We plan to obtain users' opinion on community services, especially automatic community join and proactive services.	Questionnaire, logging.		Decision making, proactivity, automatic community creation	User interface relating proactivity to the service itself	S-RO5 S-RO7
S2	H	This test aims to get users' opinion on proactive social event, do they mind disclosing their personal information (e.g., music preferences) while taking part in social activities? What kind of data and how much they would be prepared to give us?	Questionnaire, logging.		Privacy, proactivity	User interface relating potential preference edit	S-RO7 S-RO8
S4	H	Test whether users would like services to start automatically on their behalf, such as automatic friend recommendation.	Questionnaire, logging.		Proactivity	User interface relating proactivity to the service itself	S-RO4 S-RO5 S-UO1 S-UO3
S6	H	Test whether students would like to join or even be automatically added into sub-communities based on shared criteria.	Questionnaire, logging.		Context inference, preference learning, community orchestration		S-RO1 S-RO3 S-RO4 S-RO5 S-RO7
S9	H	Test how the students trust the decision making mechanism provided by SOCIETIES, should the system ask users for confirm first before any action?	Questionnaire, logging.		Decision making, preference learning	User interface relating possible popups	S-RO4 S-RO5
S11	H	Test whether students want the university screens to display content specific to their interests on walking past them. If they want, how much personal/privacy data would they disclose?	Observation, questionnaire.		Privacy, location management		S-RO2 S-RO6 S-RO7 S-RO8
S13	H	Test what kind of criteria would impact users' decision on joining in a new community, current members, their own interests, or any other criteria.	Questionnaire, logging.		Context inference, preference learning		S-RO1 S-RO3 S-RO10
S17	H	Test whether the student would like to join a random community without a particular interest.	Questionnaire, logging.		Community orchestration	Service efficiency	S-RO10
S18	H	Test whether knowing about other community members would affect students' willingness to join the community.	Questionnaire, logging.		Community orchestration	Service efficiency	S-RO3 S-RO10

S19	H	Test whether the user is able to start up the SOCIETIES platform with relative ease. Can they perform the tasks of: • Creating their CSS and adding nodes? • What did the user have to perform to get up and running	Observation, questionnaire	CSS Management		Ease of use	S-UO3
S20	H	Test how well interactions with other CIS's are performed automatically, and how easy it is to do manually, with criteria such as: • Was the user able to search for other CIS's to join? • Was the user automatically joined to CIS's based on preferences/interests?	Observation, questionnaire	CIS Management		Ease of use	S-UO1 S-UO3 S-RO3
S21	H	Test and evaluate the user's interaction with the configuration screen, specifically regarding some basic settings that the user will need to define there. Test criteria include: • Was the user able to select the appropriate settings? • Were there configuration items that they felt were missing?	Observation, questionnaire	CSS Management		Ease of use	S-UO1 S-UO3
S22	H	Test the user experience with Android devices: • How usable did the users find the Android as a device for their CSSs? • What limitations of Android were found that prevented users from participating fully in the student trials?	Observation, questionnaire			Ease of use	S-UO1 S-UO3 S-TO1 S-TO2 S-TO3
S23	H	Test the scalability of the system when we have large numbers of users interacting with each other. Where are the perceived bottlenecks? • Is our design proving to be scalable? Or • The prototype implementation was not scalable due to minimal deployment and partial system development.	Observation, questionnaire, data logging	Dependability			S-TO2
S24	H	Test how well the framework caters to the student user group in particular: • Does the platform seamlessly interact with the student user group? • Does the platform provide appropriate unique functions for this group compared to the other user groups?	Observation, questionnaire			Ease of use	

S25	H	Test interoperability of each of the software components with real world environments, criteria includes: • Performance of platform on each OS's of the mobile devices • Several components deployed on different nodes in the cloud. • Where do failures occur?	Observation, questionnaire, data logging	Dependability			S-TO2
S26	H	Test user publishing aspects: • The user is able to publish an "item": (similar to Facebook Wall) • The user can publish a comment/video/file/etc to a known location • Allow users to provide feedback and comments	Observation, questionnaire, data logging			User interface, ease of use	S-UO1
S27	H	Test whether the service container API is satisfactory: • Did it allow the creation/deployment of services with easy deployment? • What limitations did 3rd party developers find when creating shared services for the community?	Observation, questionnaire, data logging	Service Infrastructure			S-UO1 S-UO3
S28	H	Test how easily users of the platform can interact with existing services and how well this works: • Easy to browse/search service instances within a registry of services. • Ability to search within own CSS/targeted CSS/multiple CIS's [This is as opposed to downloadable software from an appstore or marketplace]	Observation, questionnaire, data logging	Service Infrastructure			S-UO1 S-UO3
S29	H	Given that users are getting to interact with software, evaluate their desire for future applications/technology solutions based on our presented system: • What features would users like to see? • Features that were felt to be missing?	Observation, questionnaire				S-TO3 S-RO10 S-UO1 S-UO3

S30	H	Test and evaluate system recommendations, both CIS and non-CIS-based: • Do “communities to individual” recommendations improve over time during the course of the trials? I.e. does the user acceptance rate improve? • Do 'individual to community' recommendations improve over time during the course of the trials? I.e. does the acceptance rate improve? • Are recommendations based on data gathered from several sources better than from each of the sources alone? • What is the difference in acceptance rates for different recommendation schemes?	Data logging		Community orchestration		S-TO2
S31	H	Test and evaluate how users handle and make use of CISs: • Why do users create CISs? E.g. is it for family, interests, an ad-hoc meeting, etc? • Do user's create more long-term or ad-hoc communities? • What is the most relevant criteria for a user to join an existing community? E.g. friends already members, community purpose, etc. • How many CISs does the average user create? • How many CISs is the average user a member of?	Questionnaire, data logging		Community orchestration		S-RO3 S-RO10 S-UO1 S-UO3
S32	H	Test and evaluate automatic community orchestration, and how the users feel about it: • Do users see the benefit in the automatic creation of ad-hoc communities? • Do users see the benefit in the automatic creation of long-term communities? • Do users like to be automatically added to ad-hoc communities? • Do users like to be automatically added to long-term communities? • How much information is needed to create an automatic community?	Questionnaire, data logging		Community orchestration		S-RO5 S-RO10 S-UO3

S33	H	Test how community context data is handled and how well it works: • How accurate is community context data? • What are the delays involved in requesting context data from a community? • What are the delays involved in requesting context data from another individual? • Predefined context data distribution policies will exist, on CIS and CSS level. User will be able to further parameterize these policies. To which extend are predefined polices changed?	Questionnaire, data logging		Context		S-TO2 S-UO3
S34	H	Test proactivity- and context history-based computation concerns: • What is the average size of necessary training data sets for efficient proactivity related functionality? • What is the overall processing and storage resources requirements of the context history on an average day of usage? Is this affecting other components?	Data logging		Context, proactivity		S-TO2
S35	H	Test user acceptance of context inference and context management • What context inference mechanisms are producing the most acceptable, by the user, results? • How easy is it for the user to manage their conceptual location information?	Questionnaire, data logging		Context		S-UO1 S-UO3
S36	H	Evaluate user acceptance of monitoring and peripherals: • Do users accept to wear sensors (obtrusiveness)? • Do users accept to be monitored (privacy)?	Data logging		Privacy		S-UO1 S-RO7
S37	H	Test and evaluate context prediction and its performance: • What is the rate of acceptance of the predicted context by the users? How accurate are the predictions? • Is there an impact in terms of memory and processing load, or serious delay to the user?	Questionnaire, data logging		Context		S-UO3 S-TO2

S38	H	Test and evaluate context inheritance, including its conflict resolution: • What is the rate of acceptance of the context inherited by the users from CISs? • How accurate are the context inheritance conflict resolution mechanisms?	Data logging				S-RO1 S-UO3
S39	H	Test the ability of the system to make evaluations on the user and user actions: • Can meaningful evaluations be achieved in a (near) real time solution? • Can evaluations model the dynamic CSS/CIS interactions to provide useful data sets to use as a base for further predictive analytics?	Data logging				
S40	H	Test how well preferences work and user acceptance of them: • Do preferences become more accurate over the course of the user trials? • Do users abort the preferences fewer times? • On average, how many preferences will users acquire over the course of the trials?	Questionnaire, data logging		Personalisation, preferences		S-RO8
S41	H	Test the learning algorithms involved in the system: • Are incremental learning algorithms able to run continuously for the duration of the trial? • How often are batch preference learning executions scheduled? • What learning algorithm performs best for preference learning?	Data logging		Personalisation, preference learning		S-RO8
S42	H	Evaluate how the personalisation GUI is used: • Do users use the Personalisation GUI to create, edit and delete preferences? • How often do the users use the Personalisation GUI to create preferences? • How often do the users use the Personalisation GUI to edit learnt preferences?	Observation, questionnaire, data logging		Personalisation	User interface	S-RO8

S43	H	Test and evaluate the user intent model and user acceptance of it: • Does the user intent model become more accurate over the course of the trials? Do users abort the intentions fewer times? • What is the confidence level at which users will allow an intent prediction to be performed automatically? • What is the average size of the intent model for all trial participants at the end of the trial? • How does the user's intent model change (grow?) during the course of the trial?	Observation, questionnaire, data logging		Personalisation - user intent		S-RO8
S44	H	Test the data requirements of personalisation: • How much data is necessary to achieve noticeable personalisation? • How do the proposed criteria perform?	Data logging		Personalisation		S-RO8
S45	H	• Does the user trust the privacy management of the system? If yes, what is the main feature providing this trust?	Observation, questionnaire		Privacy		S-RO7
S46	H	Test to what degree the user accepts automated privacy policy decisions: • How often does the user make any changes to the privacy policy suggested by the service? • How many times does the privacy policy negotiation fail because of the user not accepting the terms and conditions of the service?	Questionnaire, data logging		Privacy		S-RO7
S47	H	• Does the user use the privacy GUI to create new privacy preferences? • Does the user use the privacy GUI to change existing preferences? • Do users understand the concept of privacy preferences? If yes how much it is appreciated?	Observation, questionnaire, data logging		Privacy		S-RO7
S48	H	Test and evaluate how users handle data obfuscation and how they feel about it: • Do users understand the concept of data obfuscation? If yes how much it is appreciated? • How many times does the user modify the obfuscation level in their privacy preferences? • Do most users use data obfuscation? • What is the average level of obfuscation per data type chosen by users?	Observation, questionnaire, data logging		Privacy		S-RO7

S49	H	Evaluate user desire for warnings on the sharing of sensitive data: • Would end user benefit from a tool that could warn him that he is about to share media containing sensitive information to inappropriate users or user groups?	Observation, questionnaire				S-RO4 S-RO7
S50	H	Evaluate whether the user would appreciate certain visual aspects for parts of the system: • Would it be helpful for end user to have services fit with a clear visual estimation (e.g. a semaphore and/or progress bar) of trust for each service so he could filter the services by a trust threshold and sort services by trust level? • Would it be helpful for end user to have other users fit with a clear visual estimation (e.g. a semaphore and/or progress bar) of trust augmented for each user within SAG view?	Observation, questionnaire			User interface	S-RO4
S51	H	Test and evaluate personalisation conflict resolution, user agent decision making, and user acceptance of automated actions based on these things: • How often do the personalisation sources suggest conflicting actions? • Which personalisation source suggests the correct action most often? • Is the User Agent able to resolve the conflict among different internal proactive components? • Do users accept automatic decision making? To what extent? • Do users accept the automatic installing, execution, provision of services? To what extent?	Observation, questionnaire, data logging		Personalisation, user agent, decision making		S-RO8
S52	H	Test and evaluate user feedback notifications and how the user feels about them: • Do user feedback notification alerts become less frequent as personalisation models become more accurate? • As notifications alerts become less frequent, is the user relieved not to receive so many or upset at not them? • Does the user find notification alerts annoying and as a result ignore them?	Observation, questionnaire, data logging		Personalisation, user feedback		S-UO3
S53	H	Test community governance	Data logging		Intelligent Community Orchestration		S-TO2

S54	H	Test and evaluate how well the system does in inviting specific individuals to join a community, and the user acceptance of this	Questionnaire, data logging		Intelligent Community Orchestration		S-RO3 S-UO1 S-UO3
S55	H	Test how well the system handles detecting and removing obsolete community members	Questionnaire, data logging		Intelligent Community Orchestration		S-RO3
S56	H	Test the resource sharing functionality of the system	Data logging				S-RO10 S-RO6
S57	H	Test conflict resolution or prioritisation in resource sharing	Data logging		User Agent		S-RO6
S58	H	Test the worth of the techniques for discovery of relationships and behaviours within communities	Interview, questionnaire, data logging		Intelligent Community Orchestration		S-RO3
S59	H	Test the orchestration of multiple communities	Interview, questionnaire, data logging		Intelligent Community Orchestration		S-RO1 S-RO3
S60	H	Test intelligent conflict resolution among community members	Data logging		User Agent		S-TO2
S61	H	Evaluate whether users are happy with the way they receive advertisements for CISs?			Intelligent Community Orchestration		S-UO1 S-UO3
S62	H	Test whether the system is able to identify potential new CIS members based on information derived from social network sites		CSS and SNS integration			S-RO1
S63	H	Test if the system deals well with various pervasive systems coming together to co-operate in an interoperable fashion		Dependability			S-TO1
S64	H	Test and evaluate whether the users are able to communicate ad-hoc with one-another, both intra- and inter-CIS, or without CIS involvement		Communications and Device Asbstractions			S-RO2 S-RO6
S65	H	Test certain technical factors of the system, namely that: • The system is robust, able to deal with all situations capably and without crashing • The system is open • The system is scalable		Dependability			S-TO2
S3	M	Test whether the community “cooking” service is of benefit to users.	Observation, questionnaire.		Community preference learning	Service efficiency	S-RO8 S-UO3
S5	M	Test whether the student would like join a SOCIETIES community to find job or make money.	Observation, questionnaire, logging.			Service efficiency	S-RO6
S7	M	Test whether the device should be able to predict user behaviour and make appropriate suggestions based on this.	Logging.		Sensor, user behaviour learning	Service efficiency, ease of use	S-RO1 S-RO4 S-RO7 S-RO8 S-RO9

S8	M	Test whether students would appreciate technology to help them with their studies by allowing them to join communities dedicated to it	Questionnaire, logging.			Service efficiency	S-RO6
S10	M	Test whether students would like to be aware of actions or contexts of other community members, and whether such information would affect their own decisions.	Questionnaire, logging.		Context management, user action monitor	User Interface for user actions and contexts presentation	S-RO6 S-RO8 S-RO9 S-RO10
S15	M	Test whether students would like the public screens display some information like exam timetable, course-related information and personal advertisements.	Observation, questionnaire.			Service efficiency	S-RO2 S-RO6 S-RO10
S16	M	This test aims to get users' opinion on community participation, do they mind disclosing their personal information? What kind of data and how much they would be prepared to disclose within the community?	Questionnaire, logging.		Privacy		S-RO1 S-RO6 S-RO10
S12	L	Test whether students want a public screen to display their personal information. If they want, do they have any specific consideration?	Questionnaire, logging.		Privacy	Service efficiency	S-RO2 S-RO7
S14	L	Evaluate students' opinion on augmented reality glasses.	Observation, questionnaire.			Service efficiency, ease of use	S-RO2 S-RO10

3.9.2 Analysis of the evaluation points set by WP4 with feedback to WP4, 5 & 6

The following table presents a set of evaluation points defined by WP4. For each point, we indicate how they can be related to our storyboard and to our questions. Some evaluation points require extensions of the storyboard; these extensions will be considered in the realisation of the trial prototype. Finally we also provide important feedback to WP4, 5 & 6. For instance, in some cases we describe functionality needed for performing the evaluation or we suggest an alternative approach to trial for evaluation.

