



Project Acronym: **CATALYST**  
Project Full Title: **Collective Applied Intelligence and Analytics for Social Innovation**  
Grant Agreement: **6611188**  
Project Duration: **24 months (Oct. 2013 - Sept. 2015)**

## **D2.4 Collective Intelligence Software for Social Innovation Networks: Testbed Deployments**

Deliverable Status: **Final**  
File Name: **CATALYST\_D2.4.pdf**  
Due Date: **February 2014 (M5)**  
Submission Date: **March 2014 (M6)**  
Dissemination Level: **Public**  
Task Leader: **Imagination for People**



This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

The CATALYST project consortium is composed of:

<b>SO</b>	Sigma Orionis	France
<b>I4P</b>	Imagination for People	France
<b>OU</b>	The Open University	United Kingdom
<b>UZH</b>	University of Zurich	Switzerland
<b>EN</b>	Euclid Network	United Kingdom
<b>CSCP</b>	Collaborating Centre on Sustainable Consumption and Production	Germany
<b>Purpose</b>	Purpose Europe	United Kingdom
<b>Wikitalia</b>	Wikitalia	Italy

### **Disclaimer**

*All intellectual property rights are owned by the CATALYST consortium members and are protected by the applicable laws. Except where otherwise specified, all document contents are: "© CATALYST Project - All rights reserved". Reproduction is not authorised without prior written agreement.*

*All CATALYST consortium members have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the owner of that information.*

*All CATALYST consortium members are also committed to publish accurate and up to date information and take the greatest care to do so. However, the CATALYST consortium members cannot accept liability for any inaccuracies or omissions nor do they accept liability for any direct, indirect, special, consequential or other losses or damages of any kind arising out of the use of this information.*

## Revision Control

Version	Author	Date	Status
0.1	Benoit Grégoire (I4P)	March 3, 2014	Initial Draft
0.2	Jean-Michel Cornu (I4P) Benoit Grégoire (I4P) Anna de Liddo (OU) Rosa Groezinger (CSCP) Joseph Holt (EN) Lee-Sean Huang (Purpose)	March 2014	Draft
1.0	Stéphanie Albiéro (Sigma Orionis)	March 28, 2014	Quality Check and submission to the EC

## Table of Contents

<b>Executive summary .....</b>	<b>6</b>
<b>Introduction.....</b>	<b>7</b>
<b>1. The testbeds for the first cycle of tests .....</b>	<b>8</b>
<b>1.1 Task 4.1: Social Network Analytics.....</b>	<b>8</b>
1.1.1 Tested features.....	8
1.1.1.1 Initial social network analytics.....	8
1.1.2 Testing community.....	8
1.1.3 The testbed.....	9
1.1.4 Requirements.....	9
1.1.5 Expected improvement to be measured, and testing methodologies .....	9
1.1.5.1 How these features are expected to improve the Collective Intelligence process?.....	9
1.1.5.2 How to measure the improvement? .....	9
1.1.5.3 Technological dependencies of the tests to gather data:.....	10
1.1.6 Testbed deployment plan .....	10
<b>1.2 Task 4.2: Argument Mapping &amp; Deliberation Analytics .....</b>	<b>10</b>
1.2.1 Tested features.....	10
1.2.1.1 Collective Deliberation and argument mapping .....	10
1.2.1.2 Deliberation analytics:.....	11
1.2.2 Testing community.....	11
1.2.3 The testbed.....	11
1.2.4 Requirements.....	12
1.2.5 Expected improvement to be measured, and testing methodologies .....	12
1.2.5.1 How these features are expected to improve the Collective Intelligence process?.....	12
1.2.5.2 How to measure the improvement? .....	12
1.2.5.3 Technological dependencies of the tests to gather data.....	12
1.2.6 Testbed deployment plan .....	13
<b>1.3 Task 4.3.1: Harvesting, Mapping &amp; Analysing Arguments, test 1.....</b>	<b>13</b>
1.3.1 Tested features.....	13
1.3.1.1 Email bridge.....	13
1.3.1.2 Message idea harvesting.....	14
1.3.1.3 Conversation mapping .....	14
1.3.1.4 Idea mapping to a synthesis.....	14
1.3.2 Testing community.....	14
1.3.3 The testbed.....	15
1.3.4 Requirements.....	15
1.3.5 Expected improvement to be measured, and testing methodologies .....	15
1.3.5.1 How these features are expected to improve the Collective Intelligence process?.....	15
1.3.5.2 How to measure the improvement? .....	15
1.3.5.3 Technological dependencies of the tests to gather data:.....	15
1.3.6 Testbed deployment plan .....	16
<b>1.4 Task 4.3.2: Harvesting, Mapping &amp; Analysing Arguments, test 2 .....</b>	<b>16</b>
1.4.1 Tested features.....	16
1.4.1.1 Graphical argument mapping by community managers and harvesters .....	16
1.4.1.2 Web annotation .....	17
1.4.2 Testing community.....	17
1.4.3 The testbed.....	17

1.4.4	Requirements.....	17
1.4.5	Expected improvement to be measured, and testing methodologies .....	17
1.4.5.1	How these features are expected to improve the Collective Intelligence process?.....	17
1.4.5.2	How to measure the improvement?.....	17
1.4.5.3	Technological dependencies of the tests to gather data:.....	18
1.4.6	Testbed deployment plan .....	18
<b>1.5</b>	<b>Task 4.4: Online Creativity Support .....</b>	<b>19</b>
1.5.1	Tested features.....	19
1.5.1.1	Virtual creativity card facilitation .....	19
1.5.1.2	Collocation widget.....	19
1.5.2	Testing community.....	19
1.5.3	The testbed.....	19
1.5.4	Requirements.....	19
1.5.5	Expected improvement to be measured, and testing methodologies .....	19
1.5.5.1	How these features are expected to improve the Collective Intelligence process?.....	19
1.5.5.2	How to measure the improvement?.....	20
1.5.5.3	Technological dependencies of the tests to gather data:.....	20
1.5.6	Testbed deployment plan (synchronized with 4.3.1 test).....	20
<b>1.6</b>	<b>Task 4.5: Improving Engagement &amp; Pledging .....</b>	<b>20</b>
1.6.1	Tested features.....	20
1.6.1.1	Multi-axis voting widget.....	20
1.6.1.2	Random option ordering widget.....	21
1.6.2	Testing community.....	21
1.6.3	The testbed.....	22
1.6.4	Requirements:.....	22
1.6.5	Expected improvement to be measured, and testing methodologies .....	22
1.6.5.1	How these features are expected to improve the Collective Intelligence process?.....	22
1.6.5.2	How to measure the improvement?.....	22
1.6.5.3	Technological dependencies of the tests to gather data:.....	22
1.6.6	Testbed deployment plan .....	22
<b>1.7</b>	<b>Task 4.6: Collective Intelligence Analytics Dashboard Usability Evaluation .....</b>	<b>23</b>
1.7.1	Tested Features.....	23
1.7.1.1	Collective Intelligence Analytics Dashboard.....	23
1.7.2	Testing community.....	23
1.7.3	The testbed.....	23
1.7.4	Requirements.....	24
1.7.5	Expected improvement to be measured, and testing methodologies .....	24
1.7.5.1	How these features are expected to improve the Collective Intelligence process?.....	24
1.7.5.2	How to measure the improvement?.....	24
1.7.5.3	Technological dependencies of the tests to gather data:.....	24
1.7.6	Testbed deployment plan .....	25
<b>2.</b>	<b>The second phase of tests.....</b>	<b>26</b>
	<b>List of Tables.....</b>	<b>27</b>

