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)**

D4.1.0 Evaluation of CI Software: Work Status

Deliverable Status:	**Final**
File Name:	**CATALYST_ D4.1.0.pdf**
Due Date:	**September 2014 (M12)**
Submission Date:	**October 2014 (M13)**
Dissemination Level:	**Public**
Task Leader:	**Euclid Network**

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:

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

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

Revision Control

Version	Author	Date	Status
0.1	Alexandra Yaghil (Euclid Network)	August 2014	Initial Draft
0.2	Alexandra Yaghil (Euclid Network)	September 2014	Second Draft
0.3	Jean-Michel Cornu (I4P) Rosa Strube (CSCP) Lee-Sean Huang (Purpose), Alberto Cottica (Wikitalia), Anna de Liddo (OU)	September 2014	Inputs from partners
0.4	Alexandra Yaghil (Euclid Network)	September 18, 2014	Third Draft
0.5	Stéphanie Albiéro (Sigma Orionis)	September 22, 2014	Quality check
0.6	Rosa Strube (CSCP), Anna de Liddo (OU) and Lee-Sean Huang (Purpose)	September 24, 2014	Sent additional inputs
0.7	Benoit Grégoire (I4P)	September 29, 2014	Sent additional inputs
0.8	Alexandra Yaghil (Euclid Network)	September 29, 2014	Fourth draft
0.9	Anna de Liddo (OU), Frank Escoubes (I4P)	October 20, 2014	Peer-review
1.0	Marta Arniani (Sigma Orionis)	October 21, 2014	Final review and submission to the EC

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

Table of Contents

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

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).

In June 2014, CATALYST invited the project's community partners to introduce a series of specific testing tools in real-world settings. The goal was to test each specific technology developed by the different parties over a period of three months (until month 12 - September 2014). These tests consist in involving genuine participants in conversations and debates on pertinent topics and issues to foster discussion and allow each tool to be tested on an on-going basis.

CATALYST is committed to producing tools that are innovative, effective and user-friendly. Therefore, the testbeds' evaluation, which this document reports on, focused on two key aspects: the usability and the usefulness of the tools. In other words: is the tool convenient for use? And is the tool helpful and in correspondence with its initial objectives? The evaluation process has consisted in asking all partners to fill in a preliminary questionnaire in order to gather feedback on the strengths and weaknesses of the features they tested.

Thanks to a careful and efficient follow up of technical discrepancies and users' comments on the tools, all partners involved have been able to identify pain points and bugs throughout the running of the tests. Moderators and users have indeed raised numerous points for consideration, e.g. facilitate the navigation, modernise the design, increase the font's size, an overly time-consuming harvesting task. Partners have also gained insights into the necessary development to address missing features that are compromising the usage of the tools either by the participants to the discussions or to the community managers. Although the feedback received was specific to each tool, all partners will reflect on the commentary in order to ensure that their tools meet the needs of the targeted community.

Each partner analysis, based on both quantitative (percentage and/or number of responses, participation, views, etc.) and qualitative results (answers to surveys, comments sent, etc.), has provided all parties with an opportunity to proceed with all the improvements and modifications necessary to come up with more efficient and user-friendly tools for the on-gong testing process and the second phase of testing to be initiated at month 19.

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

Introduction

Based on the previous deliverable D2.4 (Collective intelligence software for social Innovation Networks: Testbeds deployments), CATALYST launched at month 9 (June 2014) a series of real world tests of specific technologies for communities from within the project's community partners. In order to engage users and support dialogue on societal issues, the tests involved real participants and focussed on topics and issues of importance to them and their communities. As such, the majority tests were conducted in real communities as oppose to lab settings.

Over a three-month period, these initial tests sought to analyse each tool's usefulness in terms of online debates facilitation and usability. Through this testing, CATALYST partners looked to understand whether the features of each tool tested helped the group and/or individual to attain the improvement objective specifically targeted by the tools' features. In addition, the tests analysed potential usability issues or missing features to assess whether they significantly hinder the usage of each tool by the participants and/or community managers and moderators.

This report summarises the status of the work that has been done so far in WP 4 - Evaluation of Collective Intelligence Software - by providing an evaluation of the testbeds that have been launched in a first cycle of tests. In order to provide CATALYST with a consistent summary of the work done and draw on lessons learned, each partner has been invited to complete a preliminary report whose results are presented within this document. After each test phase, the evaluation carried out aims to identify:

- Lessons learned at a strategic level
- Key functionalities formally validated
- Key functionalities formally invalidated
- Key functionalities to be fine-tuned
- New testing avenues for further cycle(s)
- Key success factors
- Major implementation obstacles
- Usability issues
- Risk mitigation strategies

Despite the early stage development of the technologies tested, this first evaluation phase and the conclusions drawn will allow each partner to improve the tools, both for the first cycle of test in progress and before the second round of tests. They will also help in the designing of the testbeds for the second cycle of tests to be started at month 19. Finally these tests have been useful to refine use-cases for the different CATALYST tools.

It is to note that among the seven tests programmed in this first phase, four have not yet been launched. Testbeds number 2 - "Argument Mapping & Deliberation analytics" - will be launched in the middle of October 2014. The testbeds numbers 4 and 5 - "Online Creativity Support" and "Improving Engagement and Pledging" - are currently being launched (September 2014), therefore no feedback is yet available at this point. The testbed number 6 – Collective Intelligence Analytics Dashboard Usability Evaluation – will be launched in October 2014 as testers will be a subset of community members and community managers from testbeds communities involved in all other testbeds.

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

1. Listing of testbeds

TESTBED	Launching date	Community Partner	Technical Partner	Deliberation Environment	Tested feature	Community
1	Edgesense dashboard: May 23rd 2014 Tutorial & Survey: July 16th 2014	Wikitalia	Wikitalia	Custom Drupal forum	Initial Social network analytics	Facilitators / Community managers
2	*Will be launched mid-October 2014*	Purpose	Open University	Debate Hub	Argument Mapping & Deliberation Analytics	Moderators / Participants
3.1	June 2nd 2014	Imagination for People	Imagination for People	Assembl	Email bridge, Message idea harvesting, Conversation mapping, idea mapping to a synthesis	Pre-existing discussion groups' participants: Anim-fr & Fablab-fr
3.2	July 3rd 2014 (until August 22nd)	CSCP	Open University	LiteMap	Graphical argument mapping, Web annotations	Utopia Community (German website and online community)
4	*Launched in September 2014*	Imagination for People	Imagination for People	Assembl	Virtual creativity card facilitation, Collection widget	Communities and discussions from testing 3
5	*Launched in September 2014*	Imagination for People	Imagination for People	Assembl	Multi-axis voting widget, Random option ordering widget	Communities and discussions from testing 3
6	*Will be launched in October*	All partners	Open University, Imagination for People	Debate Hub, Assembl, Litemap	Collective Intelligence Analytics Dashboard	A subset of community members and managers from the testbeds communities involved in testbed 1 and 5

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

2. Assessment of the first testing phase

All community partners have filled in a report with an evaluation of the work that has been done in each test-bed, and drawn on the main lessons learned and foreseen risks. The report on each tesbed is detailed below.