Evaluation Group	Evaluation points	Storyboard relation to evaluation points / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
SOCIETIES platform startup	Is the user able to start up the SOCIETIES platform with relative ease? Can they perform the tasks of: • Creating their CSS and adding nodes? • What did the user have to perform to get up and running	CSS creation/installation not covered by storyboards. Adding nodes is not covered by storyboards. Also not covered by storyboards. Priority: TBD	Tutorial will be given to students at the beginning of the trial	WP 4 should provide clear platform startup introduction, all the WPs should provide user-friendly GUI.
CIS interaction	Interactions with other CIS's: • Was the user able to search for other CIS's to	Scene 1 – Arrival on Campus		WP 4 should support this function.

	join? • Was the user automatically joined to CIS's based on preferences/interests?	Scene 7 – Hobbies and Interests Priority: TBD		
Usability of setting up platform	There will be some basic settings that the user will need to define in a configuration screen: • Was the user able to select the appropriate settings? • Was there configuration items that they felt were missing?	Not covered by storyboards. Not covered by storyboards. Priority: TBD		WP4 should clarify what is meant by basic settings. It has to be clarified whether all the WPs would provide a GUI for configuration.
User Experience on the mobile device	Based on targeted feature rich device being Android based: • What is the experience for non-Android based devices? • What limitations were found that prevented user from participating fully in each of the user trials?	Systematic, not a scenario specific evaluation Priority: TBD	Each of the user would receive a mobile device, they should also be encouraged to use different kind of mobile devices	
Scalability of Platform	How scalable is the system when we have large numbers of users interacting with each other. Where are the perceived bottlenecks? • Is our design proving to be scalable? Or • The prototype implementation was not scalable due to minimal deployment and partial system development.	Scene 2 – Evening Activities Priority: TBD		Certain system scalability should be performed before the user trial.
Portability of platform across all scenarios	We are developing a framework to satisfy 3 separate user groups: • Does the platform seamlessly interact with each user group? • Is the platform segmented with separate functions for each divide?	Systematic, not a scenario specific evaluation Priority: TBD		How much work needs to be done by WP4 and WP5 in order to support different user groups.
Interoperability across multiple platforms	Investigate interoperability of each of the software components with real world environments. • Performance of platform	Systematic, not a scenario specific evaluation		Performance should be addressed through simulations first.

	on each OS's of the mobile devices • Several components deployed on different nodes in the cloud. • Where do failures occur?	Priority: TBD		We suggest WP4 and WP5 to provide log functionality so that failures can be traced.
Sharing an "item" with feedback	Ability to publish an "item": (similar to Facebook Wall) • Publish a comment/video/file/etc to a known location • Allow users to provide feedback and comments	Scene 6 – Dining out Scene 7 – Hobbies and Interests Priority: TBD		WP4 should not only support sharing but also annotating
Service Lifecycle	Was the service container API satisfactory? • Did it allow the creation/deployment of services with easy deployment? • What limitations did 3rd party developers find when creating shared services for the community?	Systematic, not a scenario specific evaluation Priority: TBD	The trial may also ask users to develop some 3rd party services	The 3rd party services developed by WP6 should be in line with the service lifecycle management mechanism provided by WP4.
Service Registry	Users of the platform interacting with existing services: • easy to browse/search service instances within a registry of services. • Ability to search within own CSS/targeted CSS/multiple CIS's [This is as opposed to downloadable software from an appstore or marketplace]	Systematic, not a scenario specific evaluation Priority: TBD		The 3rd party services developed by WP6 should be in line with the service registry mechanism provided by WP4.
Future Feature Requests	With users getting to interact with software, understanding their desire for future applications/technology solutions based on our presented system: • what features would users like to see? • Features that were felt to be missing?	Systematic, not a scenario specific evaluation Priority: TBD	Users should be encouraged to give feedback or requests during the user trial.	

CSS/CIS Management - Advertisement	• Are users happy with the way they receive advertisements for CISs?	Scene 5 - Part time employment Scene 6 - Dining out Priority: TBD		Should collect users' feedback for this
CSS and SNS Integration	• Is the system able to identify potential new CIS members based on information derived from social network sites?	Not covered in the storyboard Priority: TBD	Extend the storyboard to cover this issue	
Multiple CSS Interoperability Management	• Does the system deal well with various pervasive systems coming together to co-operate in an interoperable fashion?	Systematic, not explicitly covered Priority: TBD	Extend the storyboard to cover this issue	
Communications and Device Abstractions – ad-hoc communications	• Are the users able to communicate ad-hoc with one-another, both intra- and inter-CIS, or without CIS involvement?	Systematic Priority: TBD		How to evaluate this? Based on user feedback?
System-wide	Technical factors in the system: • The system is robust, able to deal with all situations capably and without crashing • The system is open • The system is scalable	Systematic Priority: TBD		We should propose appropriate metrics to evaluate the robustness and the scalability of the platform

3.9.3 Analysis of the evaluation points set by WP5 with feedback to WP4, 5 & 6

The following table presents a set of evaluation points defined by WP5. For each point, we indicate how they can be related to our storyboard and to our questions. Some evaluation points require extensions of the storyboard; these extensions will be considered in the realisation of the trial prototype. Finally we also provide important feedback to WP4, 5 & 6. For instance, in some cases we describe functionality needed for performing the evaluation or we suggest an alternative approach to trial for evaluation.

Evaluation Group	Evaluation points	Storyboard relation to evaluation points / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
Community/Individual Recommendations	• Do "communities to individual" recommendations improve over time during the course of the trials? I.e. does the user acceptance rate improve? • Do 'individual to community' recommendations improve over time during the course of the trials? I.e. does the acceptance rate improve?	Scene 7 – Hobbies and Interests Scene 7 – Hobbies and Interests		WP5 should provide user-friendly GUI for this

	• Are recommendations based on data gathered from several sources better than from each of the sources alone? • What is the difference in acceptance rates for different recommendation schemes?	It is not easy to determine what kind of data sources will be used according to scenario description It is not easy to determine what kind of recommendation schemes will be used according to scene description Priority: TBD		
Manual Communities	• Why do users create CISs? E.g. is it for family, interests, an ad-hoc meeting, etc? • Do user's create more long-term or ad-hoc communities? • What is the most relevant criteria for a user to join an existing community? E.g. friends already members, community purpose, etc. • How many CISs does the average user create? • How many CISs is the average user a member of?	Scene 1 – Arrival on Campus Scene 7 – Hobbies and Interests Scene 8 – Studying for exam Scene 7 – Hobbies and Interests Scene 8 – Studying for exam Scene 5 – Part time employment Scene 6 – Dining out Scene 7 – Hobbies and Interests Scene 8 – Studying for exam Systematic, all the scenes Systematic, all the scenes Priority: TBD		WP5 should give introduction for this as well as provide ease to use GUI
Automatic communities	• Do users see the benefit in the automatic creation of ad-hoc communities? • Do users see the benefit in the automatic creation of long-term communities? • Do users like to be automatically added to ad-hoc communities? • Do users like to be automatically added to long-term communities? • How much information is needed to create an automatic community?	Scene 7 – Hobbies and Interests Scene 1 – Arrival on Campus Scene 6 – Dining out Scene 7 – Hobbies and Interests Scene 1 – Arrival on Campus Scene 7 – Hobbies and Interests Systematic Priority: TBD		
Community based context data distribution	• How accurate is community context data? • What are the delays involved in requesting context data from a community?	Systematic Systematic, Scene 5 – Part time employment		WP5 should explain how to evaluate the accuracy of community context, so that WP6 is able to develop 3rd party

	• What are the delays involved in requesting context data from another individual? • Predefined context data distribution policies will exist, on CIS and CSS level. User will be able to further parameterize these policies. To which extend are predefined polices changed?	Systematic, Scene 2 – Evening Activities Systematic Priority: TBD		services to collect the necessary information
Intelligent context history management	• What is the average size of necessary training data sets for efficient proactivity related functionality? • What is the overall processing and storage resources requirements of the context history on an average day of usage? Is this affecting other components?	Scene 4 – Directions and Introduction Systematic, all the scenes Priority: TBD		
Context Inference	• What context inference mechanisms are producing the most acceptable, by the user, results? • How easy is it for the user to manage their conceptual location information?	Systematic, it is not easy to understand what kind of context inference mechanism will be used according to scene description. Not covered by storyboards Priority: TBD		How can we estimate users' acceptability of the inference results?
Sensors and Monitoring	• Do users accept to wear sensors (obtrusiveness)? • Do users accept to be monitored (privacy)?	Scene 2 – Evening Activities Scene 2 – Evening Activities Priority: TBD	What kind of sensors will be used during the trial? What kind of information will these sensors collect? How would these information be used?	
Context Prediction	• What is the rate of acceptance of the predicted context by the users? How accurate are the predictions? • Is there an impact in terms of memory and processing load, or serious delay to the user?	Scene 4 – Directions and Introduction Systematic Priority: TBD		WP5 should develop corresponding monitoring components to collect this kind of information

Context Inheritance	• What is the rate of acceptance of the context inherited by the users from CISs? • How accurate are the context inheritance conflict resolution mechanisms?	Scene 6 – Dining out Scene 2 – Evening Activities Priority: TBD		When users inherit context from CISs, the system should notify them.
Context Similarity	• Can meaningful evaluations be achieved in a (near) real time solution? • Can evaluations model the dynamic CSS/CIS interactions to provide useful data sets to use as a base for further predictive analytics?	Scene 6 – Dining out Scene 2 – Evening Activities Not covered Priority: TBD		WP5 should explain the second point in more detail.
Preference Management	• Do preferences become more accurate over the course of the user trials? • Do users abort the preferences fewer times? • On average, how many preferences will users acquire over the course of the trials?	Scene 2 – Evening Activities Scene 3 – Meal Preparation Scene 7 – Hobbies and Interests Systematic Systematic Priority: TBD	Users should be notified that how would their preference be managed	WP5 should leverage user feedback to evaluate these points
Preference Learning	• Are incremental learning algorithms able to run continuously for the duration of the trial? • How often are batch preference learning executions scheduled? • What learning algorithm performs best for preference learning?	Scene 2 – Evening Activities Scene 3 – Meal Preparation Scene 7 – Hobbies and Interests Systematic Systematic Priority: TBD	Same as the above	
Personalisation GUI	• Do users use the Personalisation GUI to create, edit and delete preferences? • How often do the users use the Personalisation GUI to create preferences? • How often do the users use the Personalisation GUI to edit learnt preferences?	Systematic Priority: TBD		The GUI should be easy to use

User Intent	• Does the user intent model become more accurate over the course of the trials? Do users abort the intentions fewer times? • What is the confidence level at which users will allow an intent prediction to be performed automatically? • What is the average size of the intent model for all trial participants at the end of the trial? • How does the user's intent model change (grow?) during the course of the trial?	Scene 4 – Directions and Introduction Systematic Systematic Systematic Priority: TBD	Users should be notified that the platform might discover as well as predict their intent	
Personalization	• How much data is necessary to achieve noticeable personalisation? • How do the proposed criteria perform?	Systematic Systematic Priority: TBD		
Privacy Management	• Does the user trust the privacy management of the system? If yes, what is the main feature providing this trust?	Systematic, not covered Priority: TBD	Extend the storyboard to cover this issue	
Privacy Policy Negotiation	• How often does the user make any changes to the privacy policy suggested by the service? • How many times does the privacy policy negotiation fail because of the user not accepting the terms and conditions of the service?	Systematic, not covered Systematic, not covered Priority: TBD	Extend the storyboard to cover this issue	
Privacy Preference Manager	• Does the user use the privacy GUI to create new privacy preferences? • Does the user use the privacy GUI to change existing preferences? • Do users understand the concept of privacy preferences? If yes how much it is appreciated?	Systematic, not covered Systematic, not covered Systematic, not covered Priority: TBD	Extend the storyboard to cover this issue	
Data Obfuscation	• Do users understand the concept of data obfuscation? If yes how	Not covered by storyboards Should be Systematic	Extend the storyboard to cover this issue	

	much it is appreciated? • How many times does the user modify the obfuscation level in their privacy preferences? • Do most users use data obfuscation? • What is the average level of obfuscation per data type chosen by users?	 Systematic Systematic Systematic Priority: TBD		
Active Privacy Assistant	• Would end user benefit from a tool that could warn him that he is about to share media containing sensitive information to inappropriate users or user groups?	Not covered by storyboards Priority: TBD	Extend the storyboard to cover this issue	
Visual Trust Indicators	• Would it be helpful for end user to have services fit with a clear visual estimation (e.g. a semaphore and/or progress bar) of trust for each service so he could filter the services by a trust threshold and sort services by trust level? • Would it be helpful for end user to have other users fit with a clear visual estimation (e.g. a semaphore and/or progress bar) of trust augmented for each user within SAG view?	The scenes do not cover this explicitly. The scenes do not cover this explicitly. Priority: TBD	Extend the storyboard to cover this issue	
User Agent	• How often do the personalisation sources suggest conflicting actions? • Which personalisation source suggests the correct action most often? • Is the User Agent able to resolve the conflict among different internal proactive components? • Do users accept automatic decision making? To what extent? • Do users accept the automatic installing, execution, provision of	Systematic Systematic Systematic Scene 7 – Hobbies and Interests Scene 4 – Directions and Introduction		User Agent should provide user-friendly and easy to use GUI

	services? To what extent?	Priority: TBD		
User Feedback	• Do user feedback notification alerts become less frequent as personalisation models become more accurate? • As notifications alerts become less frequent, is the user relieved not to receive so many or upset at not them? • Does the user find notification alerts annoying and as a result ignore them?	Systematic Systematic Systematic Priority: TBD	The storyboard should explain to the users how would their feedback be used.	
User Feedback	• Does UAV steering work with an acceptable delay for remote controllers?	N/A to Student Group trial		
User Agent	• Are users satisfied with the decisions made following intelligent conflict resolution among community members?	Scene 2 – Evening Activities Scene 3 - Meal Preparation Priority: TBD		Once a decision has been made, User Agent should collect the user's feedback by multiple channels.
User Agent	• Are users satisfied with decisions made based on cconflict resolution or prioritization in resource sharing?	The scenes do not cover this explicitly. Priority: TBD	Extend the storyboard to cover this issue	
Intelligent Community Orchestration	• Are community governance actions carried out efficiently on appropriate CSSs, to the satisfaction of CIS members?	Systematic Priority: TBD		The system should collect necessary information and design appropriate metrics to evaluate the satisfaction of CIS members
Intelligent Community Orchestration	• To what extent does the user receive recommendations to join CISs that are personally relevant? • Are users satisfied with the CIS recommendations they receive?	Scene 5 - Part time employment Priority: TBD		User Agent should collect relevant user feedbacks

Intelligent Community Orchestration	• Have users ever been removed from CISs without wanting to leave, and to what extent?	Systematic, not covered Priority: TBD	Extend the storyboard to cover this issue	
Intelligent Community Orchestration	• Are users satisfied with the resource sharing functionality on offer?	Systematic Priority: TBD		Collect relevant user feedbacks through User Agent
Intelligent Community Orchestration	• Do the techniques for discovery of relationships and behaviors within communities produce suggestions, decisions, and actions that are meaningful to the users?	Scene 2 – Evening Activities Scene 5 - Part time employment Scene 3 - Meal Preparation Priority: TBD		Collect relevant user feedbacks through User Agent
Intelligent Community Orchestration	• Does the system orchestrate multiple communities to the satisfaction of members and other relevant CSSs?	Systematic, not explicitly covered Priority: TBD	Extend the storyboard to cover this issue	

4 Disaster Group Trials

4.1 Summary of the trials

This trial focuses on assessment missions as a response of our society to major natural and man-made disasters in any area of the world. The users are assessments experts who arrive at such a disaster scene in order to assess damage, available resources, and so help coordinate the disaster relief actions. Typically, such a mission lasts one to three days and is the first phase of the response to the disaster. The Societies use case targets two ends of this scenario: the on-site segment (i.e. at the physical scene of the disaster) and a novel component, that of volunteers from outside the disaster area. The significant novelty is that these volunteers can participate in crowd-sourced tasks (e.g. translation, decoding, and operation of autonomous aerial vehicles for damage assessment).