## Executive summary

The present document is a deliverable of the CATALYST project, funded by the European Commission's Directorate-General for Communications Networks, Content & Technology (DG CONNECT), under its 7th EU Framework Programme for Research and Technological Development (FP7).

It describes the testing environments where the consortium will test, in real world conditions (real participants working on real issues), the first technologies released as part of WP3. It outlines the main improvements anticipated from these early technologies, and the methodologies employed to determine if the expected improvements have been achieved.

## Introduction

This document describes the first cycle of tests of the technologies developed by the consortium. These tests will start at month 9 and continue (depending on the test) until month 11 to 15.

As outlined in the proposal, the tests described here consists in a series of real world testing of specific technologies by communities from the different Catalyst community partners. That is, real participants engaged in reflection around issues of importance to them in an on-going process in their community, not in a lab setting (with few exceptions).

The goal of these tests is to:

1. Measure if the feature tested helps the group or individual reach the improvement objective the feature is designed to provide (tool usefulness)
2. Measure if usability issues or missing features significantly hinder the usage of the tool (tool usability), regardless of perceived usefulness towards the objective.

One of the methodological difficulties posed by these tests is that in many cases we do not have a good baseline to compare to. It is not always possible in real world testing to split the group in two comparable and isolated subgroups (as real communities expect to always be able to communicate), and even less to create a control group that does not use a certain tool or process at all.

Improvements frequently will have to be measured through either:

- Surveys of the subjective perception of participants: users, facilitators and community managers (ex: Did this visualisation help you understand issue X better)
- Measurable fluctuations in user/facilitator/CM behaviour or activity (ex: Do people participate more actively right after a synthesis is published? What percentage of users takes the action recommended by an attention mediation notification?)
- Bringing a subset of a community in a controlled setting and measure their ability to perform predetermined tasks.
- A/B testing, when having two subgroups see different interfaces does not disrupt the overall activity of the group.

The lessons learned from these tests will allow improving the tool, both while this first cycle of test is in progress (fixing usability issues), and before the second round to tests (major changes, additional features).

They will also allow designing the testbeds for the second cycle of tests to be started at month 19

## 1. The testbeds for the first cycle of tests

The technologies tested are very early (it is anticipated that less than a third of the software development effort will have been spent by the time the first round of tests start). Despite this, at least one feature from each of the nine tasks in WP3 will be tested in real-world conditions.

Some Tasks in WP3 are however not directly measured in these testbeds, as they are enabling technologies. Namely:

- 3.1: Software architecture and cross platform interoperability specification
- 3.2: Social network interconnection layer

While they will not be the objects of specific measures, their proper operation (or lack thereof) will be tested by these real world deployments.

Specifically:

- The Analytics and visualisations are dependent on the ecosystem to get data and be displayed to users, moderators and provide objective data for some tests.
- The social network interconnection layer is used to acquire messages from outside the ecosystem, and share visualisations and messages back out.

The test report for the first phase of testing will cover any major problems encountered in those features.

### 1.1 Task 4.1: Social Network Analytics

#### 1.1.1 *Tested features*

##### 1.1.1.1 Initial social network analytics

The first test will expose users to a relatively broad variety of social network indicators, which will be visualised over a time-line so that not only the community's network characteristics are visible, but its changes over time are visible as well.

The users will be facilitators / community managers faced with the task to animate a city's online community and to animate the citizen's participation in the forums. They will use the network metrics and visualisations to get an understanding of the community dynamics and will try to derive ways to better engage the users.

This feature is built in task 3.4: Social network analytics

#### 1.1.2 *Testing community*

*Catalyst community partner(s):* Wikitalia

*Technical partner(s):* Wikitalia

*Deliberation environment(s):* Custom Drupal forum, with added technology(s) to drive the analytics collection and present a panel of metrics and visualisations to the forum administrators.

*Community:* The city of Matera, in southern Italy, has applied to be European City of Culture in 2019. As part of building its application form, it launched an online community in spring 2013 to discuss in public the emerging themes for the application. Several members of the scientific committee participated alongside with citizens, mostly from Matera itself but with a significant presence of people from elsewhere in the region and the diaspora community. The community took an interesting turn when citizens spontaneously began to use it to propose deploying small projects and coordinate them, like planting trees in a local park or survey the unfinished construction site of a railway line, which they wanted to turn into a biking and running track.

The community was highly active until late 2013, then quieted down as the Comitato Matera 2019, in charge of the process, took time out to plan its 2014 activities. Meanwhile, Matera has been short-listed for ECOC 2019. At the time of writing, the community



had 346 manually approved registered users, 191 “missions” (sort of blog posts written by users and pertaining to the city’s strategy in the ECOC 2019 bid) and 2,003 comments.

### **1.1.3 The testbed**

Wikitalia will introduce the functionality in an existing online community initiated by a mid-sized city’s municipal authority and designed for citizen participation. The focus of the test is to find out which information about community structure (as captured by network analytics) is most meaningful to community managers and individual users; and how possessing such information affects their behaviour. The technology that will be deployed will allow the technical partners to present to the community managers a set of network metrics collected from the conversations happening on the forum platform. The metrics dashboard will also present visualisations of the social network defined by the interactions between the community users, as configured by the administrators.

### **1.1.4 Requirements**

- Discussion scenarios required to test technologies (if any): None
- Required test duration (how long an actual discussion run has to be, how many cycles of some methodology, etc.): Sufficient time must be allowed for the social networks measured to reconfigure after moderator intervention. Typically a week or two.