2.1 Testbed 1: Social network analytics

2.1.1 The testbed

Edgesense is a dashboard which enriches Drupal forum and community sites with social network analytics. By augmenting online conversations with network analytics, the goal is to foster collective intelligence processes. The vision behind is to contribute to building a format for participatory democracy that works at the global scale.

Wikitalia introduced the functionality in an existing online community initiated by Matera, an Italian mid-sized city's municipal authority. The Matera's online community was 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.

2.1.2 Description of the testbed implementation from testing and technical partners

The Edgesense dashboard went live on May 23rd, 2014 while the interactive tutorial, which includes the survey for the users to complete, was activated on July 16th.On July 20th, Wikitalia held a workshop in Matera, to demonstrate the use of Edgesense. It was mostly targeted to moderators, but also open to the public. There were 17 registrations, of which about 12 turned up in person.

Edgesense is a tool for community managers, rather than ordinary users. In the Matera's online community there are two main community managers and about 20 volunteers who coordinate through a mailing list. 9 people have admin credentials for the community, so about 10 participants were expected.

The dashboard is live on http://matera2019.edgesense.spazidigitali.com/ and is updated daily with data from the community. People started using Edgesense right from the beginning of the testing round, and the tool was well received by the users. In the three months period between June 1st and September 1st, the Edgesense has been used by over 230 people in about 400 sessions. The average session duration was less than 10 minutes, with the noticeable exceptions of the first week of June and the last week of July (coinciding with the workshop in Matera) where the average session was over 20 minutes long. Over the three months period, on average, the tool has seen about 57% of new users vs. about 43% of returning users. There is a slight tendency observed in the analytics that shows an increase in the percentage of new users over time. This, together with the bounce rate which is over 40%, and the fact that the tool is open to all the community users, **tells that probably the casual user has some interest in looking at the visualisations but that they cater more to the community administrator.**

As of today September 11th, 2014 the community includes 475 users, 193 of which have participated in the discussions (only users who have written at least one post or one comment are counted as participating). The users have produced 2 566 comments, which have been grouped into 870 individual pairwise interactions. These are the edges analysed by the SNA algorithms to extract relevant metrics. Edges are directed (A => B is not the same as B=> A) and weighted (two comments from A to B result in only one edge of weight 2 from A to B). **It can be observed that there has been an almost linear increase of the number of interactions in the online community over time.**

The share of community generated content is rising over time; this is the share of posts and comments written by ordinary users vs. administrators over time. The closer to 1, the more the conversation is self-sustained. **It is interesting to notice that these numbers are extracted straight from the tool dashboard** available at the link above.

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

2.1.3 Main pain points observed and improvements made

During the testing phase, various improvements to the dashboard have been implemented in response to direct users' requests and with the aim to make it simpler for the community manager to get used to the tool.

Among the UI/UX improvements, **Wikitalia improved the ability to handle communities with disconnected components,** e.g. small groups of people interacting with each other but not with the rest of the community. **The contextual information shown for each node have been enhanced,** showing the users and comments count and the community managers have been allowed, if they wish, to hot-link the users' personal page from the network visualisation (this allows the managers to jump from the graph to the users' page in the community.)

A series of bugs which were reported during the tests and related for instance to the handling of zero-degrees nodes have also been closed (i.e. peoples that never interacted with anyone.)

Further, **an interactive tutorial has been added, which purpose is to learn how to interpret social network analytics.** The main hope is that this will help to drive adoption and ease the new users' experience. The tutorial is embedded in the live dashboard and asks the user to imagine being a moderator of the community. The user is asked five questions during the tutorial, each addressing a concrete problem that they, as administrators, might have. Below the question, they will find a hint that contains information about how to use the Edgesense interface to answer the question.

2.1.4 Description of the results to date according to the aims of the tool

After each interactive tutorial was completed, the users were asked to rate the usefulness of the features they have been using. There have been very few complete test runs of the tutorial as of today:

- 14 replies total
- 3 replies before we implemented the survey
- 11 replies which include the survey responses

As data from incomplete tutorial runs are not collected, it has been deemed necessary to fix this and collect even incomplete answers to the survey going forward to be able to have more complete users' feedback. **Wikitalia is still processing the data collected by the dashboard analytics about the users' behaviour and patterns of interaction. This data will be very useful to improve Edgesense usability and user friendliness.**

Of all the tutorial answers, the following were the results:

Question	# correct answers	% correct answers
Look at your community's network of relationship and click on a highly central node.	13	92%
What percentage of all relationships occurring on your community do you think involve moderators?	6	43%
In the latest period for which we have data, how self-sustaining is your community?	2	14%
Over the last period, has the share of comments written by non-moderators increased or decreased?	5	36%
Does the presence of moderators increase or decrease modularity in your network?	5	36%

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

The numbers reflect that the graph visualisation is very helpful to make the most central users of the community emerge and it appears that finding such information is very clear, even to beginner users.

On the other hand, looking for instance at the results of the last three questions, the bottom part of the dashboard seems to struggle more in conveying the right information to the users. **This may also be the result of an objective difficulty of explaining the idea of sustainability in a meaningful way to a user untrained in the field of social network analysis.**

On the 11 surveys completed, users were asked how useful the features they had just used in the tutorial were perceived to be. The possible answers were: "very useful", "somewhat useful", and "useless". All the answers have been either "very useful" or "somewhat useful" with only one feature that was more polarising than the others, namely "how self-sustaining" has the community become.