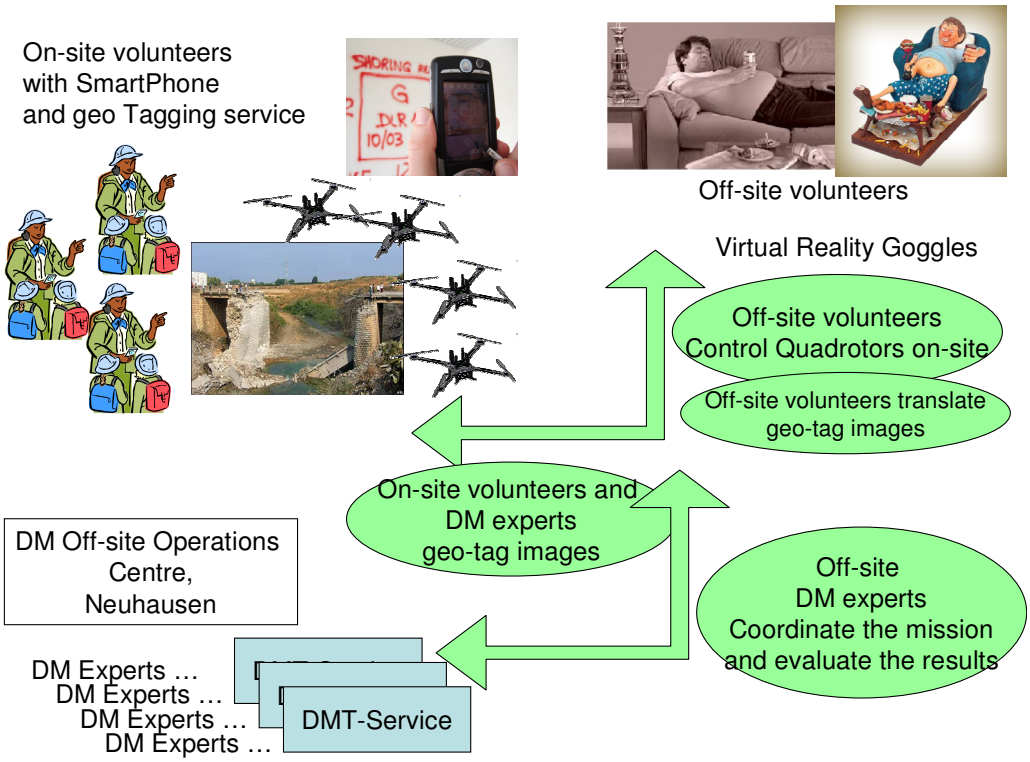

Figure 4: Overview of the DM trials

We anticipate that the second trial will offer significant advances over the first trial, which is the focus of this deliverable. Overall, however, the major research questions addressed on both trials will be directed towards illuminating the value propositions of the project. These value propositions are reflected somewhat differently across all three user groups, despite the many similarities, of course. For the disaster management group three critical issues have been identified in previous work in the project, based on user feedback:

1. Trust issues, in particular with respect to the involvement of the volunteer community.
2. Operation in environments where networking infrastructure is heavily impacted by the disaster.
3. Reliability of the services provided and efficiency gains provided to the stakeholders as a result of Societies' technologies and services.

We will focus on these in the next section.

The first trial will illuminate more the end-user related research questions, whereas we expect the second trial to validate to a greater extent the technical advancements of the platform as well as additional research questions and opportunities that arise in the time leading up to the second trial. We also expect to extend the

scope to new and additional end-users, to enhance and extend the service spectrum, and to allow a greater degree of unsupervised usage.

We are developing the use case whereby many users are involved in the "back end" to perform visual tasks that machines are bad at (e.g. visual assessment). It is a value proposition and research question of the project that we expect that trust can be modelled here by using redundancy, and seeing how many agreements people are achieving. The idea is that each task is mapped 1:n where n is the number of users for that task.

4.2 Objectives of the Disaster Management trials

This following table describes the research objectives and questions for the Disaster Management trial in relation to the project objectives. The feedback of the paper trial was taken into account during the definition of the research questions; the relation between the feedback from the paper trial and the research questions are presented in Section 4.3. Further priority is set on the research questions also based on the feedback and complexity of the prototype needed to answer the questions. A number of questions will be answered during the second trial (marked V2 in the table). To summarize, the following will be given priority in the first trial:

1. Basis support for manual/automatic community management.
2. Basis support for setting up a pervasive communication infrastructure.
3. Involvement of the volunteer community.
4. Flexible and configurable data sharing in a community.
5. Privacy and trust issues associated to the concerns above

The second trial will further investigate these questions (as the platform will be enhanced according to feedback form the first trial) and in addition will address:

1. Interactions among CIS.
2. Community management in a poor communication infrastructure.
3. Remote access to physical resources.
4. User feedback for the definition of new services (seeds from end-users).

Notation: The research objectives and research questions are numbered and prefixed "RO" and "RQ" respectively. This numbering will be used later in this section.

Project objective	**Research objective**	**Research question**	**Relation to the storyboard (towards definition of tests)**	**Prio-rity**
To facilitate the creation, organization, management and communication of communities via CSS/CIS where pervasive is integrated with social	RO_1: Investigate manual and automatic management of CISs and the transition between automatic and manual	RQ1_a: Is it useful to automatically/manually manage communities of rescuers based on expertise / availability / workload / previous collaborations / human-social preferences (additionally location / task / institution, country / language / sex / etc.)?	- Pre_scene: compare automatic management and recommendation for team members and formations. - Scene 5: coordination of teams - Scene 6: compare automatic task allocation and recommendation	H
		RQ1_b: How much automation is accepted by end users and how end users trust this automation?	As RQ1_a	H

Project objective	Research objective	Research question	Relation to the storyboard (towards definition of tests)	Priority
		RQ1_c: Is it useful to automatically/manually manage communities of rescuers and volunteers based on expertise / availability / previous collaborations (additionally location/ / task / institution / country / language /sex /etc.)?	- Scene 2: transcription audio-text - Scene 3: involvement of a quadrator pilot - Scene 4: Translation	H
		RQ1_d: What are the implications of automatic management of CISs for rescuers and volunteers regarding privacy/trust/user control?	As RQ1_c	H
	RO_2: Investigate the intersection between pervasive computing and social computing	RQ2_a: Is it useful to share access to physical resources (e.g. sensor controls, cameras, robotic arms, UAV controls, etc.) in online communities?	- Scene 3: involvement of a quadrator pilot	V2
		RQ2_b: Is it useful to share information collected from physical sensors in online communities?	- Scene 2: they are equipped with … wearable sensors that report and log their activities and interactions.	H
		RQ2_c: Is it useful to provide access to external services (e.g. translation services) in online communities?	- Scene 2: transcription audio-text - Scene 4: Translation	H
		RQ2_d: What are the implications of this type of sharing on privacy/trust/user control/security?	Scenes as above	H
		RQ2_e: What are the preferred ways of setting up a pervasive communication infrastructure for sharing physical resources?	- Scene 2: On site, a communication network is setup as the first step toward deploying a sensor network for collecting measurements and relaying data	H
		RQ2_f: Is it useful to provide support for the configuration of information flows based on type of information/task/ responsibility /etc.	- Scene 2: transcription audio-text - Scene 4: Translation	H
	RO_3: Investigate interactions among CISs (inter-CIS interactions)	RQ_3a: Is it useful to allocate tasks to a specific CIS based on CIS context (e.g. location of members, expertise, workload)?	- Scene 6: task allocation to a team	V2
		RQ_3b: What are the implications of sharing information between teams on privacy/trust/user control?	- Scene 5: information transfer between teams at end of shift.	V2
To provide an enhanced	Advanced context	TBD	Scene2: the handheld devices incorporate advanced sensor	V2

Project objective	Research objective	Research question	Relation to the storyboard (towards definition of tests)	Prio-rity
user experience for *individuals* and *user communities* based on proactive smart behavior and dynamic sharing of community resources across geographic boundaries	reasoning as an objective?		fusion techniques….	
Robust, open and scalable systems for pervasive communities	RO_4: Identify challenges of managing pervasive communities in areas with poor infrastructure	RQ_4a: How did sensor control sharing function in unstable network conditions?	- Scene 2: they are equipped with … wearable sensors that report and log their activities and interactions.	V2
		RQ_4b: How did the transition between connected and disconnected operations worked from a user perspective?	As above	V2
		RQ_4c: Did users have problems with sharing of large amounts of data?	As above	M
		RQ_4d: Did users have problems with synchronizing data in large CISs?	As above	M
To evaluate through strong involvement of end users the usefulness and acceptance of the developed CSS/CIS software prototypes	RO_5: Specific to DMS, investigate contributions that volunteer communities can make in a disaster relief scenario	RQ_5a: What services can the volunteer communities provide?		V2

4.3 Analysis DM Paper Trial Evaluation

In the following table we show how the feedback from the paper trial has influenced the specification of the research questions.

Source in D8.1	Summary of feedback	Discussion and recommendation	Research Question
Section 5.4.2.1 Section 5.4.2.9 Section 5.4.3.2.3 Annex B.1.3, scene 0 (background) Annex B.1.3.5 Annex B.1.3.6 Annex B.1.3.7 Annex B.1.3.13 Annex B.1.3.17	**Community orchestration, Team prioritistaion, Automation** Participants were familiar with concepts like crowd sourcing, and technologies like location sensing devices, but they were not so clear about how an ICT system could make smart or intelligent suggestions or decisions. Some participants did not believe a computer system would be capable of managing these team prioritisation and selection tasks as well as humans, due to the complex influences involved. These influences include: human factors like psychology, chemistry and trust, which are difficult to measure, and this was mentioned in two discussions. However the participants could see potential value in a system making intelligent suggestions.	The participants had no prior knowledge of the system. Lack of familiarity with the system may have influence the answers. We need to let participants try and experience system-based decisions to evaluate how useful the system suggestions might or might not be. Note that, in a first time, we can let the system make recommendations, but not take actual decisions. Decisions should be made by a human (i.e. the OSOCC co-ordinator).	RQ1_a RQ1_b RQ3_a
Section 5.4.2.1 Section 5.4.2.3 Section 5.4.3.2.3 Annex B.1.3.1 Annex B.1.3.3 Annex B.1.3.8 Annex B.1.3.17 Annex B.1.3.18	**Crowd sourcing, Collective Intelligence, Trust** The participants could immediately recognise the value in having access to the "wisdom of the crowd" for assisting with some tasks like audio transcripts, technical expertise (flying UAV remotely), translation or checking satellite images. However, most participants have reservations about information from untrained or unknown sources. A clear support was given for prioritising crowd sourcing information from trained and registered volunteers, particularly those who are traceable through association with existing trusted organisations (such as the Red Cross) was articulated. The participants varied in their opinion of the trustworthiness of crowd sourcing as a data source, and suggested that some conditions would have to be met before they could accept information from offsite volunteers.	We need to build into the system some mechanisms that select the "right online person for the right onsite job" and this selection process must then be trusted by our user communities.	RQ1_c RQ1_d
Section 5.4.2.4 Section 5.4.3.2.4	**Access to better information** Disaster Management participants can see a clear use for more information, and are very positive about the storyboard accessing expert insights, and sharing satellite, maps, and overview information, in general.	We need to identify what information facilitates the work of the rescue team, and what resources an services are needed to provide this information	RQ1_b RQ2_b

Source in D8.1	Summary of feedback	Discussion and recommendation	Research Question
Section 5.4.2.5 Annex B.1.3.4 Annex B.1.3.9 Annex B.1.3.18	**Translation Services** Almost all participants agree that translation constraints hinder relief efforts by slowing down emergency responses in a disaster situation. There was a broad acceptance for the potential use of a SOCIETIES translation service. However, participants also voiced a clear preference for a human translator and guide, who could also read and interpret local cultural and geographical signs in three of the four participant discussions. One participant also expressed a wariness of unknown translators, or offsite volunteers.	We need to identify what services facilitate the work of the rescue team, and what people are best to provide these services.	RQ2_c RQ1_c
Section 5.4.2.1 Section 5.4.2.6 Annex B.1.3.10	**Remote Distribution of Tasks** Audio transcription by off site volunteers was a service proposed in the storyboard, which raised different opinions. While some users saw great value in relieving first responders of administrative tasks, another user was very wary of the value as first hand experience would still be required to edit reports, and also that person had concerns about security of sensitive data. A range of other was also illustrated, e.g. assessment of satellite imagery by untrained volunteers and remote control an UAV by an expert. There was a strong potential value recognised in allocating some specialist tasks to experts off site on a case- by-case basis.	We need to identify what services facilitate the work of the rescue team, and what people are best to provide these services.	RQ2_c RQ1_c
Section 5.4.2.7 Section 5.4.3.2.1 Annex B.1.3.11 Annex B.1.3.15	**General pervasive technologies, and sharing of data - Privacy** Monitoring individuals was not viewed favourably in general, with some notable exceptions. The fear that the proposed system could introduce close monitoring akin to a panopticon 'big brother' was mentioned in two different discussions about the use of sensors. Participants indicated they would be happy to have some internal information sharing within the team, if it benefited team health and safety, but that they would like to maintain complete confidentiality with the team.	This feedback indicates that the SOCIETIES system should help members to achieve their goals, but not interfere with privacy of individuals. It shows protectiveness about the level of trust, cooperation, and inter-reliance required within teams. This is necessary for their work. We need to investigate support for configurable privacy rules.	RQ2_d RQ3_b

Source in D8.1	Summary of feedback	Discussion and recommendation	Research Question
Section 5.4.2.8 Section 5.4.3.2.1 Annex B.1.3.12	**Information flows - Status sharing, location tagging - Privacy** Most users are familiar with sharing GPS coordinates for location, but appear to make a distinction between retaining control in sharing this information by choosing to send it out, at particular points, and sharing it continuously, or making it available on demand. They made clear distinctions between public web sharing and private team or OSOCC sharing. The participants had not enough information about the flow of data collected by the sensors in the storyboards to make decisions, about whether they would wear sensors. However several are positive to wear sensors. There was much disagreement among participants as for sharing personal data collected by sensors.	We need to understand what information flows are useful, how they flows can be configured for different types of data, and how privacy rules can be set.	RQ2_d RQ3_b RQ2_f
Section 5.4.3.2.2 Annex B.1.3.16	**Security** Data can be confidential or even not intended to be recorded at all. The participants identify a potential risk for the safety of personnel or affected persons if disaster situation data is not managed securely, such as revealing geo-locations in some humanitarian missions.		RQ2_d RQ3_b RQ2_f
Annex B.1.3.2	**Using external devices** Most participants are willing to wear small devices that improve communication in the field.	We need to identify what devices facilitate the work of the rescue team and support their collaborative work.	RQ2_a, RQ2_b
Annex B.1.3.2	**Setting up ad managing infrastructure** The majority of participants can see the benefit of spending time establishing networks, but less participants seem interested in maintaining devices in the field.	This may be because there are so many demands for time and effort in a disaster situation that maintaining devices would not be considered as important as other tasks. Yet the value of knowing the status of other team members is clearly recognised as significant by a majority of participants. This is confirmed by the interest in receiving or having access to information about others.	RQ2_e RQ3_a RQ3_b
Annex B.1.3.14	**Usability** In general users accept the improvements of the SOCIETIES system. It has become clear however that such a system must be very unobtrusive, user friendly and easy to manage.		RQ1_b RQ2_e RQ3_a RQ3_b RQ4_*

4.4 Temporal aspects

The trial will address two groups of users, at both end of the value chain: the volunteers on the one hand, and the end-user professionals on the scene. The trial will consist of two phases:

Phase 1: Incorporating only the volunteer community in a variety of experiments to validate the creation of a CIS (member selection) and evaluation of the contribution these volunteers can make in a disaster mission, using specially prepared materials (e.g. translation and image processing tasks) and assignments (control of unmanned areal vehicles). This trial will be spread out over a course of several weeks.