### **1.1.5 Expected improvement to be measured, and testing methodologies**

#### **1.1.5.1 How these features are expected to improve the Collective Intelligence process?**

By deploying an easy SNA tool, we expect to gain clarity on the social structure in the Matera 2019 online community. The role of central users in connecting the community should emerge, as opposed to that of vocal users who write a lot, but communicate with a relatively small number of other users. We should be able to detect sub-communities and interpret their connectivity pattern in social terms; also, “brokering” users that connect such sub-communities to the rest of the conversation should emerge clearly. This will allow the team of animators to take steps, for example encouraging relatively isolated sub-communities to build more ties to the general conversation.

#### **1.1.5.2 How to measure the improvement?**

A very important target for this first test is to explore the user experience and usability challenges that could facilitate or make more difficult for the community managers to use the tool and get the right understanding from the data that is presented to them.

The focus of the test is thus to:

- Find out which information about community structure (as captured by network analytics) is most meaningful to community managers and individual users; and how possessing such information affects their behaviour.
- Collect feedback on the usability of the tool with respect to the ability of the users to extract information from the data presented to them.

The questions we wish to answer are:

1. **Usability testing:** The users will be asked to identify elements of the social structure that emerges from the online community, and answer some questions, e.g. who are the central users? Who are the more vocal users? Which sub-communities have been formed? This will allow to cross check their answers with the experts understanding of the network that can be obtained from the metrics. Great differences in what the novice users understand of the network would be an indication of non-clarity in the way data is presented.

2. **Impact survey:** A user test will run after deploying the tool, asking animators if and how it brought new clarity and effectiveness to their work as they perceive it, and in what sense. For example, they will be asked to name actions that they undertook as a result of studying the SNA results.
3. **Engagement:** The tool usage will be tracked through web analytics software and will count the number of times animators (and maybe general users, if it is decided to release it for general use) access the tool, how long they stay on the pages and the patterns of navigation through the tool pages. This will give us objective data to draw some conclusions on the pattern of tool usage (e.g. frequent access can be interpreted as a signal that the users found the tool useful).

#### 1.1.5.3 Technological dependencies of the tests to gather data:

- In this first phase we will install the tool ourselves, so that we must be able to operate the required changes to the forums on the hosting platform (installing Drupal modules and configuring them)
- Hosting platform must provide access to web analytics for the forum, and possibly offer a way to customize them to fit our data collection requirements.
- We must have a way to communicate with the community managers

#### 1.1.6 *Testbed deployment plan*

Below is the tentative calendar for the testing cycle:

- Step 1 (1 week)
  - June 2 to 7: Installation of the tool on the platform and Wikitalia pre-release testing.
- Step 2 (2/3 weeks)
  - June 9 to 27: Community managers' test
- Step 3 (1/2 week)
  - June 30 to July 4: Surveys
- Step 4
  - July: Surveys analysis
- Step 5
  - September: Tool improvement from user feedback
- Step 6
  - September/October: second round of testing

## 1.2 Task 4.2: Argument Mapping & Deliberation Analytics

### 1.2.1 *Tested features*

#### 1.2.1.1 Collective Deliberation and argument mapping

The overall goal of this testbed is to test whether Catalysts tools can support more informed and democratic deliberation processes through reflective online debate; and whether the Deliberation Analytics can support the moderators and participants in improving the quality of the analysis of the debate.

Specifically, the testbed will involve testing the Debate Hub, a new deliberation tool to allow large numbers of participants to collaborate online, providing mechanisms for a more structured, effective and democratic debate. To achieve this, two sets of mechanisms will be tested: the first related to the enabling of new forms of collective deliberation, the second to the mapping and distillation of deliberation outcomes.

1. **Collective deliberation:** Members of a specific community will be invited to add ideas on how to tackle civic society problems, topics or challenges. They will be also invited to comment on the proposed ideas by adding arguments for and

against each idea. Different voting mechanisms and visual statistics will also be used to allow preference making and support/challenge evaluation.

2. **Idea mapping and distillation of deliberation outcomes:** We will test specific features to lift good and supported ideas from the noise and therefore make ideas more accessible to the entire community. To do so, we will test specific features that help moderators to better organise and cluster ideas, provide visual summary and map deliberation results so that they can be better shared with the community.

These features are built in task 3.3, 3.6 and 3.9.

#### 1.2.1.2 Deliberation analytics:

The Debate Hub Dashboard will provide a subset of deliberation analytics which will consist of diverse views on issues, ideas and arguments raised by the community. Data logs from of the collective deliberation test will be used to assess to what extent and in which way the Deliberation Analytics supported moderators and participants in improving the quality of the deliberation process and its analysis.

This feature is built in task 3.5: Deliberation analytics

### 1.2.2 Testing community

*Catalyst community partner(s):* Euclid Network, Purpose

*Technical partner(s):* Open University, University of Zurich

*Deliberation environment(s):* Debate Hub

*Community:* This testbed will involve

- Primarily: Leaders/members of the European civil society sector (non-profits, NGOs, charities, etc.)
- Secondly: anybody else interested in the topic, or affiliates of the civil society sector outside of Europe.

This targeted community will be debating “Civic Society” issues.

### 1.2.3 The testbed

The targeted community members will use the Debate Hub to host the online debate. Debate Hub is different from a traditional chat forum or blog in that it can also map arguments and visualise aspects of the discussion to give users/moderators a “meta” view of what is going on in the debate and make necessary interventions, i.e. propose counter-arguments as a devil’s advocate if the debate appears one-sided.

The first testing phase will be organised in three main phases:

1. A collective deliberation phase in which participants will be invited to add, debate and vote on new ideas to solve specific civic society issues.
2. A summarisation phase in which community managers and facilitators will map and distil the main deliberation outcomes to present to the community.
3. Another collective deliberation phase in which community will be invited to debate the deliberation findings distilled by the community managers.

This last phase will be used to reflect on the effectiveness of the summary and visualisations of the debate and to assess if they improved participants awareness, reflection and understanding of the state of the debate.

The summary created in the second phase will be hand crafted by the moderators, also by interpreting and using the data analytics and visualisations created in the process and made available by the Deliberation Dashboard.

In the third phase moderators will need to clearly signal that the discussion has pivoted to talking about the findings and the more meta aspects of the debate.

#### 1.2.4 Requirements

The required test duration will be about 24 weeks. Please note that this first testing phase will overlap the summer vacation period in which a significant percentage of the users community is on holiday. This may disrupt the timing and provoke delays.

The online deliberation process will involve a pre-testing of 8 weeks (with consortium partners) and a public debate of 6 topics with end-user community members. We will host one debate per topic and each debate will last approximately 1 month (2 weeks of mixed ideation, debate and voting, and then 2 weeks of purely debating and voting)

- The topics will overlap, i.e. as the first phase of 1 topic ends, another one will begin.
- This cycle continues until the end of month 5 (approximately 20 weeks)
- The final step will be 1 month long, and requires moderators to create the summary maps and users to comment on the summaries produced for the findings of each topic.