How useful do you think this is that Edgesense lets you ...	**Very useful**	**Somewhat useful**
Discover how central each individual user is in the conversation	82%	18%
Investigate the role of moderators in the conversation	91%	9%
Measure how much moderators contribute to your community's activity	82%	18%
How self-sustaining the conversation has become	45%	55%
Discover and explore the subcommunities emerging from the conversation	82%	18%

Here are a few comments made by the users (translated from Italian):

- "*It is a little bit complicated for someone who is not used to using it*"
- "*Awesome tool*"
- *"We want to use it in our community*"

While these have only a qualitative value, they nonetheless **show interest in the tool** and seem to indicate that we are heading in the right direction.

Another encouraging sign is that **requests to install Edgesense on two communities besides our official testbed** of Matera 2019 have been received. Both run on Drupal 7. The first one, Innovatori PA, is an Italian community of (mostly) civil servants interested in open government themes. It has over 10,000 registered users. They have a live installation (http://156.54.105.206/), but it runs on an older version of Edgesense and the data is not updating as it should. They also have not received any demonstration on how to use it. The other community that requested Edgesense is that of the European project CHEST, who would like to use it for its idea platform. The discussion is ongoing and should result in the form of a proposal to be submitted in the framework of the Open Call for Collaboration.

2.1.5 Description of the main learnings and identified risks for the future

Overall, the tool seems to be working quite well and to be appreciated. The main scope for improvement concerns the interactive tutorial/test, which is a very useful and scalable way to teach oneself to interpret a network, however it is probably a slightly too difficult for beginners – and most moderators will be beginners. Wikitalia therefore intends to redesign it.

Among all the original pain points, Edgesense seemed like a good candidate to detect a potential for balkanisation (presence of multiple connected components with no bridge to each other) and individual isolation ("singleton" nodes, unconnected to the network). More generally, it seemed like a good tool for assessing the conversation's cohesion and identifying central individuals. In Matera, the network is quite cohesive and the risk of balkanisation

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

does not seem to be there. Some isolated nodes have been identified: community moderators reported those users had been active many months before Edgesense started to run, but not since. They conjectured that these users, not receiving any response on their first contribution, had dropped out of the conversation. **Edgesense, giving timely information on singleton nodes, can help direct the attention of moderators to new users who are trying to engage**.

An important result came in just at the time of delivering this report: different communities, even when based on the same software, have very different topologies, and Edgesense can convey intuition about topology simply by visual cues. This is done with no human intervention: the visualisation is built by the exactly same algorithm. Innovatori PA is more sparsely connected, with several "islands", about 100 singletons, and a very high modularity (notice the yellow cluster on the left, all connected through a single very important moderator).

Conversely, the Matera community (much smaller in terms of the number of users) is highly cohesive, with practically all active users connected to the giant component. Modularity is lower, average distance between nodes also lower. This result suggests that Edgesence could be also effectively used to compare communities' performances.

Figure 1 - The network of comments in the InnovatoriPA online community as seen by Edgesense. Color coding identifies (algorithmically) subcommunities.

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

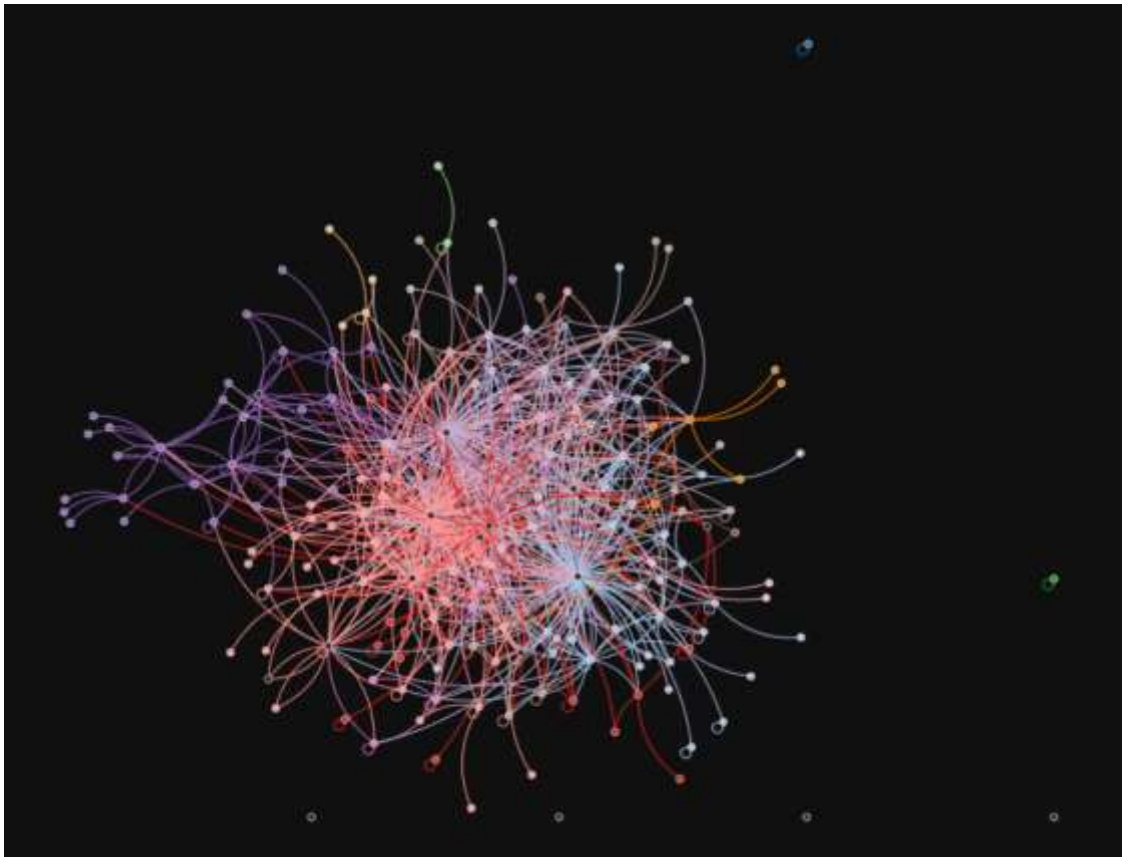


Figure 2 - The network of comments in the Matera 2019 online community as seen by Edgesense. Color coding identifies (algorithmically) subcommunities.