Phase 2: This will last one to two days and will incorporate both the volunteer community and the professional end-users.

The time flow of the Phase 1 trials is less demanding than the one of Phase 2, as no Project-external users have to be integrated. In general there is one overall trial plan which all users follow at the same time, as all users together and at the same time simulate the online volunteer community.

In the Phase 2 trials, the time flow has to be defined clearer to make most efficient use of the available time. Again here, the trial participants – both on-site and off-site – will work in parallel, representing a single team on a single, simulated mission. As such, there will be a single timeline for all actions at the two locations.

In order to collect more data, two consecutive missions shall be worked on in series.

4.5 Physical / location aspects

Phase 1 of the trial will take place virtually, by incorporating volunteers sourced from Societies partners, and possibly their family or friends, at their respective home / work location. We expect a large proportion of these volunteers to be draw from DLR staff. The disaster relief experts among DLR staff We will thereby allow to evaluate one of the findings of the paper trial where in which the interviewees stated that they rather trusted disaster experts in the off-site community than complete strangers.

Phase 2 of the trial will take place at four locations simultaneously:

1. Neuhausen, Germany "Bundesschule Technisches Hilfswerk" – this is the general school of the German disaster relief agency and the place where the DM experts will reside.
2. DLR Oberpfaffenhofen or Neuhausen (outdoors): This is where the "disaster" will be enacted with a bridge or building as the main disaster area to be assessed.
3. DLR Oberpfaffenhofen or Neuhausen respectively: room with a selection of volunteers (online and on-site).
4. "Cyberspace" / Virtual: A selection of volunteers off site (i.e. at home or at their work location)

Users will be provided with the following devices, corresponding to the trial location as described above:

1. Volunteers at DLR Oberpfaffenhofen: standard PC, one or two virtual reality goggles for remote control.
2. Disaster Site: Disaster Management Service installed on two to three notebooks. In addition, there will be a fleet of 3-4 quadrators, autonomous aerial vehicles, to assess the building or bridge, operated remotely by the volunteers. Volunteers on-site will be equipped with SmartPhones and will use the image geo-tagging service.
3. Volunteers off site "Cyberspace": standard PC; in addition, one volunteer will be provided with a set of virtual reality goggles to allow remote operation of the quadrotors on-site.
4. DM experts at Neuhausen, Germany: Disaster Management Service installed on one to two notebooks.

4.6 Third party services

The trials will incorporate the following third party services to be implemented by WP6. These are mapped to the scenes for clarity; *services shown in italics*. A subset of scenes and services will be realised in the first trial allowing the coverage of the high priority research questions (see Section 4.2). The scenes "pre-scene", "scene 1" and "scene 2" have been selected, and as shown in the table below only a subset of services in these scenes need to be implemented in the first trial.

Service Name	Prio-rity	Scene	Service Description/ Functionality	Questions / Evaluation Points	Paper trial feedback taken into account	Compelling Points/ Improvement Potential

Service Name	Priority	Scene	Service Description/ Functionality	Questions / Evaluation Points	Paper trial feedback taken into account	Compelling Points/ Improvement Potential
Service "IWantToHelp" (Volunteer registration and specification)	High	Pre-scene + Scene 1 (needs to cover initial registration of volunteers and how they set up and use the system)	- Register users as a volunteer. - Some introductory information for making it more compelling (use imagination) - Register expertise and interests. - Set up communication means (e.g. Alerts etc.)	EPs from - WP4: platform start-up		Need to consider how the web site should look like in order to promote a feeling of helping. Maybe show the results, badges, socializing with other helpers.
Service "YouRNotAlone" (Volunteer social computing support)	Medium	Pre-scene + Scene 1 (needs to cover initial registration of volunteers and how they set up and use the system)	- Propose user to be joined into some specialized groups (eg "I know Greek" for sign translation) - Crowd-sourcing functionality integrated with on-site services	- RQ1_c - RQ1_d - RQ5_a	- Crowd Sourcing, Collective Intelligence, Trust	Need to consider how the web site should look like in order to promote a feeling of helping. Maybe show the results, badges, socializing with other helpers.
Service "iDisaster" (creation and management of actual disasters by USAR)	High	Pre-scene + Scene 1	- register disaster - upload disaster data (satelite images etc.) - register tasks, specify requirements for a tasks - assign tasks to volunteers (automatic? T5.1) - specify answering inspection mode (related to crowd-sourcing method)	- RQ1_a - RQ1_b - RQ1_c	- Access to better information - Community orchestration, Automation - Usability	
Service "DDC" (Disaster Data Collector) (Used to collect onsite and offsite data about a specific disaster)	High	Pre-scene+	- Integrating sensor input with the service - Read sensor data from body-worn sensors - Store these data - Upload them to shared areas	- RQ2_b - RQ2_d - RQ2_e - RQ4_a - RQ4_c	- General pervasive techniques, and sharing of data – Privacy - Using external devices - Setting up and managing infrastructure - Status sharing & Privacy - Security - Usability	
Service "AnalyzeThis" (allows creation and distribution of analysis tasks to volunteers)	High	Scene 2	- get notified about uploaded data - View data - Annotate data - Send data to volunteers	- RQ2_c - RQ2_d	- Remote Distribution of Tasks - Crowd Sourcing, Collective Intelligence, Trust - Sharing data – Privacy - Security	
Special case of "AnalyzeThis": Service "Image recognition "	High	Scene 1				
Special case of "AnalyzeThis": Service "Photo geo-tagging"	High	Scene 2				

Service Name	Prio-rity	Scene	Service Description/ Functionality	Questions / Evaluation Points	Paper trial feedback taken into account	Compelling Points/ Improvement Potential
Special case of "AnalyzeThis": Service "Translation"	V2	Scene 4				V2 as scene 4 will not be realized in the 1st trial.
Service "Satellite Image Analysis"	High	Scene 1	- Specify images taken by satellite. - Share/import images - Visualize images (in e.g. A web page) - Annotate images	- RQ2_c - RQ2_d - RQ4_c		
Service "Quadrator control"	V2	Scene 2	- Integrating quadrator control with the service - Share control - Provide quadrator feedback to the remote user	RQ2_a		

To illustrate the potential of the service "*Photo geo-tagging*": The geo-tag Insarag markings (see Figure below) is currently used on buildings to indicate the level of damage, the status of pervious observations / actions, and necessary future actions (such as shoring, gas-leakage detection, search and rescue of person trapped in the building). These markings are usually sprayed on building walls. The objective of "*Photo geo-tagging*" is to geo-tag the image (see example in the Figure below), and to let off-site volunteers translate this into machine-readable form. The resulting information can then be used to coordinate the mission on a more global scale.

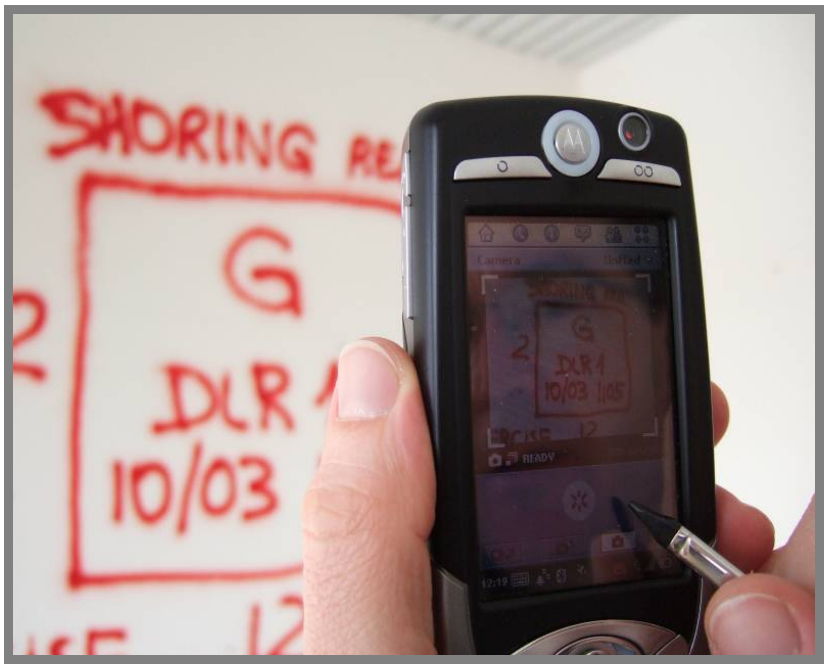

Figure 5: Service example: geo-tag with an Insarag sketch that represents a status of, e.g., a building after an assessment

4.7 User aspects

4.7.1 Number of users

We suggest the following numbers of users:

- Off-site volunteers: 6-10.
- On-site volunteers: 2-4.
- Off-site DM experts: 3-10

The inclusion criteria for the trial participants, along with the demographics of the users shall be decided in the coming months, closer to the actual trials when the availability of experts can be ensured better. This information will be described in the second version of this deliverable.

Users will be involved in planning the trials to some extent. We expect limited involvement of the DM experts and some involvement of the volunteers, by allowing them to experiment with pre-trial versions of the services and devices.

4.7.2 Person-on-person monitoring / mentoring

We expect that 2-3 Societies staff members be present at each of the three physical sites during the whole trial. This will allow some monitoring of users on a one-on-one basis. We suggest that user preparation be 1 hour at each site.

We suggest that a questionnaire on technology acceptance be given to users before the trial. For this we will use standardized models.

4.7.3 User Feedback

Users will be able to provide feedback informally during the trial by communicating directly with the Societies support staff. Formally, they will be interviewed after the trial. The interview will address the same line of questions as the paper trials.

4.8 Observation / Monitoring

Monitoring during trials will be provided by supportive staff such as students working at DLR and other partners. On-site we will cover each room with encompassing wide-angled Full-HD video and audio recording. This will also apply to the on-site disaster area. Isolated normal-range video and audio coverage will follow the person-on-person mentoring.

Volunteers off-site and on-site will be invited to write first-person accounts of their experience during the trial. Personal Monitoring during the person-on-person mentoring will allow the documentation of the users during their service usage with audio / and or written notes.

Additional observation and documentation will include still photographs, screenshots (auto-generated from the applications at specific actions), log files (auto-generated from the applications and the Societies platform at specific actions).

4.9 Formal Performance Metrics and Feedback for WP4, 5 & 6

4.9.1 Disaster Management Group Test Tables

In the following test tables we have taken the analysis from the table in Section 4.2 and created a list of tests, in order of priority. The first column gives the unique test identifier. The second column gives the test prioritisation. The next column gives a summary of the test description. The last four columns attempt to capture more detailed requirements for each of these tests, as they relate particularly to WP4-6.

Test ID	Priority	Test description	Input	WP4 Relation	WP5 Relation	WP6 Relation
D1	H	This test should be conducted at the beginning of the trial. We plan to obtain the volunteer and expert members' opinion on community services, especially automatic community join and community driven assistance for DM.	Question-naire.			
D2	H	This test aims to get users' opinion on the perceived benefit/risks of automatic management of communities of rescuers based on expertise / availability / workload / previous collaborations / human-social preferences (additionally location / task / institution, country / language / sex / etc.). Compared to manual management or semi-manual.	Observation, question-naire.		Proactivity, Context Inference	User interface relating proactivity / autom. Mgt. to the service itself
D3	H	Test whether DM volunteer users would like services to start automatically on their behalf, such as automatic recommendation that their expertise would now be of direct value.	Observation, question-naire.		Proactivity, Context Inference	User interface relating proactivity to the service itself
D4	H	Test whether volunteers and experts would like to join or even be automatically added into sub-communities based on shared criteria.	Observation, question-naire.			
D5	H	Test how the users trust the decision making mechanism provided by SOCIETIES, i.e. should the system ask users for confirm first before any action?	Observation, question-naire.		Proactivity, Context Inference	User interface relating proactivity to the service itself
D6	H	Test whether users trust the privacy control provided by SOCIETIES	Question-naire.		Privacy Mgt.	
D7	H	Test how useful it is to share access to physical resources (e.g. sensor controls, cameras, robotic arms, UAV controls, etc.) in online communities	TBD, logging; Observation, question-naire.		Sensor Mgt. Context Inference	Service efficiency, HMI

Test ID	Priority	Test description	Input	WP4 Relation	WP5 Relation	WP6 Relation
D8	H	Test how useful it is to share information collected from physical sensors (typically those that collect user related information such as location, activity) in online communities	TBD, logging; Observation, question-naire.		Sensor Mgt. Context Inference	Service may range from fully automated DM expert allocation, to just providing support for manual allocation
D8	H	Test if it is useful to provide access to external services (e.g. translation services) in online communities	Comparison with ground truth information (e.g. translation accuracy, speed)			Service efficiency, ease of use
D9	M	Test which are the best ways of setting up a pervasive communication infrastructure for sharing physical resources.	TBD	Network resources		
D10	M	Test whether value is provided by advanced context inference and personalisation.	TBD		Context Inference and Personal-isation	
D11	M	Test whether transition between connected and disconnected operations worked smoothly from a user perspective	TBD, Observation, Logging	Networking		
D12	L	Test whether users would like to be aware of actions or contexts of other community members, and whether such information would affect their own decisions.	Observation, question-naire.			
D13	L	Test whether the volunteers would like to join a random volunteer community without specifying any particular interest.	Observation, question-naire ; logging			How do services that are related to a community (e.g. translation) present themselves to the user?
D14	L	Test whether knowing about other community members would affect volunteers' willingness to join the community.	Observation, question-naire.			