#### 1.2.5 Expected improvement to be measured, and testing methodologies

##### 1.2.5.1 How these features are expected to improve the Collective Intelligence process?

The main goal of this testbed is to mainstream the concept of argument mapping for general Internet users participating in online debate/deliberation. In order to do this, we need to simplify the user interface and make it “intuitive” to users who already have an existing mental model about how online deliberation and debate work. We expect that this tests will provide in-use examples of low-entry User Interface for structured online deliberation and argument mapping; it will provide community managers with specific tools and features to make it easier to build summary of the deliberation process; and finally it will provide effective visualisations of the state of the debate to support community reflection, trigger new ideas and conversation and enhance collective awareness of civic society issues.

##### 1.2.5.2 How to measure the improvement?

We plan to measure the improvement of the tested features in three main ways:

1. **Targeted feedback gathering:** Moderators will be asked to complete the University of Zurich’s analytics feedback tools for the analytics used during the debate. Specific Semi Structured interviews will be planned to provide feedback on the various moderator features implemented (i.e. ideas organisation and mapping).
2. **Survey:** We will survey user testers, asking them to rate how “intuitive” the interface is at each stage of design iteration. If this general “intuitiveness” score increases, then it means that we are going in the right direction.
3. **Analytics on users’ logs:** Post-public launch, we will be monitoring usage statistics to see how users used the tool “in the wild,” and make any design modifications necessary at that time. If there are important features that are underutilised or misused (improper marking/tagging of arguments), we will redesign these to make them more accessible, or reassess their usefulness to end users (as opposed to moderators). Specifically:
  - Since the “dashboard” of debate visualisations is rather out of the ordinary for many forum users, we will want to carefully track usage statistics of how many users are viewing the visualisation dashboard and how they are using it. If users are not really using these visualisations, then we will iterate on the interface design to make it more accessible/visible.
  - It is the same for the final summaries; we will count numbers and type of comments to the human-summary of the debate to assess if and how users do engage with the findings of each topic, i.e. comment on the summaries, if they do not comment, then we need to consider promoting this feature better/removing it/making it more prominent.

##### 1.2.5.3 Technological dependencies of the tests to gather data

This testbed relies on the fact that Debate Hub is fully functioning and available no later than month 11. It also requires that the summarisation and mapping tool is available to test by month 12. It also requires that the Dashboard Analytics are sufficiently

implemented and that is the reason why the features related to testing Deliberation Analytics may be object of the final month of testing (months 14).

### 1.2.6 *Testbed deployment plan*

Below is the tentative calendar for this testing cycle:

The deployment plan of the first testing phase consists of 8 steps:

- Step 1 (8 weeks):
  - Since this will be the first time the proposed Debate Hub interface will be available to external users, it makes sense pre-public launch to conduct in-person, or remote user tests, where test users can give immediate feedback about the interface to the Catalyst partners. A pre-testing will be organised in which agile feature development and usability testing will be conducted in strict collaboration between the technical partners and the use case partners involved in the testbed
- Step 2 (2 weeks):
  - Publish Topic 1
  - Allow users to contribute ideas for Topic 1
  - Allow users to comment on the proposed ideas for Topic 1
- Step 3 (2 weeks):
  - Close the ideation stage for Topic 1
  - Continue to allow comments on the ideas for Topic 1
  - Publish Topic 2
  - Allow users to contribute ideas for Topic 2
  - Allow users to comment on the proposed ideas for Topic 2
- Step 4 (2 weeks):
  - Close commenting stage for Topic 1
  - Close the ideation stage for Topic 2
  - Continue to allow comments on the ideas for Topic 2
  - Publish Topic 3
  - Allow users to contribute ideas for Topic 3
  - Allow users to comment on the proposed ideas for Topic 3
- This continues until the end of Step 6 (at which point 6 topics will have been discussed)
- Step 7 (1 month)
  - Publish 'Summaries' of findings for each topic along with the visual data in the 'Dashboard' section of the platform.
  - Allow users to explore the summaries and the dashboard visualisations, comment on them and leave positive/negative feedback on the findings.

## 1.3 Task 4.3.1: Harvesting, Mapping & Analysing Arguments, test 1

### 1.3.1 *Tested features*

#### 1.3.1.1 Email bridge

Assembl allows bidirectional interaction with email participants, to allow reaching a wider range of participants than a pure web forum or pure mailing list.

This feature is built in task 3.2: Social network interconnection layer

#### 1.3.1.2 Message idea harvesting

Assembl has a specialised environment to speed up the workflow of a team of harvester harvesting incoming messages in a forum on an on-going basis.

This feature is built in task 3.3: Semantic Tagging, Annotation and Mapping

#### 1.3.1.3 Conversation mapping

Assembl allows mapping streams of chronological conversations to ideas (essentially topics in this case), whether or not the message thread was initially about this topic, or discussed multiple topics. This allows participants to read a subset of the conversation only relevant to the topic they are currently interested in.

This feature is built in task 3.3: Semantic Tagging, Annotation and Mapping

#### 1.3.1.4 Idea mapping to a synthesis

The synthesis allows sending an editorialised subset of the general concept map to participants every week, in order to re-start the conversation, and allow participants that are less involved to follow the discussion without reading every message.

This feature is built in task 3.6: Semantic map edition and visualization

### 1.3.2 **Testing community**

*Catalyst community partner(s):* Imagination for People

*Catalyst technical partner(s):* Imagination for People

*Deliberation environment(s):* Assembl

*Community 1:* Anim-fr is a group of 340+ community managers, created in 2011. The aim of this group is to empower projects, groups and social networks facilitators by discussing on existing and new methods for cooperation. Anim-fr has already tested (by handcraft) the harvesting and synthesis process tested here, produced a collective text on "Cooperation explained to my redneck brother in law" (arguments on cooperation for those who don't believe in it). The group will work on a new discussion focused on community management. While the group is an international French speaking community, some of them will meet in Brest (France) on July 2 to 4, during the "cooperative usage forum" (6th edition) where a Catalyst workshop is organized to discuss the results of the first iteration of the testbed.