The main risks identified for this testbed are:

1. **Edgesense is perceived as difficult**. The failure rates in the tutorial test are alarmingly high. This could be due to a bad design of the tutorial itself, but it could also be fundamental: **thinking about an online community in terms of network is not easy**. There could also be a language barrier as the testers are all Italians.
   Mitigation: rewrite the language in the tutorial test.
   Mitigation: rewrite the test and translate its text.

2. **The community itself loses steam and interaction therein breaks down**. Starting in the spring 2014, the dormant Matera 2019 community has awakened to a new life. In October the winner of the European City of Culture will be announced; it is possible that a defeat drives down the enthusiasm for participation into this online community.
   Mitigation: take the test to new communities, like Innovatori PA and CHEST.

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

2.2 Testbed 2: Argument Mapping & Deliberation Analytics

2.2.1 The testbed

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.

This testbed involves testing the DebateHub, a new deliberation tool to allow large numbers of participants to collaborate online, providing mechanisms for a more structured debate and analytics to mediate attention, and improve debate's understanding, exploration and summarization.

2.2.2 Description of the testbed implementation from testing and technical partners

The pretest was launched internally within Euclid Network and Purpose starting in May 2014. Due to internal staffing and resourcing difficulties, the pre-test in this phase was not converted into a full testbed.

Involved partners have therefore pivoted the testbed environment in partnership with the School of Visual Arts in New York City and an international "philosophy incubator" called Wisdom Hackers. **A test bed will be launched with these communities in the middle of October 2014.** This timeline was determined in consultation with the external testing partners. In the case of the School of Visual Arts, this timing was set by the academic calendar, placing the testing period in a time when students were well oriented into their academic program and had enough context and background to successfully complete the proposed user tests. In the case of Wisdom Hackers, this timing is due to the publishing and launch schedule of the first Wisdom Hackers content.

2.2.3 Main pain points observed and improvements made

Several improvements to the UI were to set up the tool for the need of the testing community. Since a full test has not run yet, the answer to this question is partial. We can just report on general statistics of Debate Hub usage form the CATALYST technical partners, the pre-testing with Euclid and Purpose, the initial participation from one of the class of the School of Visual Arts in New York City and some involvement with the Decarbonet FP7 project, which is considering using DebateHub as tool to host debates on energy saving dilemmas.
Debate hub was so far used by 166 unique users in 375 sessions. The average session duration is 5,13 minutes. We can record 39.7 percept of new visitors and 1760 page views. One important non-technical insight is that partners learned that recruitment and engagement of testers is much harder than what they had anticipated, so they will have to reallocate more resources to user acquisition and engagement.

2.2.4 Description of the results to date according to the aims of the tool

No formal feedback from users has been gathered to date since the testbed has been rescheduled for October 2014. Partners intend to ask participants to fill out a questionnaire after using DebateHub to take part in a group deliberation virtual event (2-3 hours, geographically distributed).

2.2.5 Description of the main learnings and identified risks for the future

The main lessons learned at this stage are:

- Community recruitment and engagement requires more time/resources than previously imagined.
- Research goals and requirements (A/B testing, using a control group, etc.) do not always align with community needs/expectations (slick "commercial" user interface design, simple 'usability' versus full-featured functionality.)
- Finding the right theme/question is difficult but essential in getting the right kind of user engagement.

Partners will mitigate these risks by focusing outreach on existing communities with an audience to engage, rather than building an audience from scratch.

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

2.3 Testbed 3: Harvesting, Mapping & Analysing Arguments

Testbed 3 consists of two tests, each testing a different CATALYST tools for online debate's harvesting and mapping: Assembl (test 3.1) and LiteMap (test 3.2). In the following we detail the result of the two testbeds.

2.3.1 The Assembl's testbed

The communities participating are all pre-existing groups of participants having discussions on mailing lists around specific topics. As such they have a pre-existing discussion culture and already know each other. They also have experienced community facilitators that will be mobilized as expert harvesters.

2.3.1.1 *Description of the testbed implementation from testing and technical partners*

The test was launched on June 2nd, 2014 (with previous presentations on November 29th, 2013 and March 27th, 2014, at an initial workshop in Montpellier, France on May 14th and two harvesters training session on May 22nd and 26th).
The first testbed is now finished, after 4 weeks of online debates and one workshop held from July 1st to 4th in Brest, France. A second testbed is planned for September to include online creativity support, improve engagement and pledge testbed inside a Harvest-Mapping cycle.

340 persons, members of Anim-fr group, have participated in the test. Four of them were chosen and trained to be harvesters. The next test will run with the same group (and with about the same number of people). Another test may be organised also with the Fablab group which has currently 323 members worldwide. The partners were not able to launch this testbed at that time because of software performance: it is indeed not possible to test a debate with more than about a hundred contributions whereas the Fablab list has had already 300+ messages since Assembl was plugged into the list.

2.3.1.2 *Main pain points observed and improvements made*

A first debate mapping was carried out and the synthesis was sent on June 9th, 2014. During harvesting, a bug considerably slowed down the time required (it was necessary to keep on relocating the following mail). The development team gathered and efficiently resolved this problem during the week, which undoubtedly helped the work of the testers.

42 comments were produced after the 4 harvesting sessions: 36 concerned the software and 6 the documentation. These comments were then sorted into 3 categories:

- A : Consideration planned for September
- B : Consideration planned for later
- C : Discussion currently underway

Some of the comments made during the test and some additional bug fix have been handled during the summer in order to have a new version of Assembl ready in September for the next tests.

The synthesis semi-automated production function was not yet implemented in July 2014. To cope with this issue, the test organiser designed a synthesis "by hand" from the table of ideas produced by the harvesters in order to send the weekly synthesis to all the members.

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

2.3.1.3 *Description of the results to date according to the aims of the tool*

All four harvesters succeeded in fulfilling their tasks, i.e harvesting and mapping. This exercise took them an average of 2 hours (the typical harvesting/mapping session without tools takes 5 hours). However, the goal is to shorten this duration up to 1 hour in order for the harvesters to be able to complete this task during their work time.

The participants (mainly group coordinators and community managers) worked on the topic of "how to release stress in a group?" The feedback provided both by the online group and the group physically attending the workshop in Brest (this one included people who were not participating in the test but discovered the results of the debate) is very positive: **this debate proved to be a major breakthrough on the topic**.