4.9.2 Analysis of the evaluation points set by WP4 with feedback to WP4, 5 & 6

The following table presents a set of evaluation points defined by WP4. For each point, we indicate how they can be related to our storyboard and to our questions. Some evaluation points require extensions of the storyboard; these extensions will be considered in the realisation of the trial prototype. Finally we also provide important feedback to WP4, 5 & 6. For instance, in some cases we describe functionality needed for performing the evaluation or we suggest an alternative approach to trial for evaluation.

Evaluation Group	Evaluation points	Relation to research question / storyboard / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
SOCIETIES platform start-up	Is the user able to start up the SOCIETIES platform with relative ease? Can they perform the tasks of:	CSS creation/installation not covered by storyboard.	We need to clarify if it is possible to require the rescue team to install their CSS. At the moment we had considered that all devices would be set up before he trial.	WP6 needs to provide clear start-up process / service
	• Creating their CSS and adding nodes?	Scene 2 Relates to RQ2_e.		
	• What did the user have to perform to get up and running			
CIS interaction	Interactions with other CIS's: • Was the user able to search for other CIS's to join?	Relates to - RQ1_a (rescue team viewpoint) - RQ1_c (rescue + volunteer viewpoint)		We assume that support for CIS interaction is provided by WP4 (it is not a 3rd party service provided by WP6).
	• Was the user automatically joined to CIS's based on preferences/interests?			
Usability of setting up platform	There will be some basic settings that the user will need to define in a configuration screen: • Was the user able to select the appropriate settings?	Priority: TBD		WP4 should clarify what is meant by basic settings. It has to be clarified whether WP4, WP4 or WP6 5 provide a GUI for configuration.
	• Was there configuration items that they felt were missing?			
User Experience on the mobile device	Based on targeted feature rich device being Android based: • What is the experience for non-Android based devices?	The rescue team members use mobile devices. The volunteers may use PCs (scene 2, 3, 4). Priority: TBD	We should clarify what devices we expect the rescue team to use. In addition we may require "volunteers" to make use of other devices.	
	• What limitations were found that prevented user from participating fully in each of the user trials?			
Scalability of Platform	How scalable is the system when we have large numbers of users interacting with each other. Where are the perceived bottlenecks? • Is our design proving to be scalable? Or • The prototype implementation was not scalable due to minimal deployment and partial system development.	Relates to RQ3_d		We suggest WP4 and to perform provide before this question is addressed in the trial.

Evaluation Group	Evaluation points	Relation to research question / storyboard / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
Portability of platform across all scenarios	We are developing a framework to satisfy 3 separate user groups: • Does the platform seamlessly interact with each user group? • Is the platform segmented with separate functions for each divide?	Priority: TBD		WP4 and WP5: How much tweaking needs to be done to the platform to support the three groups? Does this lead to hard wiring or compatibility problems? Are we loosing platform generality?
Interoperability across multiple platforms	Investigate interoperability of each of the software components with real world environments. • Performance of platform on each OS's of the mobile devices • Several components deployed on different nodes in the cloud. • Where do failures occur?	Priority: TBD		Performance should be addressed through simulations first. We suggest WP4 and WP5 to provide log functionality such as failures can be traced.
Sharing an "item" with feedback	Ability to publish an "item": (similar to Facebook Wall) • Publish a comment/video/file/etc to a known location • Allow users to provide feedback and comments	The storyboard does not use the term "publishing" but rather the sharing of data collected by different sensors. It does use the term "feedback" but rather annotations. This is covered by scene 2. This relates to RQ2_b. Priority: TBD	RQ2_b might be reformulated to describe that what is meant with data sharing. It is not only publishing but also allowing other to annotate the data.	
Service Lifecycle	Was the service container API satisfactory? • Did it allow the creation/deployment of services with easy deployment?	Priority: TBD		WP6 should provide support for service lifecycle.
	• What limitations did 3rd party developers find when creating shared services for the community?			This point should be evaluated by WP6.
Service Registry	Users of the platform interacting with existing services: • easy to browse/search service instances within a registry of services.	The storyboard does not describe explicitly how services are discovered.	This feature will be taken into account during the definition of test scenarios.	How does WP6 support the service discovery process?
	• Ability to search within own CSS/targeted CSS/multiple CIS's [NB! This is different from SW download from a marketplace]			

Evaluation Group	Evaluation points	Relation to research question / storyboard / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
Future Feature Requests	With users getting to interact with software, understanding their desire for future applications/technology solutions based on our presented system: • what features would users like to see? • Features that were felt to be missing?		This is a general point that can be covered by a questionnaire related to the trial.	

4.9.3 Analysis of the evaluation points set by WP5 with feedback to WP4, 5 & 6

The following table presents a set of evaluation points defined by WP5. For each point, we indicate how they can be related to our storyboard and to our questions. Some evaluation points require extensions of the storyboard; these extensions will be considered in the realisation of the trial prototype. Finally we also provide important feedback to WP4, 5 & 6. For instance, in some cases we describe functionality needed for performing the evaluation or we suggest an alternative approach to trial for evaluation.

Evaluation Group	Evaluation points	Relation to research question / storyboard / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
Community/ Individual Recommendations	• Do "communities to individual" recommendations improve over time during the course of the trials? I.e. does the user acceptance rate improve? • Do "individual to community" recommendations improve over time during the course of the trials? I.e. does the acceptance rate improve? • Are recommendations based on data gathered from several sources better than from each of the sources alone? • What is the difference in acceptance rates for different recommendation schemes?	Relates to RQ1_a and RQ1_c Priority: TBD	The test scenario should describe different alternatives for comparing recommendation schemes / sources.	An important question here is whether the time constraints of the trial allow us to perform a sufficient number of tests to make all required comparisons. Time constraints may not either allow us to observe improvement over time.
Manual Communities	• Why do users create CISs? E.g. is it for family, interests, an ad-hoc meeting, etc? • Do user's create more long-term or ad-hoc communities? • What is the most relevant criteria for a user to join an existing community? E.g. friends already members, community purpose, etc. • How many CISs does the average user create?	Question 3 relates to RQ1_a and RQ1_c Priority: TBD		

Evaluation Group	Evaluation points	Relation to research question / storyboard / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
	• How many CISs is the average user a member of?			
Automatic communities	• Do users see the benefit in the automatic creation of ad-hoc communities? • Do users see the benefit in the automatic creation of long-term communities? • Do users like to be automatically addded to ad-hoc communities? • Do users like to be automatically added to long-term communities? • How much information is needed to create an automatic community?	Question 1, 3 and 5 relates to RQ1_a and RQ1_c Priority: TBD		
Community based context data distribution	• How accurate is community context data?	RQ2_b relates to the question of accurateness as inaccurate data are probably not useful.		
	• What are the delays involved in requesting context data from a community? • What are the delays involved in requesting context data from another individual?	For large data set, relates to RQ4_c and RQ4_d. Priority: TBD		Performance is only covered in the case of large data set.
	• Predefined context data distribution policies will exist, on CIS and CSS level. User will be able to further parameterize these policies. To which extend are predefined polices changed?	Relates to RQ2_f Priority: TBD		
Intelligent context history management	• What is the average size of necessary training data sets for efficient proactivity related functionality? • What is the overall processing and storage resources requirements of the context history on an average day of usage? Is this affecting other components?	Priority: TBD		The time constraints of the trial may not allow us to answer these questions. We suggest WP5 to address these questions through simulations.

Evaluation Group	Evaluation points	Relation to research question / storyboard / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
Context Inference	• What context inference mechanisms are producing the most acceptable, by the user, results? • How easy is it for the user to manage their conceptual location information?	Priority: TBD	Relevance of context inference and symbolic location inference needs to be clarified in the scenes.	The time constraints of the trial may not allow us to compare different inference mechanisms.
Sensors and Monitoring	• Do users accept to wear sensors (obtrusiveness)?	Relates to RQ2_a and RQ2_b.		
	• Do users accept to be monitored (privacy)?	Relates to RQ2_d.		
Context Prediction	• What is the rate of acceptance of the predicted context by the users? How accurate are the predictions?	Priority: TBD		See context inference above.
	• Is there an impact in terms of memory and processing load, or serious delay to the user?	Will not be addressed.		A better answer to this question can be acquired through simulation than through a trial.
Context Inheritance	• What is the rate of acceptance of the context inherited by the users from CISs? • How accurate are the context inheritance conflict resolution mechanisms?	Priority: TBD	This feature is not described in the scenarios. A close feature is described in scene 5 with transfer of data between teams.	We suggest WP5 to propose extension to scene 5 so this feature can be covered.
Context Similarity	• Can meaningful evaluations be achieved in a (near) real time solution?	Will not be addressed.		This question should be answered through simulation.
	• Can evaluations model the dynamic CSS/CIS interactions to provide useful data sets to use as a base for further predictive analytics?	Priority: TBD		We suggest WP5 to provide log functionality such as data can be collected. However following the time constraints of the trial, it may be difficult to collect sufficient data sets.
Preference Management	• Do preferences become more accurate over the course of the user trials? Do users abort the preferences fewer times? • On average, how many preferences will users acquire over the course of the trials?	Priority: TBD		Time constraints in DM trial.

Evaluation Group	Evaluation points	Relation to research question / storyboard / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
Preference Learning	• Are incremental learning algorithms able to run continuously for the duration of the trial? • How often are batch preference learning executions scheduled? • What learning algorithm performs best for preference learning?	Priority: TBD	The storyboard does not explicitly describe the feature of preference learning.	WP5: Check possible resource constraints. Perhaps perform simulations.
Personalisation GUI	• Do users use the Personalisation GUI to create, edit and delete preferences? • How often do the users use the Personalisation GUI to create preferences? • How often do the users use the Personalisation GUI to edit learnt preferences?	Priority: TBD	The storyboard does not explicitly describe the feature of personalisation.	
User Intent	• Does the user intent model become more accurate over the course of the trials? Do users abort the intentions fewer times? • What is the confidence level at which users will allow an intent prediction to be performed automatically? • What is the average size of the intent model for all trial participants at the end of the trial? • How does the user's intent model change (grow?) during the course of the trial? • How much data is necessary to achieve noticeable personalisation? • How do the proposed criteria perform?	Priority: TBD	The storyboard does not explicitly describe the feature of personalisation. It needs to be updated to include personalisation aspects.	Note that for questions that require acquisition of data over time, the time constraints in the DM trial may be too restrictive.
Privacy Management	• Does the user trust the privacy management of the system? If yes, what is the main feature providing this trust?	Relates to RQ1_d, RQ2_d and RQ3_b Priority: TBD	The storyboard does not describe explicitly that the user can make use of a privacy management system. This feature will be taken into account during the definition of test scenarios if supported by the platform.	As privacy was an important concern during the paper trial, a resulting recommendation is to provide the user support for the configuration of privacy rules. WP4/5 should provide such support to the user.

Evaluation Group	Evaluation points	Relation to research question / storyboard / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
Privacy Policy Negotiation	• How often does the user make any changes to the privacy policy suggested by the service?	Priority: TBD	This is not covered by the storyboard.	WP5: We do not have a clear understanding of "privacy policy" related to a service. The main concern up to now has been related to privacy related to data sharing in a CIS. Does service privacy policy mean that a service may store data in a CIS following different policies? Does it mean that access to service/its operations is controlled by policies?
	• How many times does the privacy policy negotiation failed because of the user not accepting the terms and conditions of the service?			The question may require observation over time. The time constraints in the DM trial may be too restrictive to get an answer to this.
Privacy Preference Manager	• Does the user use the privacy GUI to create new privacy preferences? • Does the user use the privacy GUI to change existing preferences? • Do users understand the concept of privacy preferences? If yes how much it is appreciated?	Priority: TBD		WP5: These points overlap with the points under "Policy Management" above.
Data Obfuscation	• Do users understand the concept of data obfuscation? If yes how much it is appreciated? • How many times does the user modify the obfuscation level in their privacy preferences? • Do most users use data obfuscation? • What is the average level of obfuscation per data type chosen by users?	Priority: TBD	Obfuscation is not covered by the storyboard.	
Active Privacy Assistant	• Would end user benefit from a tool that could warn him that he is about to share media containing sensitive information to inappropriate users or user groups?	Priority: TBD	This can be included in a test scenario - in relation with data sharing. Currently the storyboard does not explain when rules for sharing are set up and by whom	
Visual Trust Indicators	• Would it be helpful for end user to have services fit with a clear visual estimation (e.g. a semaphore and/or progress bar) of trust for each service so he could filter the services by a trust threshold and sort services by trust level? • Would it be helpful for	Priority: TBD	Trust indicators can be used for displaying the reliability of a crowd input for a DM service.	

Evaluation Group	Evaluation points	Relation to research question / storyboard / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
	end user to have other users fit with a clear visual estimation (e.g. a semaphore and/or progress bar) of trust augmented for each user within SAG view?			
User Agent	• How often do the personalisation sources suggest conflicting actions?	Priority: TBD	The storyboard does not explicitly describe the feature of personalisation.	WP5: These points overlap with the points under "User intent" above.
	• Which personalisation source suggests the correct action most often?			
	• Is the User Agent able to resolve the conflict among different internal proactive components?		This feature does not require change to the scenario.	WP5 should provide support such that conflicts are detected and logged.
	• Do users accept automatic decision making? To what extent?	Relates to RQ1_a and RQ1_c.		
	• Do users accept the automatic installing/execution/provision of services? To what extent?		Automatic service management is not explicitly described in the scenario.	WP6 should provide support for service lifecycle
User Feedback	• Do user feedback notification alerts become less frequent as personalisation models become more accurate?	Priority: TBD		WP5: We do not have a clear understanding of what these alerts are about. Does this mean that the user should report when they are not satisfied by the system pro-activeness? The time constraints in the DM trial may be too restrictive to get an answer to this. Further it might be quite complex to determine the "accuracy" of personalisation models.
	• As notifications alerts become less frequent, is the user relieved not to receive so many or upset at not them?			The time constraints in the DM trial may be too restrictive to get an answer to this.
	• Does the user find notification alerts annoying and as a result ignore them?			Same comment as above (first question)
User Feedback	• Does UAV steering work with an acceptable delay for remote controllers?	Relates to RQ2_a. Priority: TBD		

5 Enterprise Group Trials

5.1 Summary of the trials

The Enterprise trial will take place during the annual Intel European Research & Innovation Conference (ERIC). The 2011 Intel ERIC will take place at the Intel Ireland, campus in Leixlip, Ireland from October 12th-14th. The project will you use this event to gather information about the attendees and to promote our user trial for the follow-on ERIC in 2012.

The theme of the 2011 conference is 'Building a Smart, Sustainable and Inclusive society through Research and Innovation partnership' and there will be a number of distinct focus areas included within the event. The details of this event can be found at:

http://www.intel.com/corporate/education/emea/event/irc/

The detailed planning for the 2012 ERIC conference will begin after the 2011 conference has concluded but at the time this deliverable was written the 2012 ERIC should also take place over three days in October at Intel's Ireland campus.