*Community 2:* Fablab-fr is a group of 300+ fablab members and managers, other digital fabrication places and individual hackers created in 2012. This active group (more than 30 mails a week on the mailing list) holds many diverse discussions on technical issues, economic models and political recognition. The fablab-fr group will use Catalyst tools to test capitalisation of multiple discussions on collaborative watch, sensemaking and on various topics discussed in the group. At this stage, the aim will not be to test complex debates but rather to keep track of ideas and arguments in numerous debates and discussion. This is a major issue for many groups which hold discussions, generate important info and ideas, and pass to the next discussion forgetting all that was said before. This may also lead to recruiting several groups and communities (30 groups already helped by Imagination for People are active or in the bootstrapping phase) to use Catalyst tools and for some of them to switch to more complex debates.

Both communities will test multiple iterations of — reactions/harvesting/mapping/summary sent to the group. Previous experience from I4P experimenting this method (that has been carried out without tools' support to date) showed that with 3 or more iterations, the group goes beyond the sum of individual intelligence. Individual contributions are not only based on pre-existing ideas but are also built on the global view created from the ideas of all the group (this approach need to ask contributors to go further and give additional ideas and arguments not already in the map)

### 1.3.3 *The testbed*

The communities participating are all pre-existing groups of participants having discussions on mailing lists around specific topics (sometimes for years). As such, they have a pre-existing discussion culture, and already know each other. Also, they have experienced community facilitators that will be tapped as expert harvesters.

### 1.3.4 *Requirements*

- Discussion scenarios required to test technology (if any): Any group of people discussing ideas through online messaging.
- Required test duration (how long an actual discussion run has to be, how many cycles of some methodology, etc.): Between 3 and 7 weekly synthesis iterations depending of the participation. Two such cycles will be run for each community.

### 1.3.5 *Expected improvement to be measured, and testing methodologies*

#### 1.3.5.1 How these features are expected to improve the Collective Intelligence process?

- It is expected that the email bridge will allow participants to interact with each other in a single discussion (though the mailing list, an online forum and eventually through Facebook or other social networks)
- It is anticipated that the message idea harvesting tool and synthesis tool will allow to save time for the respective tasks (typically 5 hours for a one-week cycle without tools)
- It is expected that the publication of synthesis will foster activity in the discussion.
- It is expected that conversation mapping will help end user find an entry point to the discussion, and further speed up ongoing harvesting.

#### 1.3.5.2 How to measure the improvement?

1. Compare the impact of the synthesis published (number of contributions in the attached threads after the idea is generated, number of distinct ideas generated from those threads, number of distinct participants who interacted with the idea and attached message threads) compared to the baseline activity using time series. Collect distinct statistics for those who reacted directly from the synthesis, and those who reacted transitively (responded to a thread that started on the synthesis)
2. Ask harvesters to log all the time spent harvesting and writing a synthesis for each test iteration.
3. Survey the participants twice (once after each cycle of 3 to 7 synthesis iterations) to ask if:
  1. They used the table of ideas. If so, did it allow them to find a conversation they wanted to read message in. (Tool usefulness)
  2. They find the synthesis useful to their understanding of the discussion.
4. Survey the harvesters twice (once after each cycle of 3 to 7 synthesis iterations) to ask if they feel that:
  1. Did the table of ideas and catcher's view allow them to find messages to harvest quickly?
  2. Regardless of the perceived usefulness, did they feel it was easy to perform the harvesting and write the synthesis? (Tool usability)
    - Ask to describe the top blockers (irritants that, if fixed, would allow them to use the tool better/faster)
  3. Ask them to rate (subjectively) how much faster they think harvesting from messages and writing a synthesis is compared to a mail client and a word processor. (Tool usefulness)
5. Calculate from logs if participants actively participated from the mailing list, the web, and both.
6. Calculate from logs how often participants interact directly with the synthesis.

#### 1.3.5.3 Technological dependencies of the tests to gather data:

- Logging user and moderator activity (as described in the interoperability specification) must be available in Assembl.

- Synthesis impact analytics must be available

### 1.3.6 *Testbed deployment plan*

Below is the tentative calendar for this testing cycle. (These first testbeds will overlap the vacation period, which may disrupt the timing.)

Two cycle of iterative test with Anim-Fr

- March 27 - Presentation of the testbed to Anim-fr
- May 9 - Environment pre-testing by community manager (1 week).
- May 12 - Train harvesters on the tool (2 weeks)
- May 26 - Announce the start of the process to participants
- June 2 - 3 synthesis iterations (3 weeks, including creativity test)
- June 23 - Week dedicated to creativity on one branch of the debate (task tasks 4.4) and voting on usefulness and usability of the tools (tasks 4.5)
- June 30 - Send initial survey to participants and harvesters and organise a meeting for debriefing during the Catalyst/Anim-Fr workshop during the forum in Brest (1 week)
- Adapt tools (during summer when a significant percentage of the group is on vacation)
- Announce changes
- September 8 - Second test with a longer iterative process (up to 8 weeks including a creativity week one time between 2 synthesis cycles and a voting/engagement process at the end)
- (No later than Nov 1st) - Surveys to participants and harvesters (1 week)
- Compile and analyse data from the tests (3 weeks)
- Write report (2 weeks)

Several iteration capitalisation of discussion with Fablab-fr

- September 1 - Train harvesters on the tools (2 weeks) and announce the start of the process to the group
- September 15 - One month of weekly synthesis iterations to capitalise multiple discussions on the list / web and if possible on social networks
- October 13 - End initial survey to participants and harvesters (1 week)
- October 20 - Compile and analyse data from the tests (3 weeks)
- November 10 - Write report (2 weeks)

It is expected that the fablab-fr group will continue capitalising discussions with the tools provided after the end of the formal test. This group will probably test at this stage only harvesting, mapping and analysing arguments but not online creativity tools (task 4.4) and improving engagement & pledging (task 4.5). However, if a demand occurs, these tools may be provided to the group and the report will include some feedback.

## 1.4 Task 4.3.2: Harvesting, Mapping & Analysing Arguments, test 2

### 1.4.1 *Tested features*

In this testbed we will test specific web annotation and argument mapping features provided by Catalysts tools.

#### 1.4.1.1 Graphical argument mapping by community managers and harvesters

We will test specific Catalyst tools to allow community managers and harvesters to build argument maps on top of specific discussion forum, in order to collaborate directly to build a graphical representation of an IBIS map. A group of facilitators and



harvesters will engage into directly adding IBIS nodes, and engage on other people's nodes, in order to build a common shared understanding of the issues being tackled by a specific community discussion forum.

This feature is built in task 3.6: Semantic map edition and visualization

#### 1.4.1.2 Web annotation

Specific Web annotation features will be tested that allow harvesters to highlight resources on different websites (inside and outside the one hosting the community debate), and tie these resources to the collectively edited argument map.