The testbed met its expectations on the outcomes of the debate (an unexpected solution to a complex problem with a general convergence on some of the best solutions proposed). However, it was not possible to test Assembl on a group having already a large number of mails (capitalisation of all the discussions of one group rather than focusing on only one debate). The synthesis semi-automated production function was not implemented and then tested.

One of the major expectations of the software is to shorten the weekly harvesting/mapping session from 5 hours (with the methodology but without tools) up to 1 hour. (methodology + tools). The current version of Assembl leads to an average of 2 hours per harvesting/mapping session.

2.3.1.4 *Description of the main learnings and identified risks for the future*

The training of harvesters was a major issue. During the pre-tests run by the test organisers and the two harvesters-training sessions, the use of the software was quite difficult. **The developers were very reactive in taking into account the difficulties expressed by the testers.** They were also reactive after the first real harvesting session on June 9th: corrections were made during the week so that the second harvester could work with a much more usable software. The harvesters were very willing to participate and understood quite well the first difficulties. Assembl is based on a methodology to develop Collective Intelligence, which describes an iterative process of synthesis/comments[1]. The participants used mainly the mailing list to contribute and then saw no difference on the methodology used with or without the tool. A test of the methodology without the tool had been previously done in the same group.

Developers participated in the group to see what was happening "inside". Some of them actively participated to the discussion and made some very relevant contributions to the debate; t**his was unexpected and shows how a very large group of participants with very low interest and implication can get feedback even from some of the inactive "observers"**.

The main risk is the harvesting/mapping productivity: those who agreed to be harvesters did not have a lot of time to spend on this test (most of them have their own group to manage, while Anim-fr is just a group of coordinators to help empowering its participants). The core Assembl functionalities are now working quite well, but should be associated to a strong methodology in order to produce collective intelligence.

There are still some performance improvements needed for the software (managing large numbers of mails - see above) and for users of Assembl (duration of harvesting sessions – see above). Moreover, the semi-automated production of the synthesis was still missing at the time of the test (see above) even if it has been almost completely developed since.

Other functions planned in the CATALYST project should also be added:

- Online creativity support
- Improving engagement and pledging

[1] How to Produce a Document When You are Several Hundred People: http://ebook.coop-tic.eu/english/wakka.php?wiki=HowToProduceADocumentWhenYouAreSeveral

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

- Plug in any discussion tool (mailing lists, Facebook, but also simple mails for those who want to receive only the synthesis, not all the comments).

2.3.2 The LiteMap's testbed

Different discussions on sustainable consumption and lifestyles are taking place in various discussion forums of the community on the Utopia website. Many of the discussions tackle similar topics which are used by the harvesters to connect different discussion streams and additional resources from the web in order to show the connection and interaction between the different ongoing discussions.

2.3.2.1 *Description of the testbed implementation from testing and technical partners*

Figure 3: Screen shot - Utopia website home page with advert pointing to the testing site

As already specified in D2.4, the first test round carried out by the community partner CSCP was supported by the technical partner Open University and its tool Litemap. This change of technical partner compared to what was mentioned in the DOW was decided to best fit the pain points identified by the testbed community relier in the project. Highest interested was shown for a tool that would be able to show different discussion streams in an argument map. Additionally, the nature of the online community asked for an integration of the tool into the existing website, a function, which Litemap could fulfill with the help of an iframe.

From a technical perspective the existing platform used by CSCP community members (Utopia) was controlled by a third party and not open-sourced, so it was technically impossible to build a bridge between that platform and

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

Assembl. Therefore since I4P did not have access to the platform code and involved partners thought that asking participants to switch online discussion platform was an obstacle to adoption, an alternative approach was chosen to use LiteMap, which allowed to add an argumentation layer on top of the Utopia Platform without requiring full integration.

The testbed for the LiteMap tool with the German online platform on sustainable lifestyles and consumption Utopia started on July 3rd and ran officially until August 22nd. Over the course of these six weeks, 3 argument maps were integrated into the website with the help of an iframe. The Utopia newsletter, which is sent out once a week on Thursday, announced each new map to the audience. Additionally, the newsletters of the weeks in between focussed on a topic related to the topic addressed in the map. Over the entire period of the testing, an advert at the side bar of the page pointed to the testing site (*see screen shot above*).

The site in which the maps were embedded also explained how to use the map, information on the CATALYST Project, as well as a link to a discussion forum for the content of the argument map and a feedback survey.
During the testing period, over 800 people navigated to the argument maps. Most traffic was generated after the newsletters were sent: after each one, around 130 visitors came to see the argument maps on the same day).
In total, 57 people started and 27 completed the survey on the usability of the tools. Additionally, 4 people posted feedback on the usability directly in the discussion groups.

The first of the three maps generated most discussions in the forum. However, this can also be attributed to the nature of the question addressed in the map, which was more a debate between two ideas, whereas the others showed different aspects of one complex issue. Three expert harvesters created the argument maps as planned before. The number of engaged audience exceeded the 100 people planned.

Cooperation between community and testing partners worked very well. Exchange of emails was sometimes up to 10 times a day (especially related to bugs in the programming); the experience of creating the map was also shared between partners through several telephone calls and interviews.

2.3.2.2 *Main pain points observed and improvements made*

In the internal testing phase between testing and technical partners, many smaller bugs and usability issues were changed. This included, for example, the size of the text displayed in the boxes of the argument map, the content that could be seen via a roll-over, the colour scheme of the different layers of the map, the possibility to move the arrows around freely in order to connect boxes which belong to different ideas, etc.

Based on the users' feedback from Utopia, some navigation items were increased in size to **make navigation easier**. As some people explained that they found the map's structure sometimes confusing and were interested in a linear view as well, this option was also included via a switch to other view button.

For the harvesters, a function of **an open comment was added** as it proved over the process of creating maps that this made the initial collection of web quotes before the structure of the map was developed easier.

2.3.2.3 *Description of the results to date according to the aims of the tool*

In total, 57 users started to fill in the survey. While 30 of them filled it in partially, 27 persons completed the survey entirely. Twice as many women as men (voluntary response/ 24 answers of possible 27 answers) participated in the survey.

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

Most of the users who filled in the survey visit the Utopia-Website several times a month. A lot of users visit the website several times or, at least, once a weak. Nevertheless, most of them do not participate very often in discussions on the website: 44 % participate less than once a month. 32 % even never participate in discussions on the website. A very small number of the users (4 %) participate once a week, while 8 % participate several times a week.