As this is a short window of opportunity we need to ensure that proper planning is in place well in advance. Our scenarios include the conference organiser as a stakeholder as well as the delegates at the conference. This deliverable will specify what Societies will want this stakeholder to test as part of our first user trial. We will need to establish a close liaison with this stakeholder so as to ensure that if we require the conference organiser to test any conference organisational tools (as described in the Pre-conference storyboards) in advance of the October date that the consortium will have these tools available in a timely manner.

During the three days of the conference we will need to ensure that sufficient training material has been prepared and that there are enough people available on the ground to offer assistance to conference delegates if and when required.

The trial itself will involve presenting the conference attendees with a selection of the services that were previously presented to them as part of our paper trial storyboards.

These services will be described in more detail in section 5.6 "Third Party Services".

5.2 Objectives of the trial

Objectives of the first Enterprise evaluation user trial:

a. Research Objectives: where we construct demonstrators that can be used to measure the value of our system

b. End User Objectives: where we evaluate the usability of our system from our users perspective

c. Technical Objectives: where we implement integrated prototypes of our higher system so that we can evaluate technical constraints or issues

5.2.1 Research evaluation objectives for our Enterprise Community

- *E-RO1: To extract context rich data sets so as to create communities of interest.*
- *E-RO2: To investigate how interaction with physical resources, using physical context can be used for initiating and supporting social interactions.*
- *E-RO3: To investigate how are relevant members (people) of a CIS discovered.*
- *E-RO4: To investigate how we can address trust issues, in particular, how professional end-users trust the integrity and accuracy of the services.*

- *E-RO5: To understand how much automation is accepted by end-users and how the end-users trust this automation.*
- *E-RO6: To investigate the characteristics of communication and collaboration in a conference scenario.*
- *E-RO8: To understand the impact of privacy concerns to the functionality of the system.*

5.2.2 User evaluation objectives for our Enterprise Community

- *E-UO1: To investigate what levels of usability and what user interaction metaphors should be supported.*
- *E-UO2: To evaluate an easy-to-use privacy layering interface.*

5.2.3 Technical evaluation objectives for our Enterprise Community

- *E-T01: To investigate the level of interoperability constraints of the software components within real-world environments*
- *E-T02: To evaluate the performance and scalability of the designed system (but not solely of the implemented prototype)*
- *E-T03: To understand the user's desire for future applications and technical solutions of the presented system*

5.3 Analysis of the Enterprise Paper Trial Evaluation

This section takes the recommendations identified in Annex B of the deliverable D8.1 "Paper Trial Evaluation Report" and uses them to produce test descriptions that we will use to specify, measure and evaluate a series of tests in our first User Trial.

5.3.1 D8.1 Enterprise Storyboard Responses

In our Paper Trial in April 2011 we presented to our Enterprise Community a low-fidelity graphical representation of our storyboards. During this presentation a number of questions were inserted that related to the storyboard scene, and we asked our users to give us their individual feedback to these questions. We summarised these responses in Annex B of the deliverable D8.1 "Paper Trial Evaluation Report" and we have used these responses to build the following tables. The first column gives a reference to the scenarios, which were detailed in deliverable D2.2 "Scenario description, use cases and technical requirements specification". The second column gives a summary of the paper trial feedback from our users for the referenced scenario. The third column presents a summary of the recommendations from the Enterprise paper trial, which was relayed back to the consortium. The forth column gives the unique identifier for the future User Trial test relating to this scene. The next column lists the evaluation objective references (identified in section 1.2) that relate to this test. The final column gives a prioritisation (H=high, M=medium and L=low) for this test.

Source in D8.1	Summary of user feedback from D8.1	Summary of D8.1 discussion and recommendation	Test ID	Priority
COS_01 Scene 1	The **users surveyed recognised the value** of pre-conference data capture of users' requirements including; dietary, logistical information, such as conference access details, and accommodation.	This forms part of our system's ability to create "context rich data sets" and we could test this to see how much data they would be prepared to give us	E1	H
COS_01 Scene 2	Most respondents were favourable to sharing their preferences and requirements with conference organisers. This reaction is not surprising given that the attendees want to minimise problems such as accommodation and food choice, while maximising their effectiveness at the conference by determining in advance meeting opportunities and points of interest at the conference.	As above	E2	H
COS_01 Scene 3	Respondents were mostly interested in receiving information that reflects their expressed interests, i.e. conference events, speakers and **people and/or organisations that they perceive as useful to meet** at the conference. The production of a personalised conference agenda	**Perceive as useful to meet**: This is similar to our notion of "finding relevant people and things" and a key selling feature of the system	E3	H

	that gives more information related to the speakers and other attendees, along with the highlighting of possible agenda clashes where multiple tracks are being presented was found to be beneficial. Allied to this, the navigation between multiple venues and average time to travel between these locations were perceived as being useful. Interest was also shown in personalised direction from the attendee's hotel to the conference and other amenities such as restaurants and sightseeing.			
COS_02 Scene 1	Attendees should be able to review the group's members before joining the group rather than being automatically added and causing offence by removing themselves from the group afterwards. There should be a co-ordination in the formation of these groups rather than every attendee forming groups and attendees being subscribed to many overlapping groups. It is unclear whether the conference or selected attendees would act as the organisers of these groups and it does highlight the main problem with static group formation. The formation of dynamic groups based on criteria such as location, i.e. "flash mobs" was also shown to be desirable. One strong requirement was the **ability to filter the formed groups** and create smaller sub-groups to make the networking experience more focussed.	Auto-add test. Have a sub-group who are auto-added and see how they react to this. Who do we want to allow to create communities? Test "Group Filtering" functionality.	E4 E5	M M
COS_02 Scene 2	There is a potential issue with users accepting invitations to join communities, generally based around how well known (trust - e.g. eBay gives users the rating of the merchant) the invitee is.	Test how much of an issue this is.	E6	H
COS_02 Scene 3	The mechanics of sharing taxi fares with strangers, the preference to share with colleagues and the perceived offence in turning down the invitation to share the taxi were all cited. The ability to screen the travel arrangements and decline **without the other taxi occupants**	Test how much of an issue this is.	E7	L

	being aware was the main requirement. The stronger case of being allowed to opt-out of the shared transport arrangements was also expressed.			
COS_03 Scene 1	It should be noted that there were privacy concerns raised with movement tracking systems.	Test how much of an issue this is.	E8	M
COS_03 Scene 2	One principal finding is that while attendees want as much information as possible **they are more reluctant to disclose information and want to be consulted on disclosures rather than the Societies devices making the decisions**.	Auto-decisions being taken by the system should be evaluated.	E9	H
COS_03 Scene 3	A general comment was added by one user, which stated that 'security is the big question and would they trust sensitive data to a third party'.	How do we get users to trust that their data is secure?	E10	H
COS_03 Scene 5	A user questioned whether the SAG **view would get cluttered** with all this information and suggested that sharing this level of information should be voluntary and the **users should have control** over who or what groupings or people can see it. **Privacy settings** would be very important for all of this.	We can test this	E11	L
COS_03 Scene 5	The users suggested some alternatives to the SAG including; just using a device, e.g. smartphone, with a camera and an overlay view on the screen; an overhead projector onto a huge wall, highlighting people every few minutes; or a digitised room mapping with info pins as you get closer to people.	Do we want to test these alternatives?	E14	M
COS_03 Scene 5	Some users question the practicality of using the goggles for people who already wear glasses and also due to the reduced eye contact.	How do we test whether this is practical or not?	E15	L
COS_03 Scene 6	Some of the **useful aspects** included being able to quickly find people with similar interests, getting to know what others are doing, fast tracking into conversations of interest, making face-to-face connections with	Can we measure how useful this is?	E16	H

	relevant people, business and exchange, understanding which attendees are in which zone and 'check-in' to let others know you are joining the discussion.			
COS_03 Scene 8	Others stated that **technology may get in the way of natural conversations** and that **'trust is a very human behaviour or inference and probably can't and shouldn't be inferred through technology'**.	We definitely have to prove/disprove or at the very least measure this	E17	H
COS_04 Scene 1	Many users said **this service would be useful** as paper based notes either get lost or disconnected over time. There are many situations where conversations occurred where you may have failed to exchange business cards and there are many conversations at these types of events. It is difficult to remember everything, although one user would still like to have their own notes.	Should we test how useful	E18	L
COS_04 Scene 3	Some users stated that **this feature would be useful** but with some conditions, such as if it was only related to their interests/expertise, depends on how busy they are.	We should test how useful this is.	E19	L
COS_04 Scene 3	**'As a user I would be inclined to disable features that attempt to "push" information to my device too often.'**	Is this a WP5 test? If so, more detail required by WP5 on this.	E20	H
COS_04 Scene 3	There **should be an opt-out clause to stop receiving information**, subject to company policy conflicts (email notifications/calendar invites), follow on meetings perhaps but would be less inclined towards group travel etc.	Is this a WP5 test? If so, more detail required by WP5 on this.	E21	H
COS_04 Scene 3	'It is **unlikely that I would have complete confidence in a system** that attempts to predict the next steps in terms of meetings and interactions.'	In our trial, how do we go about "growing" user confidence so that they do trust the system? Is this something we want to do anything with in the 1st trial?	E22	H

5.3.2 D8.1 Enterprise Questionnaire Responses

In addition to the individual questions asked as part of the storyboard presentation, as described in the previous section, we also pitched some questions as part of a group discussion with our Enterprise Community. These responses have also been summarised in Annex B of the deliverable D8.1 "Paper Trial Evaluation Report" and we have used these responses to build the following tables. The first column gives the question that was asked to the group. The second column gives a summary of the feedback from the group. The third column presents a summary of our recommendations, which were relayed back to the consortium. The forth column gives the unique identifier for the future User Trial test relating to this recommendation. The next column lists the evaluation objective references (identified in section 1.2) that relate to this test. The final column gives a prioritisation (H=high, M=medium and L=low) for this test.

Question	Summary of feedback	Discussion and recommendation	Test ID	Priority
Would you find it useful to share resources and applications within a community? And if so would you have any privacy concerns? What would you need to know to be reassured?	This user group is **not convinced about the value of application sharing**. This is possibly because the examples we have in the scenarios are just not compelling enough. The sharing of resources depends on the type of resource.	What do we want to do with this? Maybe T24 has some revisions to make to the scenarios to make this more compelling	E23	L
As above	Sharing information would be useful, e.g. an attendee's travel information, but **sharing applications may not be as useful**.	Obvious test would be to find out how useful.	E24	M
As above	Ensure that the default setting is private for everything from the outset. By inference it is important that the UI for privacy preferences is **easy to use and understand**, which is an important consideration for the project.	Test how easy. People are quite comfortable with the sharing of public information, e.g. with LinkedIn. Uncomplicated **privacy preferences where the default setting is private on all aspects would provide a level of reassurance to users, and thus allowing them to make information public as they wish**. Company policies may prohibit the sharing of sensitive information and resources particularly if it is to applications that are externally hosted - federation of data storage could help here.	E25	H
As above	The "**IM type status**" or some other solution should be seriously noted as a recommendation, so	If this is implemented then we need to test how successfully it works	E26	M

Question	Summary of feedback	Discussion and recommendation	Test ID	Priority
	as to get around our user concerns regarding showing availability and information sharing.			
How would you like the SOCIETIES system to manage your home and work profiles?	It would useful if you could tag people as work/personal but there still could be overlaps of information that you may want to share. This is a common issue in current social networks, although aspects like Facebook groups do help with this. There is **a grayness between 'home' and 'work' connections**, thus there is a need to address this.	Does the project want to look into addressing this? If so, do we want to test anything in relation to this for the 1st user trial?	E27	L
As above	Similarly, with the **data storage privacy concerns**.... **'knowing that you are in control' is more reassuring** and **extra complexity with preferences doesn't always help**. Again once there is **'an easy opt-in/opt-out process'** and you can easily choose what you want to share. Under certain situations you may not want to share your location, e.g. if you are on holidays, but in other situations you will, e.g. if you want to meet people.	Are we implementing this "opt-in/opt-out process"? If so, we can test for ease-of-use.	E28	M

5.3.3 D8.1 Enterprise General Discussion

We concluded the Enterprise Community Paper Trial with a general discussion with the group and a summary of this can be found in Annex B of the deliverable D8.1 "Paper Trial Evaluation Report". We have used this summary to build the following tables. The first column presents a summary of our recommendations, which were relayed back to the consortium. The second column discusses further this recommendation and where applicable offers a potential test for future trials. The next column gives the unique identifier for the future User Trial test relating to this recommendation. The next column lists the evaluation objective references (identified in section 1.2) that relate to this test. The final column gives a prioritisation (H=high, M=medium and L=low) for this test.

Group Discussion Recommendation	Test discussion/description	Test ID	Priority
The consortium needs to consider and discuss how “trust” of the system will be presented to the users. 'Trust' was a big discussion point within the group where it was suggested that there are different levels of trust and it is a term that can be mis-used. Also how trust is rated could be something difficult to always get right, e.g. 'how does a trust level infer a quality of engagement?'	How do we test whether our users trust the system. We need to discuss this further	E29	H
Review the practicalities and value on whether we really should be investing project resource on the socially aware goggles. This group seemed to say that we shouldn’t, which would mean revising the scenarios accordingly.	We should discuss the value in investing project resources on the socially aware goggles.	E30	L
It was stated that **a lot of human interactions are random and cannot be easily predicted. This system may inhibit these natural interactions**, e.g. like when users follow a Sat Nav but don't use their own natural navigation skills. Thus "**a user could become a slave to their preferences and may lose the opportunity to discover new opportunities".**	Is this even measurable?	E31	M
"The system as a whole has a lot of innovation although **some services in isolation can be achieved today**" and "there are existing applications that are similar but the SOCIETIES system would be well beyond the state-of-the-art”.	This appears to be a user perception, but it might be novel to try and see how far beyond SoTA they think the system is.	E32	L
It was suggested that **if any attendee can create a sub-community for any topic then this may not be scalable**. It could be better for the organisers to create sub-communities or **existing communities could be advertised to avoid duplication**. Organisers could send invites to particular people so sub-communities could be clear from the start or **communities could be theme or relationship based**.	Scalability test	E33	H
CIS SPAM filter required	Compliancy test	E34	L

5.4 Temporal aspects

The trial will address the attendees at Intel's 2012 European Research & Innovation Conference (ERIC). The conference will be over three days in October, most likely at Intel's campus in Leixlip. There will be a requirement to set up the conference space with the equipment necessary to conduct the trial (e.g. indoor location tracking system) and this will probably happen the week before the actual conference. At the time of writing it is not intended to have any pre or post conference services available for the trial but we do intend to include conference organisers as evaluators of the system from an attendee perspective.

There will also be sufficient numbers of Societies personnel on the ground to support the attendees with any issues as they might arise.