This feature is built in task 3.3 Semantic Tagging, Annotation and Mapping

### 1.4.2 *Testing community*

*Catalyst community partner(s):* CSCP

*Catalyst technical partner(s):* Open University

*Deliberation environment(s):* Litemap

*Community:* In this testbed, we will engage with the Utopia Community. Utopia is a German website and online community on sustainable consumption with over 60 000 members. The Utopia community consists of German speaking users, with a private interest or professional background in the topics of sustainable lifestyles and consumption. In different discussion forums of the website, chats and Facebook feeds, the users are discussing topics based on current societal developments, but also concrete consumption choices and products.

#### 1.4.3 *The testbed*

The community debate will be hosted on a web forum called the Utopia platform. As described above, different discussions on sustainable consumption and lifestyles are taking place in the various discussion forums of the community on the Utopia website. Many of the discussions tackle similar topics. These will be used by the harvesters to connect different discussion streams and additional resources from the web to show the connection and interaction between the different discussions that are ongoing.

For the duration of the test, the main menu of the site will point to the Argument maps, created on the basis of the different discussions and brought together by the harvesters. The small group of harvesters will be predefined. If engaged users over the course of the testing show interest in becoming harvesters, they can also become part of that group.

#### 1.4.4 *Requirements*

The community has to be exposed repeatedly to the generated map. A moderator post will regularly prompt them to do so.

### 1.4.5 *Expected improvement to be measured, and testing methodologies*

#### 1.4.5.1 How these features are expected to improve the Collective Intelligence process?

The goal of the collectively edited argument maps is to create better shared understanding (a richer mental model) of the issues tackled by the participants, compared to the understanding that would have been generated without it. Argument maps also aim at better summarising and communicating the state and outcomes of a collective debate; and can therefore be used as tool for collective reflection and awareness. Argument maps can be also used to connect issues, ideas and resources that sit into different debate threads, thus improving issues of platform islands and balkanisation.

#### 1.4.5.2 How to measure the improvement?

The testbed will not rely into the technical interoperability between the Utopia platform and the Catalyst tools (that will be loosely couples by using a Web annotation paradigm). This clearly affects the tests that can be performed.

1. Survey the community moderator(s) with the following questions:
  - Did they find it easy to harvest content from the Forum data? (Web annotation feature)
  - Did they find it easy to create and edit argument maps? (Tool usability)?
  - Did they feel that the use of the argument map allowed them to communicate better to the forum participants the main ideas and arguments raised in the debate, compared to more standard linear summaries? (Tool usefulness)
2. Survey the forum participants with the following questions:
  - Did they feel that the view and exploration of the argument maps helped them (individually) understanding the issue better than a chronological text discussion? (Tool usefulness)
  - Did they feel the argument map exploration helped them (collectively) generating more/better contributions? (Tool usefulness)
  - Regardless of the perceived usefulness from map creators, did they feel it was easy to navigate the map? (Tool usability)
  - Did they navigate to any of the harvested web content? (Tool usability)
    - If so, did they find it useful? (Tool usefulness)
3. Survey 3 groups: a group of harvesters who edited the argument map, a group of participants who viewed the map, a group of people who did not participate in the process in any way (forum or harvesting), but have similar interests. Ask them to:
  - Rate how understandable the map is to them (Tool usability, usefulness), and why (or why not).
  - Rate if the map represents a reasonably complete overview of the central issue. (Tool usefulness)
  - Calculate how frequently, and how many different users visit the map.

#### 1.4.5.3 Technological dependencies of the tests to gather data:

The hosting platforms must provide pseudonymised usage statistics for the forum.

#### 1.4.6 *Testbed deployment plan*

Below is the tentative calendar for this testing cycle:

- Environment pre-testing by community manager/harvesters (2 weeks) Community managers/harvesters familiarise themselves with the ICT tool and the goals of the testing phase
- Train community manager how to use, and create an initial map to avoid the “blank page” syndrome (1 week): An initial map will be created in this week to have some demonstration material available once the map is displayed to the community
- Announce the tool to the participants. The community not being current users of the tool, this process will take a while. (3 weeks): The tool and the first map will be advertised and presented to the community
- Run the process (6 weeks): New discussions will be fed into the tool and be shown in the map to the community
- Send survey to those who participated (1 week): To gather feedback on the practicability and usefulness of the tool, a survey will be send to the participants
- Send survey to community managers/harvesters (1 week): To gather feedback on the practicability and usefulness of the tool, a survey will be send to the community managers/harvesters
- Gather a group of outside evaluators, and send comparative survey (2 weeks): To gather feedback on the practicability and usefulness of the tool, a survey will be send to outside evaluators
- Compile and analyse data from the test (3 weeks): Based on the feedback received in the survey and the data showing the visitors to the argument map, and analysis of the practicability and usefulness of the tool and the application for this community will be performed
- Write report (2 weeks)

## 1.5 Task 4.4: Online Creativity Support

### 1.5.1 *Tested features*

#### 1.5.1.1 Virtual creativity card facilitation

The card facilitation widget is an online adaptation of a common offline technique consisting of presenting participants with a certain number of cards selected from a set, and ask them to have either individual or group interactions with the card, and present their findings to the group. Online asynchronous interactions differ from in-presence synchronous ones in three major ways:

1. There is no facilitator that actively guides the participants through the process.
2. The participants cannot realistically form groups and perform the activity at the same time, they do so individually over longer time period (typically 48 to 72 hours instead of 30 minutes)
3. The number of participants is typically much larger

The widget will replace the facilitator to guide the participants in the activity. The results of the activity will either be shared directly through the widget (for group activities) or as ideas or posts (for individual activities).

This feature is built in task 3.7: Creative ideation

#### 1.5.1.2 Collocation widget

The collocation widget is designed to foster creative ideation by presenting users with a list of videos related to the ideas expressed in the current list of ideas considered by the group, in order to help them find inspiration.

This feature is built in task 3.7: Creative ideation

### 1.5.2 *Testing community*

*Catalyst community partner(s):* Imagination for People

*Catalyst technical partner(s):* Imagination for People

*Deliberation environment(s):* Assembl

The communities, and the discussions are selected from those testing task 4.3, as we need a baseline activity to compare to.

We may also test on other communities using Assembl in an ad-hoc manner, as for some measurements it is not necessary to run the process more than once.

### 1.5.3 *The testbed*

Same as task 4.3

### 1.5.4 *Requirements*

- Discussion scenarios required to test technology (if any): Any situation where the group or moderator considers that the process is generating an insufficient number of options in a “branch” of a debate.

### 1.5.5 *Expected improvement to be measured, and testing methodologies*

#### 1.5.5.1 How these features are expected to improve the Collective Intelligence process?

The goal of the facilitation widget is to have participants create a richer set of options on some topics than would have been generated without it.