According to the survey, the users of the "Argument Map" did not agree on the question if the software helps to structure discussions more efficiently: while more than one third thought that the software was helpful, the same amount of people did not categorise the software as helpful. Only 12, 5 % said that the software is definitely helpful. A few users added additional comments about the map:

- Almost one third (26, 09 %) liked the map more or less, 17, 39 % definitely liked it and the same amount did not like it at all. More than half of the participants of the survey would use the "Argument Map" again while almost 40 % do not think that they will use it again.
- They mentioned that it would be good to improve the visualisation of the map and that it was not that clearly structured. In general, they liked the idea of such an argument map but they also mentioned that it was not that easy to use this tool. Besides, the majority of the users (54, 17 %) thought that the use of the "Argument Map" was not easy. Only a minority (4, 17 %) said that it was definitely easy. The navigation was also criticised due to the fact that the users did not find it very easy. In this context, the implementation of a search function was suggested. The users mentioned that it would be great to be able to search for a certain argument in the map without having to study all the other arguments before. The opinions about the user-friendliness were rather negative (28 % answered with "rather not" and 16 % with "not at all", 28 % found that the user-friendliness would be average).
- More than a third of the users had the impression that the software helped them to clearly understand the course of online discussions. 12, 5 % were convinced that it definitely helped them, while the same amount of users was not sure about the utility of the software. Almost 17 % said that the software did not help them at all.
- Not everybody approved the structure of the "Argument Map". Furthermore, other starting points and ideas as a basis of the map were desired. The font size was considered as too small. One contributor felt that the map would be good for pupils but not for older users.
- About one third gained a better understanding of the topic by using the software. Nevertheless, 36 % said that it did not help them to get a better understanding. The majority of the users found that the software did not help them to make better contributions in online discussions (36 % answered with "rather not" and the same amount answered with "not at all").

To put these answers into perspective, it has to be seen that, so far, the Utopia platform has not been using any features other than forums with linear commenting functions for online discussions. This means that users are not used at all to debating in other ways or to seeing complex topics in more structured forms like the IBIS model. Often, understanding the model and appreciating its structure take some time and practice.
Further, it could be seen that design and easy usability are criteria that the Utopia-users rank very high. **These users might be much more willing to spend time on a tool which looks modern and is very easy to navigate.**

Two people who saw the argument maps approached the testing partner asking if they could use the program for their own work. This can be seen as a very positive feedback, as the tools are perceived as useful up to a point that **people want to bring them into their own communities**.

2.3.2.4 *Description of the main learnings and identified risks for the future*

Balkanisation was mentioned as one of the key pain points by Utopia's users before. While the map brought together the arguments in one place, it could obviously only capture the content debated on different places and could not physically connect the discussions. When trying to inform the people of the different discussion threads about the map and the other content on it, these posts were seen as spam by some users and were hence deleted. However, the argument maps could show the different aspects of complex topics, which had been raised all over Utopia.

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

What has not been sufficiently taken into account is the tool's design. Feedbacks showed that users are looking for very well designed and modern online tools. This is what they are used to experience on the Utopia site. Several remarks showed that the visuals and the navigation on the argument map as it is now were not as modern as users expected.

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

2.4 Testbed 4: Online Creativity Support

2.4.1 The testbed

Virtual creativity card facilitation consists in presenting participants with a certain number of cards selected from a set and asking them to have either individual or group interaction with the card, and present their findings to the group. The colocation widget is designed to foster creative ideation by presenting users with 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.

2.4.2 Description of the testbed implementation from testing and technical partners

For deliverable D3.7.1 « co-occurrence creativity tool », Imagination for People is developing 2 scenarios:

1. A function called “Inspire me” makes it possible for participants in a debate to get inspired by videos on YouTube prompted by key words
2. A function called “Creativity session” in which the harvester prompts a specific collective session (animated by the harvester) around YouTube videos

With regards to the “inspire me” function, it is a self-standing functionality that did not require a large testing environment other than individual users in the I4P community. Indeed, the testing took a few minutes per user, and all feedbacks proved to be positive.

As per the “Creativity session” on YouTube, I4P realized that a functional and experiential modification were required before launching the test. Indeed, unlike D3.7.2 « gamification creativity tool » (based on cards), the YouTube widget does not integrate any social functionality that would allow group discussions on the video itself. It means that such informal exchanges would have to be mixed with the general discussion, with associated pollution and disparity of "tone" with the more "serious" main discussion. Moreover, if adoption of the tool by a large number of participants indeed materialised, the interaction between the participants would be diluted throughout many videos, probably leading to too few participants for a meaningful brainstorming on each video, or only the first videos posted becoming hosts to significant interactions.

Imagination for People mitigated that risk by changing its value proposition. I4P anticipated that positive results could be achieved by asking the moderators or some subgroup to find inspirational videos with the co-occurrence tool (card-based widget), from which the harvester would generate a deck of "cards" usable in D.3.7.2, where each card is a video. This widget is specifically designed to foster quick group interactions on a limited number of common "cards", enabling conversations to develop. Since the interaction between the two tools had not been planned or budgeted, Imagination for People first intends to validate the concept against a panel of community managers before investing in its development. But it is very likely that such an interaction be quickly developed and implemented.

The D3.7.2 « gamification creativity tool » based on cards is technically ready. Nevertheless, preliminary internal tests (within I4P) proved that the initial set of cards was not a sufficient kickstart for new creative insights. Indeed, creative brainstorming sessions on the Web require very specific prompters (text-based rather than image-based). Therefore, a finer selection of the right textual cards, which is key to stir up a fruitful creative discussion, was required. To deal with this problem, Imagination for People started to prepare a deck of generic cards (around 80) directly applicable to generating as many options as possible for any topic.

2.4.3 Main pain points observed and improvements made

For the reasons mentioned above, the testbed involving 340 members of the group Anim-fr has been delayed to September in order to:

1. Develop the right video interface in D.3.7.2 for use in video-based creativity sessions;
2. Finalize the right set of textual cards for card-based creativity sessions.

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

2.4.4 Description of the results to date according to the aims of the tool

Waiting for availability

2.4.5 Description of the main learnings and identified risks for the future

Waiting for availability

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

2.5 Testbed 5: Improving Engagement & Pledging

2.5.1 The testbed

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. Random option ordering widget will be used when the group wants to choose projects to focus on.
Multi-axis voting widget will also be used to rate the ideas themselves when their level of formulation is stabilized in the collective intelligence process.