ID	Task Name	Duration	Start	Finish
1	**1st Enterprise User Trial plan**	**271 days?**	**Fri 30/09/11**	**Fri 12/10/12**
2				
3	**D82**	**131 days**	**Fri 30/09/11**	**Fri 30/03/12**
4	D82 version 1 Board release	0 days	Fri 30/09/11	Fri 30/09/11
5	D82 version 1 CEC release	0 days	Fri 14/10/11	Fri 14/10/11
6				
7	WP4 input to D82	100 days	Fri 14/10/11	Thu 01/03/12
8	WP5 input to D82	100 days	Mon 17/10/11	Fri 02/03/12
9	Finalise D82	10 days	Mon 05/03/12	Fri 16/03/12
10	D82 version 2 Board release	0 days	Fri 16/03/12	Fri 16/03/12
11	D82 version 2 CEC release	0 days	Fri 30/03/12	Fri 30/03/12
12				
13	WP7 release to WP8	0 days	Fri 17/08/12	Fri 17/08/12
14				
15	**WP8**	**40 days?**	**Mon 20/08/12**	**Fri 12/10/12**
16	WP8 conformance testing	10 days	Mon 20/08/12	Fri 31/08/12
17	M15: meets conformance	0 days	Fri 31/08/12	Fri 31/08/12
18	Preparation materials for trial	20 days	Mon 03/09/12	Fri 28/09/12
19	Onsite setup, Leixlip	5 days	Mon 01/10/12	Fri 05/10/12
20	ERIC 2012	3 days?	Wed 10/10/12	Fri 12/10/12

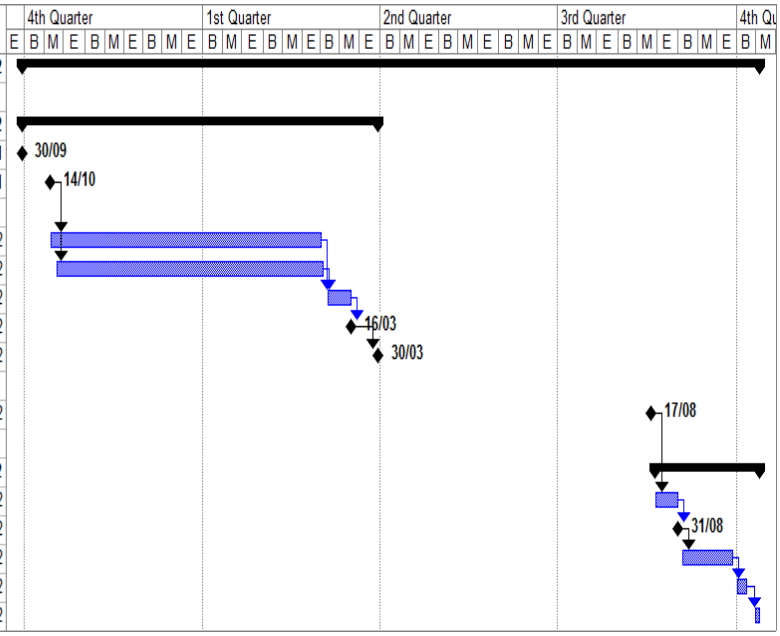

Figure 5: GANTT chart for the Enterprise User Trial

5.5 Physical / location aspects

It is not possible to say precisely at this time where the trial will take place but it is most likely to take place at the Intel campus in Leixlip, Ireland. Confirmation of the exact location of the 2012 ERIC conference will not be known until June/July 2012.

There is a requirement that the site will include an indoor tracking system and Intel have confirmed that this will be in place for the trial, most likely to be an RFID based system. Wireless networks and large displays (to present the Networking Zones) are already in situ at the Leixlip campus. There will be up to 600 delegates at the conference and this will also include our users from the paper trial. However, our intention is to try and include as many of the delegates as we can in this first trial. We will be using the 2011 ERIC conference as a springboard to advertise our intentions and enthuse support for the following year. We know from analysis of previous conferences that this group will be carrying their own devices, including laptops, notebooks and tablets but more importantly we also know that the mobile OS statistics were; Blackberry 39%, iOS 39%, Android 14% and other 8%. It is our intention to use the 2011 ERIC to update these statistics.

5.6 Third party services for WP6

The trials will incorporate the following third party services to be implemented by WP6. These are mapped to the scenes for clarity; *services shown in italics*. A subset of scenes and services will be realised in the first trial allowing the coverage of the high priority research questions (see Section 4.2). Part of the "Registration & Welcome" scene and all of the "Professional Networking" scene have been selected.

Service Name	Priority	Scene	Service Description	Evaluation Points	Paper trial feedback taken into account	Compelling Points and Improvement potential
Service "Conference Registration"	High	COS_03 Scene1: Welcome & Registration	- Mary has already pre-registered for the conference - As such, her CSS should already be known - The conference CIS has already been established - The conference infrastructure auto-senses Mary's arrival - So Mary is sent a "Welcome message"	- Ensure that the default setting is private for everything from the outset. By inference it is important that the UI for privacy preferences is easy to use and understand, which is an important consideration for the project.	- Privacy concerns were raised in ref to movement tracking systems	The user should only see subtle notifications on their device on community suggestions, with user interaction restricted to an optional button to click for joining and one for finding out more about the CIS

Service "Hello world"	High	COS_03 Scene6: Professional Networking	- Mary sets the profile info about herself that she wants the other delegates to see, or is limited to access to only "selected community of attendees" - This can include location awareness, professional interests etc	- People are quite comfortable with the sharing of public information, e.g. with LinkedIn. Uncomplicated privacy preferences where the default setting is private on all aspects would provide a level of reassurance to users, and thus allowing them to make information public as they wish. Company policies may prohibit the sharing of sensitive information and resources particularly if it is to applications that are externally hosted - federation of data storage could help here.	- Similarly, with the data storage privacy concerns.... 'knowing that you are in control' is more reassuring and extra complexity with preferences doesn't always help. Again once there is 'an easy opt-in/opt-out process' and you can easily choose what you want to share. Under certain situations you may not want to share your location, e.g. if you are on holidays, but in other situations you will, e.g. if you want to meet people.	- "Some useful aspects included being able to quickly find people with similar interests, getting to know what others are doing, fast tracking into conversations of interest, making F2F connections with relevant people, business and exchange, understanding which attendees are in which zone and 'check-in' to let others know you are joining the discussion"

Service "Network linkages"	High	COS_03 Scene6: Professional Networking	- Mary's CSS presents her with info about people around her that she has a pre-identified linkage with - These people might be part of her existing SNS contacts - The list will be ranked - These people might not be part of any existing SNS with her, but they might unknowingly share common professional interests - These could be presented as business cards - The list will be updated as other delagates arrive at the conference - Mary "can trigger a linkage" as she "scans the list of current attendees"	- "The system as a whole has a lot of innovation although some services in isolation can be achieved today" and "there are existing applications that are similar but the SOCIETIES system would be well beyond the state-of-the-art".	- This appears to be a user perception, but we will to measure how far beyond SoTA they think the system is.	
Service "SAG"	Low-Medium	COS_03 Scene6: Professional Networking	- The service above might be presented to Mary via the Socially Aware glasses	- We should measure the value in investing project resources on the socially aware goggles.	- Review the practicalities and value on whether we really should be investing project resource on the socially aware goggles. This group seemed to say that we shouldn't, which would mean revising the scenarios accordingly.	

Service "Ad hoc meetings"	High	COS_03 Scene6: Professional Networking	- Ad hoc meeting suggestions will be schedule aware - That is to say the service will know when Mary is free for a meeting - Mary can configure not to want to be invited to meetings (eg to avoid causing offence)	- Evaluate how users feel about the system negatation of these natural human engagements	- It was stated that a lot of human interactions are random and cannot be easily predicted. This system may inhibit these natural interactions, e.g. like when users follow a Sat Nav but don't use their own natural navigation skills. Thus "a user could become a slave to their preferences and may lose the opportunity to discover new opportunities".	
Service "Networking zone"	High	COS_03 Scene6: Professional Networking	- "Are available for users of common interests" - Based on "pre-defined most common interests" - "Themes and their zones are communicated to users as they are activated" - "Overlay virtual signposting can direct attendees to the correct zone" - "Attendees can check-in to show who they are, when they are in the zone area"	- Evaluate whether our users trust the system.	- The consortium needs to consider and discuss how “trust” of the system will be presented to the users. 'Trust' was a big discussion point within the group where it was suggested that there are different levels of trust and it is a term that can be mis-used. Also how trust is rated could be something difficult to always get right, e.g. 'how does a trust level infer a quality of engagement?'	

5.7 User Aspects

5.7.1 Number of users

We suggest the following numbers of users:

- Minimum number of delegates supporting the trial will be 15.

As mentioned in a previous section, our intention will be to engage as many delegates at the conference in the trial as we can. At the time of writing there are discussions in the technical work-packages about targeting the Android platform for the first trial. This may or may not have implications for our user base and we will be in a much better position to know how many delegates this might affect when we update our mobile ownership statistics at ERIC 2011.

5.7.2 Support

Societies staff members will be present at ERIC 2012 during the whole trial. This could allow some monitoring of users on a one-on-one basis but this will be detailed later in deliverable D8.3 "Specification of usability testing for 1st prototype".

5.8 Observation / Monitoring

Monitoring during trials will be provided by Societies personnel (4-5 people). From previous experience we know that there are issues with how we monitor and observe this group. We were allowed to take photographs of people in the paper trial but not video. Audio recording during interviews was permissible but the tapes could not leave the campus and Intel personnel had to do transcripts, which was obviously very time consuming. Personal Monitoring was acceptable for the initial user requirements gathering and will also be supported for our first trial as well. Questionnaires will be used to capture user related feedback.

Log files (auto-generated from the applications and the Societies platform at specific actions) from the delegates' mobiles will be used to capture other evaluation data (to be specified in version 2).

5.9 Formal Performance Metrics and Feedback for WP4, 5 & 6

5.9.1 Enterprise Test Tables

In the following test tables we have taken the analysis from the tables in Section 1.3 and created a list of tests, in order of priority. The first column gives the unique test identifier. The second column gives the test prioritisation. The next column gives a summary of the test description. The last four columns are attempting to capture more detailed requirements for each of these tests, as they relate particularly to WP4-6.

Test ID	Priority	Test description	UI description	WP4 (Component description)	WP5 (Component description)	WP6 (Service description)
E1	H	This is a Conference Organisation test, so would be tested in	Email could be enough here.		What personal data is needed to	TBD

		advance of the actual conference event. Subject to Intel approval, we could capture email addresses of delegates at the 2011 ERIC who would be interested in participating in our 1st user trial at the 2012 ERIC. This could be a very simple test to measure acceptance to share pre-conference personal data. We could set up an invite to join the Conference Community and as part of the invite ask for this personal data.			create the context richness required?	
E2	H	Also a Conference Organisation test.	Again, could be done through email.		In addition to the above, is there any other personal data is needed, e.g. preference data	TBD
E3	H	From the data gathered in E1 & E2 the system should be able to produce a personalised agenda (events, speakers, people/organisations, etc). What will be achievable here by Y2 from the system to realise this? The answer to this will help to shape this test. Ultimately it would be necessary to test how accurate our user community felt the system was in meeting this requirement.	TBD	TBD	TBD	TBD
E6	H	This is a Pre-conference test but we will probably have to do it as an "At the Conference" test - TBD. What we are trying to test is how much of an issue being invited to join a community actually is. TBD but could include invites	TBD – through Smartphone app	TBD	TBD	TBD

		being sent to join various communities ranging from a community with a specific interest, to one with some generic interest, to one with absolutely no interest. What data exactly do we want to collect with this test? Is it just the level of user response to the invite or do we want to capture other data (e.g. our users thoughts about this)				
E9	H	TBD how will we design a test that evaluates our users concerns over system decision making?	TBD	TBD	TBD	TBD
E10	H	This test relates back to E1 & E2 in terms of personal data. The test is different in so far as we are trying to measure how much users trust the integrity of the system to keep their data secure.	TBD	TBD	TBD	TBD
E16	H	We want to set up some services (as per the WP6 column) and measure their usefulness.	TBD – through smartphone app	TBD	TBD	- finding people with similar interests - getting to know what people are doing - fast tracking into conversations of interest - business and exchange - understanding which attendees are in which zone - 'check-in' to let others know you are joining the discussion
E17	H	We are trying to measure how our users would interact with our systems suggestions on so-called "trusted connections" from data-mining SNs etc.				Service that mines SNs of delegates and graphically represents users connectedness.

E20	H	Users were concerned at being bombarded with too many requests.			WP5 to provide more details on this test description	
E21	H	This refers to users wanting a 'Opt-out' to stop receiving information. Similar to E20.			WP5 to provide more details on this test description	
E22	H	Refers directly to WP5 learning and users confidence in the system to make accurate predictions about them and their needs.			WP5 to provide more details on this test description	
E25	H	We need to test and measure how easy our privacy UI is to use	TBD			
E29	H	TBD – this is probably a suite of tests on how much our users would trust our system enough to use it fully				
E33	H	Scalability test around Community creation. Do we want to let any of our delegates to create communities or will this just be limited to the Conference Organiser.				
E4	M	Auto-add test. Have a sub-usergroup who are auto-added to a community and observe their reactions to this.				
E5	M	There are a series of tests around community creation needed; static versus dynamic community creation; testing the "group filtering" functionality.				
E8	M	A test should be devised that evaluates the privacy concerns raised with movement/location tracking.				
E14	M	Testing alternative devices to the SAG device; e.g.: using the smartphone to present the				

		overlay to the user, projected onto a wall, digitised room maps with info pins.				
E24	M	Set up a test that offers users the ability to share their information with other members of their community and test how useful this sharing functionality is.				
E26	M	"IM type status" functionality needs to be implemented and tested so as evaluate user concerns regarding showing their availability and information sharing.				
E28	M	Similar to previous test, we should offer our users an easy way to "Opt-in/Opt-out" of letting the system know and track their location and test what situations (if any) they use this.				
E31	M	The test involves evaluating how much our system inhibits natural human interactions (i.e. it offers to automate this process on ours users behalf).				
E7	L	The test involves evaluating the concerns around sharing of services, such as the taxi service. Paper Trial users were concerned about how this would work and how it would be presented, i.e. was there some way to decline sharing the service without the other person knowing.				
E11	L	This test involves evaluating the user response to the SAG display, it's a usability test.				
E15	L	Test the practicality of the SAG for people who				

		already wear eyeglasses.				
E18	L	Test the usefulness of the important information recorded by the system at the conference (i.e. all of the pieces of information a person would write down or try to remember).				
E19	L	Test the usefulness of post-conference ad hoc networking (e.g. creating follow-on discussion communities).				
E23	L	This tests whether users see a value in sharing their applications				
E27	L	This tests the greyness between 'home' and 'work' connections.				
E30	L	Test the overall practicality of investing project resources implementing services for the SAG.				
E32	L	Evaluate how our user community feels that our system is beyond SoTA.				
E33	L	Test the CIS spam filter				

5.9.2 Analysis of the evaluation points set by WP4 with feedback to WP4, 5 & 6

The following table presents a set of evaluation points defined by WP4 (quantitative in orange and qualitative in blue). For each point, we indicate how they can be related to our storyboard and to our questions. Some evaluation points require extensions of the storyboard; these extensions will be considered in the realisation of the trial prototype. Finally we also provide important feedback to WP4, 5 & 6. For instance, in some cases we describe functionality needed for performing the evaluation or we suggest an alternative approach to trial for evaluation.