#### 1.5.5.2 How to measure the improvement?

1. Present a different set of cards to two subgroups, and compare the results. This allows objectively measuring the effectiveness of a certain card set, or guidance instructions compared to each other.
2. Survey the participants to ask if they feel that:
  - The intervention helped them (individually) generate more/better contributions. (Tool usefulness)
  - The intervention was/wasn't disruptive to the normal collaboration flow. (Tool usefulness)
  - Regardless of the perceived usefulness, did they feel it was easy to perform the interaction? (Tool usability)
3. Survey the community moderator(s) to ask if they feel that:
  - The intervention helped them (collectively) generate more/better contributions. (Tool usefulness)
4. Replace a synthesis with a card facilitation intervention published on the same channels, and measure the relative impact on the conversation compared to the synthesis.
5. Calculate the percentage of additional ideas generated by participants during the period when the widget is used compared to the number of ideas pre-existing at the time the discussion was considered "stuck". Also compare to the number of ideas generated by participants who chose not to use the widget, though selection bias is a factor there.

#### 1.5.5.3 Technological dependencies of the tests to gather data:

- Card widget must be capable of A/B testing (randomly split users into subgroups and present two different interfaces to them)
- Synthesis impact analytics must be available

#### **1.5.6 Testbed deployment plan (synchronized with 4.3.1 test)**

Below is the tentative calendar for this testing cycle:

- March 27 - Presentation of the testbed to Anim-fr
- May 26 – Announcement of the intervention to participants (together with task 4.3 and 4.5)
- June 2 - Environment pre-testing by community manager (1 week).
- May 9 - Card deck adaptation (2 weeks)
- June 23 - Creativity week during the Anim-fr debate
- June 30 - Send survey to those who participated and Catalyst/Anim-fr workshop (1 week)
- Adapt tools (several weeks)
- Some time between September 22 and October 13 - 2nd creativity week (1 week)
- No later than November 1 - Surveys to participants and harvesters (1 week)
- Compile and analyse data from the tests (3 weeks)
- Write report (2 weeks)

## **1.6 Task 4.5: Improving Engagement & Pledging**

### **1.6.1 Tested features**

#### 1.6.1.1 Multi-axis voting widget

Some questions are related but distinct. A good example from this very document are the two questions:

- "Is this functionality useful to you"
- "Did you find that this functionality easy to use".

Only asking the first question will make the results very difficult to interpret: If someone says that the functionality is not useful, it may be because he did not manage to get it to work, not because it is not useful to him.

One can (and typically does) go one step further and ask the two questions successively. We can now interpret the results more easily, but:

- The voter now has two questions to answer
- It is unlikely he will think of the second question when answering the first. At best he will backtrack

Several sets of questions are more useful when asked together, such as: Is this argument true vs. is this argument relevant? Is this document clear vs. is this document complete?

A classic one actually gave us the idea for the interface: is this risk likely to occur vs. is the potential impact major?

The classic risk matrix is used for reporting the results:

Risk Matrix		← IMPACT →				
		1 Insignificant	2 Minor	3 Moderate	4 Major	5 Catastrophic
↑ LIKELIHOOD ↓	5 Almost Certain	5	10	15	20	25
	4 Likely	4	8	12	16	20
	3 Possible	3	6	9	12	15
	2 Unlikely	2	4	6	8	10
	1 Rare	1	2	3	4	5

Table 1: Risk matrix

It seems the solution is to reverse the interface, put the two questions on a grid or a two axis chart and ask the user to place a pin to tell how useful, and how easy to use the functionality is at the same time (equivalent to choosing a cell in the example above)

1. It is faster (One single click to vote on both questions)
2. It actually forces the user to think about both questions together, which for related questions such as this matters: the voter will hopefully try to determine if he does not like the feature because he finds it too slow, or difficult to use, or if he simply finds it superfluous.
3. It is more visual.

This feature is built in task 3.8: Pledging and voting

#### 1.6.1.2 Random option ordering widget

It has been shown that simple voting for options tends to lock-in early options and introduces a bias. Biases also occur when the list of options presented to each voter is in the same order. Worse biases occur when the list is sorted by the number of votes cast so far. If the number of options is large enough that users will not read through all the options, the results can be especially pathological.

However, presenting a list in a different order for every participant can easily be done in a web context, but is rarely done in an online discussion for lack of adequate tooling.

### 1.6.2 Testing community

*Catalyst community partner(s):* Imagination for People

*Catalyst technical partner(s):* Imagination for People

*Deliberation environment(s):* Assembl

*Community:* As Ashoka could not be part of the consortium, the communities from task 4.3 and possibly Ad-hoc focus groups will be asked test these features, in order to make progress on the important precondition to collective action that is collective decision.

Hopefully, additional testers might be recruited from the open call to more focus on collective action for the second cycle of tests.

### **1.6.3 The testbed**

Same as task 4.3. The option ordering widget will be used when the group wants to choose projects to focus on. The multi-axis voting widget will be used for participants to rate the usefulness of the synthesis as a whole, using various sets of questions the synthesiser will find useful (Is the synthesis accurate vs. concise? I read the messages vs. The synthesis is clear, etc.)

Anim-fr group will have two sets of multi-axis vote:

1. During the first test (June) on usefulness/usability of Assembl process, and online creativity tools
2. At the end of the second test on usefulness/usability of the full process

A tentative random option ordering vote may be added to let participants vote on which actions they are prepared to be involved (from those identified during the discussion)

### **1.6.4 Requirements:**

- Discussion scenarios required to test technologies (if any): The community must feel that a decision point has been reached that must be resolved by a vote, or that feedback on options can be obtained by a vote.
- Required test duration: The tests are very short (days at most), but are best repeated over time to observe their influence on the discussion.

### **1.6.5 Expected improvement to be measured, and testing methodologies**

1.6.5.1 How these features are expected to improve the Collective Intelligence process?

The main benefits anticipated is to make more sophisticated means to vote available to communities, to provide quicker voting and avoid common biases in order to support the Collective Intelligence process.

1.6.5.2 How to measure the improvement?

1. Survey the participants to ask if they feel that:
  - Voting on two axis was easy to understand. (Tool usability)
  - Voting on two axis helped them think of the relations between the two questions (Tool usefulness)
  - Ordering options to vote was quick and clear. (Tool usability)
  - Did they find the fact that vote results were not available until voting was over annoying? If so, why? (Tool usefulness)
2. Survey the community moderator(s) to ask if they feel that:
  - The votes helped them (collectively) improve the discussion or their own work. (Tool usefulness)

1.6.5.3 Technological dependencies of the tests to gather data:

Hosting platform must be able to identify distinct users.