2.5.2 Description of the testbed implementation from testing and technical partners

The software has been released. The integration with the synthesis and the idea panel will allow for a test in December with the participation of 340 members of the group Anim-fr.
The test will cover both the voting on the synthesis and on the idea panel in the Assembl interface.

2.5.3 Main pain points observed and improvements made

Waiting for availability

2.5.4 Description of the results to date according to the aims of the tool

Waiting for availability

2.5.5 Description of the main learnings and identified risks for the future

Waiting for availability

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

2.6 Testbed 6: Collective Analytics Dashboard Usability Evaluation

2.6.1 The testbed

The Collective Intelligence Analytics Dashboard aims to make multiple visualisation of a deliberation available to moderators or end users in one place to allow better access and interpretations of the visualisations. The testing of the dashboard will consist of three main phases:

- Testing Dashboard Visualisations
- Testing Dashboard Use
- Testing Dashboard Usability and Usefulness

2.6.2 Description of the testbed implementation from testing and technical partners

The first testing round of the Collective Analytics Dashboard will be completed and reported on in November (M14, D4.6), so this test has not yet been finalised. Hence the information provided below regards mainly the planning, evaluation design and preparations carried out so far but they do not contain statistics and results in terms of usability results and users' participation and feedback.

In a nutshell, the collective intelligence analytics dashboard usability evaluation will test the usefulness and usability of several deliberation analytics visualisations provided by the CI dashboard (T3.9). The dashboard visualisations are generated from data specified in the CATALYST interchange format and the CATALYST metrics server.

The evaluation will make use of a mix of qualitative and quantitative methods in lab and field experiment settings. The experiments are designed in a way that participants work on realistic tasks in which they will answer questions about the debate with the help of the analytics visualisation. The methods used for the evaluation will range from questionnaires, video recordings, to web analytics, providing a rich data source about aspects of usefulness, usability, and use of the collective intelligence dashboard visualisations. Three main settings are evaluated, which structure the evaluation task.

- The first setting will be a study in a usability lab. It will collect rich data about the usefulness and usability of different visualisations in a controlled environment and with a small group of participants.
- The second study will be a real-world experiment, as it will involve participants form other testbeds, which have used at least one of the CATALYST deliberation tools, to provide feedback on the potential usefulness of various analytics visualisation to enhance their deliberation process. Usability and usefulness data will be gathered via questionnaires and possibly from a larger group of users.
- The third setting will gather usage data in an unobtrusive way. Data will be generated by users during the interaction with the Debate Hub dashboard visualisations in a trial organised by Purpose.

2.6.3 Main pain points observed and improvements made

The testbed is expected to be launched in October.

Currently, Open University is preparing the set-up for the different evaluations and are organising the logistics for the conduct of the experiments. Several visualisations have been prepared and are about to be integrated into the Collective Intelligence Analytics Dashboard. Several realistic tasks have been prepared for the evaluation of each visualisation. Open University is preparing the questionnaire in order to collect information about the usefulness and usability of the analytics visualisation.

In order to provide with an example of the type of tasks and questionnaire designed, below is, for instance, the task set up for testing the usefulness and effectiveness of the Topic Space Visualisation.

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

Usefulness and usability study of the Catalyst Collective Intelligence Dashboard

Welcome to our survey about your experience of the analytics dashboard.

Your contribution will be of great help to improve the analytics dashboard and as a consequence it will help to improve your experience of debating with PURPOSE.

This short survey will not take longer than 1X minutes.

This research is conducted by the Open University, UK in the scope of the CATALYST project (http://catalyst-fp7.eu) funded by the European Commission. If you have any questions about this questionnaire or the accompanying research, please do not hesitate do contact Thomas Ullmann (thomas.ullmann@open.ac.uk).

Consensus
Are you willing to take part in this research project, and do you give your permission for the data collected to be used in an anonymous form in any written reports, presentations and published papers relating to this study?

☐ Yes ☐ No

Background information

Gender:
☐ Male ☐ Female

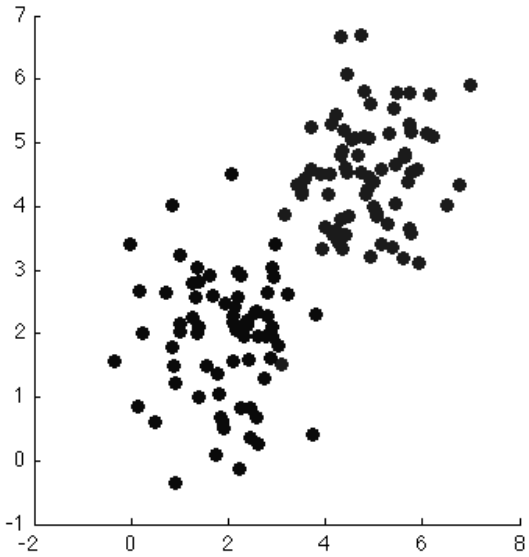
How familiar are you with analytics dashboards?
☐ Expert ☐ Advanced ☐ Average ☐ Basic experiences ☐ Novice

How familiar are you with visualisations for analysing and exploring data?
☐ Expert ☐ Advanced ☐ Average ☐ Basic experiences ☐ Novice

How familiar are you with visualisations for analysing and exploring debates?
☐ Expert ☐ Advanced ☐ Average ☐ Basic experiences ☐ Novice