Evaluation Group	Evaluation points	Storyboard relation to evaluation points / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
SOCIETIES platform startup	Is the user able to start up the SOCIETIES platform with relative ease? Can they perform the tasks of: • Creating their CSS and adding nodes? • What did the user have to perform to get up and running	CSS creation/installation not covered by storyboards but is assumed in COS_01Scene1 Adding nodes is not	The creation of CSS is not explicitly described in the storyboard.	TBD

		covered by storyboards Priority: TBD		
CIS interaction	Interactions with other CIS's: • Was the user able to search for other CIS's to join? • Was the user automatically joined to CIS's based on preferences/interests?	See: COS_01Scene1 COS_01Scene2 COS_02Scene1 COS_02Scene2 COS_02Scene3 COS_03Scene1 COS_03Scene3 COS_03Scene4 COS_04Scene3 COS_04Scene4 Priority: TBD	Evaluates our users concerns over system automated decision making a.	TBD
Usability of setting up platform	There will be some basic settings that the user will need to define in a configuration screen: • Was the user able to select the appropriate settings? • Was there configuration items that they felt were missing?	See: COS_01Scene1 COS_01Scene2	Similarly to CSS creation, this is not explicitly described in the storyboards. We should clarify what is meant by basic settings.	Who provides the GUI for this?
User Experience on the mobile device	Based on targeted feature rich device being Android based: • What is the experience for non-Android based devices? • What limitations were found that prevented user from participating fully in each of the user trials?	Priority: TBD Covered by all the "At the Conference" COS_03 scenes	ERIC delegates will have a mixture of hardware devices so this can be evaluated.	
Scalability of Platform	How scalable is the system when we have large numbers of users interacting with each other. Where are the perceived bottlenecks? • Is our design proving to be scalable? Or • The prototype implementation was not scalable due to minimal deployment and partial system development.	Priority: TBD Covered by all the "At the Conference" COS_03 scenes Priority: TBD	Exact evaluation "How to" needs further description	There will be ~1000 delegates at ERIC so what are the minimum trial numbers do we need for this evaluation

Portability of platform across all scenarios	We are developing a framework to satisfy 3 separate user groups: • Does the platform seamlessly interact with each user group? • Is the platform segmented with separate functions for each divide?	Not a scenario specific evaluation Priority: TBD	Exact evaluation "How to" needs further description	WP4 and WP5: How much tweaking needs to be done to the platform to support the three groups? Does this lead to hard wiring or compatibility problems? Are we loosing platform generality?
Interoperability across multiple platforms	Investigate interoperability of each of the software components with real world environments. • Performance of platform on each OS's of the mobile devices • Several components deployed on different nodes in the cloud. • Where do failures occur?	Not a scenario specific evaluation but is probably more suited to evaluation from the "At the Conference" scenes. Priority: TBD	Exact evaluation "How to" needs further description	
Sharing an "item" with feedback	Ability to publish an "item": (similar to Facebook Wall) • Publish a comment/video/file/etc to a known location • Allow users to provide feedback and comments	TBD	TBD	Not sure what is meant by this evaluation group – further explanation required. What is meant by "a known location" What is meant by "provide feedback and comments"
Service Lifecycle	Was the service container API satisfactory? • Did it allow the creation/deployment of services with easy deployment? • What limitations did 3rd party developers find when creating shared services for the community?	Priority: TBD See: COS_01Scene1 COS_02Scene1 COS_03Scene1 COS_03Scene2 COS_03Scene3 COS_03Scene4	This looks as if it is more suited to the Conference Organisation and Pre-conference Activities scenarios	Part of this evaluation has to include feedback from our 3P developers and not our user trials.
Service Registry	Users of the platform interacting with existing services: • easy to browse/search service instances within a registry of services. • Ability to search within own CSS/targeted CSS/multiple CIS's [This is as opposed to downloadable software from an appstore or marketplace]	Priority: TBD Covered by most of the Conference scenes Priority: TBD	The storyboard does not describe explicitly how services are discovered. Should we detail the storyboard? Or is it sufficient to describe the feature or search in a test scenario?	

Future Feature Requests	With users getting to interact with software, understanding their desire for future applications/technology solutions based on our presented system: • what features would users like to see? • Features that were felt to be missing?	Priority: TBD	This was already part of the feedback into both D2.1 & D8.1. More feedback will be expected from the 1st trial. T24 has been set up to capture this feedback.

5.9.3 Analysis of the evaluation points set by WP5 with feedback to WP4, 5 & 6

The following table presents a set of evaluation points defined by WP5 (quantitative in orange and qualitative in blue). For each point, we indicate how they can be related to our storyboard and to our questions. Some evaluation points require extensions of the storyboard; these extensions will be considered in the realisation of the trial prototype. Finally we also provide important feedback to WP4, 5 & 6. For instance, in some cases we describe functionality needed for performing the evaluation or we suggest an alternative approach to trial for evaluation.

Evaluation Group	Evaluation points	Storyboard relation to evaluation points / priority	Enhancement /clarification needed for trial description	Additional Feedback to WP4, 5, 6
Community/Individual Recommendations	• Do "communities to individual" recommendations improve over time during the course of the trials? I.e. does the user acceptance rate improve? • Do 'individual to community' recommendations improve over time during the course of the trials? I.e. does the acceptance rate improve? • Are recommendations	No specific scenario reference here. We need to discuss this further between WP6 & WP8 how best to evaluate this point	An important question here is whether the time constraints of the trial allow us to perform a sufficient number of tests to make all required comparisons. Time constraints may not either allow us to	Further dialogue needed with WP6 on this point

	based on data gathered from several sources better than from each of the sources alone? • What is the difference in acceptance rates for different recommendation schemes?	Priority: TBD	observe improvement over time. [MRo] agreed, this might be better managed by the Student trial.
Manual Communities	• Why do users create CISs? E.g. is it for family, interests, an ad-hoc meeting, etc? • Do user's create more long-term or ad-hoc communities? • What is the most relevant criteria for a user to join an existing community? E.g. friends already members, community purpose, etc. • How many CISs does the average user create? • How many CISs is the average user a member of?	The first evaluation bullet point maps in some ways to the Conference Organisation scenario. Priority: TBD	We do need to discuss with Intel whether they intend to use our system for Conference Organisation. This would impact possibly on the trial plan. The second bullet is hard to evaluate given the trial time constraints. The other bullets might also be difficult to measure, given the nature of the trial.
Automatic communities	• Do users see the benefit in the automatic creation of ad-hoc communities? • Do users see the benefit in the automatic creation of long-term communities? • Do users like to be automatically added to ad-hoc communities? • Do users like to be automatically added to long-term communities? • How much information is needed to create an automatic community?	Most of these bullets map well to the “At the Conference” scenario. Priority: TBD	The bullet on long-term communities might be better evaluated in the Student trial
Community based context data distribution	• How accurate is community context data? • What are the delays involved in requesting context data from a community? • What are the delays involved in requesting	These bullets also map well with the “At the Conference” scenario	

	context data from another individual? • Predefined context data distribution policies will exist, on CIS and CSS level. User will be able to further parameterize these policies. To which extend are predefined polices changed?	Priority: TBD	
Intelligent context history management	• What is the average size of necessary training data sets for efficient proactivity related functionality? • What is the overall processing and storage resources requirements of the context history on an average day of usage? Is this affecting other components?		Would these not be things that need to be analysed and answered in advance of the trial, i.e. at the design stage.
Context Inference	• What context inference mechanisms are producing the most acceptable, by the user, results? • How easy is it for the user to manage their conceptual location information?	Priority: TBD Both these bullets look as if they map quite well to the "At the Conference" scenario	How do we measure "user acceptability" of accuracy? Relevance of context inference and symbolic location inference still needs to be clarified in the scenes.
Sensors and Monitoring	• Do users accept to wear sensors (obtrusiveness)? • Do users accept to be monitored (privacy)?	Priority: TBD These two bullets might also map well to the "At the Conference" scenario	Are we giving the ERIC delegates an RFID badge? If so, then we can measure the first bullet. The second bullet can also be evaluated in terms of user feedback
Context Prediction	• What is the rate of acceptance of the predicted context by the users? How accurate are the predictions? • Is there an impact in terms of memory and processing load, or serious delay to the user?	Priority: TBD These two bullets might also map well to the "At the Conference" scenario	2nd question: This would be better answered through a pre-trial simulation than through a trial.

		Priority: TBD		
Context Inheritance	• What is the rate of acceptance of the context inherited by the users from CISs? • How accurate are the context inheritance conflict resolution mechanisms?	Priority: TBD	Does this imply that we need a critical number of CISs to be offered to the conference delegates.	This needs further discussion with WP5 people so as to try to map to an Enterprise scene, which is not immediately obvious at the minute
Context Similarity	• Can meaningful evaluations be achieved in a (near) real time solution? • Can evaluations model the dynamic CSS/CIS interactions to provide useful data sets to use as a base for further predictive analytics?	Priority: TBD	1st question: can be better analysed through simulation than through a trial.	Second bullet needs further discussion with WP5 people so as to try to map to an Enterprise scene, which is not immediately obvious at the minute
Preference Management	• Do preferences become more accurate over the course of the user trials? • Do users abort the preferences fewer times? • On average, how many preferences will users acquire over the course of the trials?	Priority: TBD	For the first bullet, are there any specific time constraints here being imposed to the Enterprise trial.	Second and third bullets need further discussion with WP5 people so as to try to map to an Enterprise scene, which is not immediately obvious at the minute
Preference Learning	• Are incremental learning algorithms able to run continuously for the duration of the trial? • How often are batch preference learning executions scheduled? • What learning algorithm performs best for preference learning?	The scenes do not explicitly describe the feature of preference learning. Priority: TBD		WP5: Check possible resource constraints. Perhaps perform simulations.
Personalisation GUI	• Do users use the Personalisation GUI to create, edit and delete preferences? • How often do the users use the Personalisation GUI to create preferences? • How often do the users use the Personalisation GUI to edit learnt preferences?	The scenes do not explicitly describe the feature of personalisation. Priority: TBD		Discussion required with WP5. Perhaps perform simulations.

User Intent	• Does the user intent model become more accurate over the course of the trials? Do users abort the intentions fewer times? • What is the confidence level at which users will allow an intent prediction to be performed automatically? • What is the average size of the intent model for all trial participants at the end of the trial? • How does the user's intent model change (grow?) during the course of the trial?	The scenes do not explicitly describe the feature of user intent. Priority: TBD	Time constraints probably for the Enterprise trial.	Discussion required with WP5. Perhaps perform simulations.
	• How much data is necessary to achieve noticeable personalisation? • How do the proposed criteria perform?	The scenes do not explicitly describe the feature of personalisation. Priority: TBD		Discussion required with WP5.
Privacy Management	• Does the user trust the privacy management of the system? If yes, what is the main feature providing this trust?	The scenes do not explicitly that the user can make use of a privacy management system. Priority: TBD	However this point relates to feedback from the paper trial; one recommendation there is to investigate support for configurable privacy rules.	TBD if the storyboard is to be extended.
Privacy Policy Negotiation	• How often does the user make any changes to the privacy policy suggested by the service? • How many times does the privacy policy negotiation fail because of the user not accepting the terms and conditions of the service?	The scenes do not cover this explicitly. Priority: TBD		TBD if the storyboard is to be extended.

Privacy Preference Manager	• Does the user use the privacy GUI to create new privacy preferences? • Does the user use the privacy GUI to change existing preferences? • Do users understand the concept of privacy preferences? If yes how much it is appreciated?	The scenes do not cover this explicitly. Priority: TBD		TBD if the storyboard is to be extended.
Data Obfuscation	• Do users understand the concept of data obfuscation? If yes how much it is appreciated? • How many times does the user modify the obfuscation level in their privacy preferences? • Do most users use data obfuscation? • What is the average level of obfuscation per data type chosen by users?	The scenes do not cover this explicitly. Priority: TBD		TBD if the storyboard is to be extended.
Active Privacy Assistant	• Would end user benefit from a tool that could warn him that he is about to share media containing sensitive information to inappropriate users or user groups?	See: COS_03Scene4 “People finder & service sharing”. Priority: TBD	This can be included in a test scenario - in relation with data sharing. The storyboard talks about data sharing, but does not explain when rules for sharing are set up and by whom.	Further discussion needed.
Visual Trust Indicators	• Would it be helpful for end user to have services fit with a clear visual estimation (e.g. a semaphore and/or progress bar) of trust for each service so he could filter the services by a trust threshold and sort services by trust level? • Would it be helpful for end user to have other users fit with a clear visual estimation (e.g. a semaphore and/or progress bar) of trust augmented for each user within SAG view?	See: COS_03Scene5 COS_03Scene6 COS_03Scene8 Priority: TBD		

User Agent	• How often do the personalisation sources suggest conflicting actions? • Which personalisation source suggests the correct action most often? • Is the User Agent able to resolve the conflict among different internal proactive components? • Do users accept automatic decision making? To what extent? • Do users accept the automatic installing/execution/provision of services? To what extent?	Questions 1, 2, 3 & 5 are not explicitly described by the scenes. Question 4 can be evaluated in COS_03Scene1 Priority: TBD		Further discussion needed.
User Feedback	• Do user feedback notification alerts become less frequent as personalisation models become more accurate? • As notifications alerts become less frequent, is the user relieved not to receive so many or upset at not them? • Does the user find notification alerts annoying and as a result ignore them?	If there are no time constraints then these points map well to the "At the Conference" scenario Priority: TBD	Are there any time constraints for this evaluation being imposed on the trial?	Need further data from WP5 on what time constraints are being imposed here
User Feedback	• Does UAV steering work with an acceptable delay for remote controllers?	N/A to Enterprise trial		

6 Conclusions

In this deliverable we have presented guidelines and expectations towards the first SOCIETIES user trials. The wide nature of the trials and the three user groups at the heart of the project have meant that we followed a combined approach of including common methodologies as well as those unique to each user group. As a basis for our work we have tried to identify the research questions that are relevant, both at a global and user-group-wide level. An example of the former is the question of how much the user would like automated community membership suggestions – and associated privacy concerns. The student scenario, on the other hand, will allow us to address questions that pertain the users' exposure over a longer time period than the other two trials.

After motivating our general methodology and the relation of the trial to the rest of the project structure and time-plan, we devote a chapter to each of the three user groups of the project. Picking up on the outputs from the paper trials in D8.1 we formulate the key research objectives from an intrinsic (i.e. "project view") point of view, the user point of view and a technical point of view, with a clear focus at this point in time on the first two. Each objective is linked to the relevant scene of Deliverable D2.2. This will allow us to see which scenes are suited to answering which questions, and allow fine tuning of the trials to best effect. Each chapter then describes temporal, physical, user and third party service aspects for that user group, including a first specification of hardware targets such as infrastructure (e.g. networking), user devices, sensors (e.g. location) and actuators (e.g. speakers). Another aspect is described, namely monitoring of the trials, such as person-on-person monitoring and user-feedback. In a first attempt at defining and anchoring formal performance tests we have taken the evaluation points specified by the two platform-level-development work packages, WP4 and WP5, prioritised them, linked them to individual tests, and furthermore related them back to all in WP4, WP5, and WP6. The result is a denser network of feedback from the envisioned tests to the work packages, in this critical stage in the build up prior to the first trials.

The final version of this deliverable (for M18) will provide more detailed quantitative tests – things that can be measured using suitable logs or tests, that is performance metrics of important components and functions, including their types (e.g. delay measurement in milli seconds) and threshold values (e.g. no less than 100 ms). Version 2 of this deliverable will also compare the three trials from a basis of their specifications, and the differences and commonalities of the research questions they will address. We shall also document, as an outlook of the deliverable, the project's plans in terms of the second trials.