### **1.6.6 Testbed deployment plan**

Below is the tentative calendar for this testing cycle:

- June 23 - First test of multi-axis vote plan during the first test for Anim-fr (creativity and voting week see task 4.3.1 and 4.4) a second one will be done at the end of test 2
- After September 8 - First test of random option ordering widget during the second test of Anim-fr or first test of Fablab-fr

## 1.7 Task 4.6: Collective Intelligence Analytics Dashboard Usability Evaluation

### 1.7.1 Tested Features

#### 1.7.1.1 Collective Intelligence Analytics Dashboard

The dashboard aims to make multiple visualisations of a deliberation available to moderators or end users in one place to allow better access and interpretations of the visualisations.

In task 3.9 the OU carries on the design and implementation of several visualisations of a specific subset of Collective Intelligence analytics (See T2.2 for the full list of possible Collective Intelligence analytics identified by the project; between these just a small subset will be implemented in the Dashboard)

### 1.7.2 Testing community

*Catalyst community partner(s):* All Partners

*Catalyst technical partner(s):* Open University, Imagination for People

*Deliberation environment(s):* Debate Hub, Assembl, Litemap

*Community:* Testers will be a subset of community members and community managers from the testbed communities involved in Tasks 4.1 to 4.5; there is also a scope for expert recruiting by using platforms for crowdsourcing testing such as <http://www.crowdfunder.com/>

### 1.7.3 The testbed

The testing will consist of three main phases:

#### 1) Testing Dashboard Visualisations

This testing will be carried out in a lab-setting. The testing tasks will look at assessing the usability and usefulness of the diverse analytics visualisations which the Collective Intelligence Dashboard consists of. A small number of users will be screen and audio recorded while performing realistic tasks with the Dashboards. A post-hoc analysis of the videos and short semi-structured interview will provide rich qualitative feedback on the Dashboard user experience, and usefulness of the various visualisations.

#### 2) Testing Dashboard Use

This testing will be carried out in real world conditions, but does not involve specific user participations. As many of the visualisations from the Dashboard as possible are made available to the other testbeds (described 4.1, 4.2, 4.3.1 and 4.3.2) and quantitative analysis of the usage will be performed on users logs gathered in the other testbeds of the Catalyst tools.

#### 3) Testing Dashboard Usability and Usefulness

The third testing will involve surveying a larger group of users who have been involved in the testing of the different ecosystem of tools that include all or part of the Dashboard visualis Below is the tentative calendar for this testing cycle:

June 23 - First test of multi-axis vote plan during the first test for Anim-fr (creativity and voting week see task 4.3.1 and 4.4) a second one will be done at the end of test 2

After September 8 - First test of random option ordering widget during the second test of Anim-fr or first test of Fablab-fr ations (e.g the Debate Hub, Assembl, Litemap, etc.)

#### **1.7.4 Requirements**

This testing will take place in the latest months of the testing for three main reasons:

1. It requires the full implementation of Dashboard Visualisations in the different platforms,
2. It requires integration of the Dashboard Visualisations with the Collective Intelligence analytics tools
3. It requires users logs from the other testbeds and therefore it needs that the users engagement and participation process is already advanced.

#### **1.7.5 Expected improvement to be measured, and testing methodologies**

##### 1.7.5.1 How these features are expected to improve the Collective Intelligence process?

The Collective Intelligence Dashboard will make analytics visible, usable and interpretable by real users and community managers. By presenting multiple visualisations together, the Dashboard will enable easy browsing and selection of visualisations that can serve different user needs.

The Dashboard will improve the Collective Intelligence process mainly by providing context and visualisations to the Collective Intelligence analytics. This will be done by providing users with holistic views of the state and outcomes of the online debate and will eventually help improving users' awareness, reflection and understanding of the deliberation process.

##### 1.7.5.2 How to measure the improvement?

There are different improvements the partners aim to measure:

1. Measuring user experience with the Dashboard. (Do people understand the dashboard? Can they effectively use the visualisations?)
  - Each Dashboard visualisation will be the object of targeted user experience trials in which users will be video recorded while completing realistic tasks. This will provide rich, qualitative, process data of users' interactions. Some users will participate in the OU's usability lab, while others will be recruited from partners and recorded remotely, using the dashboard to complete a range of meaningful tasks in their own context. This will include:
    - Guided tasks
    - Unguided tasks
    - Semi-structured Interviews
2. Engagement (Do participants consult the dashboard? How often? How much?)
  - User's logs gathered from community-partner testbeds will be used to assess which visualisations are more used, how often and by whom. This quantitative analysis will provide insights into the engagement level and use of the Dashboard visualisations.
    - Count first and subsequent visits
    - Auditing viewing activity for each dashboard visualisation
    - Count actions performed per visit and per dashboard visualisation
3. Dashboard Usefulness (Do participants take action using dashboard tools?) Small usability trials will be organised, in order to assess the user experience when working with multiple visualizations and analytics.
  - Surveys and Usability questionnaires will be used to gather feedback on Usability.
  - Open coding of the interaction videos by activity type

##### 1.7.5.3 Technological dependencies of the tests to gather data:

This task depends on:

- The implementation of Dashboard Visualisations in the different platforms,



- It requires that the tools that include Visualisations are appropriately developed and in-use so that users logs and deliberation data are ready for the visualisation to load,
- It requires integration of the Dashboard Visualisations with the Collective Intelligence analytics tools.

#### **1.7.6 Testbed deployment plan**

Below is the tentative calendar for this testing cycle:

- Testing Dashboard Visualisations (Duration 12 weeks - Month 9 -11)
- Testing Dashboard use (Duration 12 Weeks - Month 11-13)
- Testing Dashboard Usability (Duration 12 Weeks - Month 12-14)

## 2. The second phase of tests

As outlined at length in the original proposal, the results of the first phase of tests will inform software development, and thus features tested in the second phase.

This makes it impossible to describe the next cycle of tests except in very broad strokes at this stage. The test for the second cycle will be similar in general nature to the first. However, the context will be different:

- As community partners develop a stronger culture of online discussion, and get more familiar with the potential of those technologies, we expect that they will have much more specific request of software.
- Unlike this first round of tests, where a fraction of expected software development has been done, the bulk of software development will be complete by the time the second round start. The features tested will be more refined, and many additional ones will be made available.
- The growing number of metrics, visualisations and representations of the concept maps should allow at least some comparative testing to be done. That is, within a single community subgroup A uses tool X, subgroup B uses tool Y. Both subgroups are assigned the same task, and the results are compared.

## List of Tables

Table 1: Risk matrix ..... 21