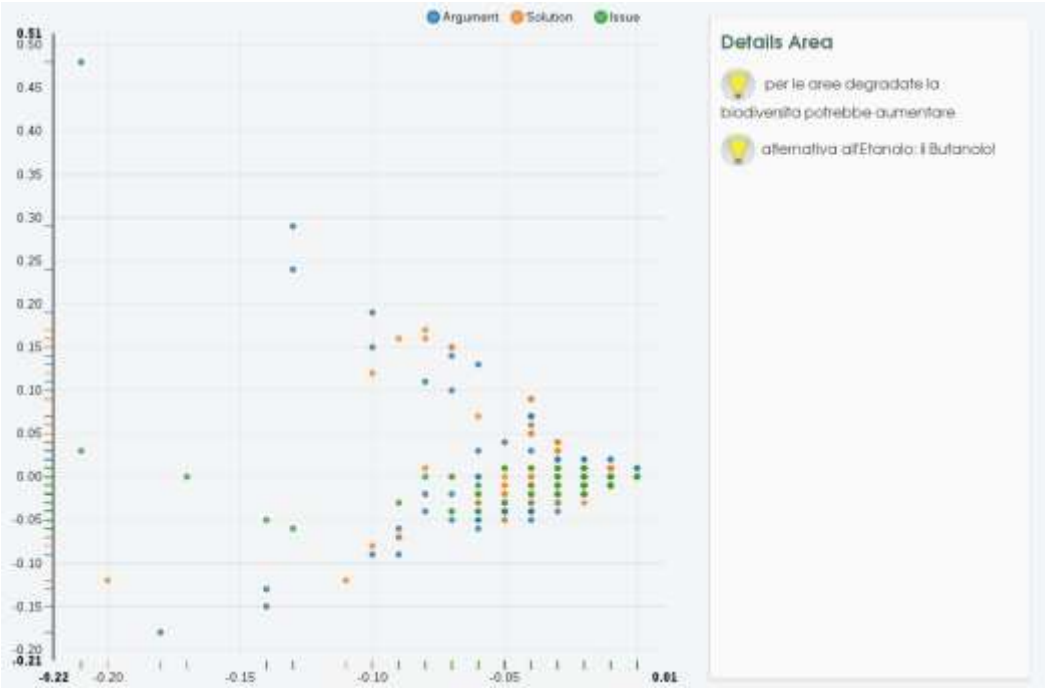
Topic space visualisation
The following visualisation shows posts arranges on a space. The important bit about the visualisation is to look out for groups of dots. The following example contains two groups (or clusters). One group is on the bottom left, and one group is on the top right. Often there are no distinct groups.

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



Please familiarise yourself with the visualisation clicking on the dots.

Once you are ready proceed to the questions. You will be asked a few questions which you can answer by using this visualisation.



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

About the visualisation:
How many groups can you see? ___

In your opinion:
Is the debate dominated by one or more group?
☐ Yes ☐ No ☐ Cannot tell
Please explain:
____

Is the debate balanced?
☐ Yes ☐ No ☐ Cannot tell
Please explain:
____

What can you understand about the debate from the information provided by the visualisation?
____

Can you identify a pattern or tendency by using this visualisation?
____

Grade from 1 to 6 if the visualisation shows a good (1) or a bad (6) debate
☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6
Please explain:
____

What actions would you recommend after seeing this visualisation?
____

If any, what further information should the visualisation provide in order to help you determining if the debate is a good one or a bad one?
____

Usability questions (Strongly disagree (1) to Strongly agree (5))
Please answer the following questions after you have completed the task

I think that I would like to use this visualisation frequently
Strongly disagree (1) ☐ ☐ ☐ ☐ ☐ Strongly agree (5)

I found the visualisation unnecessarily complex
Strongly disagree (1) ☐ ☐ ☐ ☐ ☐ Strongly agree (5)

I thought the visualisation was easy to use
Strongly disagree (1) ☐ ☐ ☐ ☐ ☐ Strongly agree (5)

I think that I would need the support of a technical person to be able to use this visualisation

Specific tasks and questions will be designed for each Dashboard's visualisation to be assessed.
Negotiations with the usability lab are under taking place and will be soon concluded. For the collection of the usage data, a dual approach will be taken by tracking the analytics visualisation usage with common web-analytics software, but also with a custom facility to track visits.

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

2.6.4 Description of the results to date according to the aims of the tool

The results will be reported in D4.6, due November 2014, after the experiments have been conducted.

2.6.5 Description of the main learnings and identified risks for the future

The analytics visualisations mainly aim at reducing the pain points of poor visualisation and poor summarisation (see D2.1 for a list of pain point by importance): The analytics visualisations are designed in a way that they allow to get a quick overview of important aspects of a debate and they highlight important facets of a debate. Furthermore they help users to explore the data in a variety of ways in order to spot important patterns, which may inform about potential problems, or allow comparing debates.

The evaluation of the analytics visualisations depends on a good amount of participants. Therefore one of the potential risks for this testbed is that not enough data can be collected in order to conduct a thorough quantitative analysis (especially for the third setting context reported above, which requires tracking visualisation usage in the context of other testbeds). To mitigate this risk, the efforts will be shifted to analyse the qualitative data in more detail.

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

Conclusions

In a first cycle of tests, which evaluation is presented in this document, several specific technologies were tested through the launching of real world tests by different CATALYST community partners. The participants of the tests were called upon to take part in discussions on topics and issues of importance to them in a community setting.

Despite the fact that the technologies tested in this first cycle are in a very early phase, this real life deployment of the seven testbeds is an essential step of the CATALYST project's development. Even if all the testbeds have not yet started, because some technologies still need to be further developed to ensure efficient tests, it is to note that, as predicted, the initial round of trial did provide each partner with rich learning opportunities and keys to better reach their tools' objectives.

By spotting the major discrepancies, pain points and administrators/users' difficulties, this first testing round has allowed both technical and community partners to correct, modify and ameliorate their tools and features, on a usability level. The feedback given by the tools' users, the community managers and/or "harvesters" in some cases, have been taken into consideration by the community partners and followed by concrete actions. The information gathered during the testing sessions has also permitted all partners to measure if the feature tested indeed corresponds to the objectives for which the tools were created.

From a community engagement perspective this first phase of testing was key to demonstrate for the first time all the newly developed CATALYST's tools in action. Edgesense, DebateHub, Assembl and LiteMap have been used by several real communities and community's feedbacks are overall positive in terms of usability and usefulness; tools were tested both with community members and with more skilled community's moderators. A main result of this first round of testing is also the identification of possible risks and barriers to engagement (such as community training and recruitment processes) and provided a better understanding of what should be improved in the second testing round to reach wider participation.

One of the main lessons learned of this exercise is also that collaboration between the technical side and the purely testing partners has smoothly and efficiently worked. Several feedback questionnaires indeed highlight that a constant dialogue between the two parties has flown naturally. The reactivity of the technical partners when a bug was detected has also been stressed upon and appreciated by the non-technical side.

All modifications and enhancement brought to the tools significantly contribute to constantly improve the tools and their usability in this first cycle of tests that is still in progress. Further, these improvements are also crucial for the launching of the second round of tests (month 19) on both technical and design levels.

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